Education Reform and Development, 2025, Volume 7, Issue 1
https://ojs.bbwpublisher.com/index.php/ERD
Online ISSN: 2652-5372
Print ISSN: 2652-5364

A Study on the Errors of Machine Translation in Political Propaganda Texts of the Energy Field

Man Chen*

China University of Mining and Technology, Beijing, Beijing 100083, China

***Corresponding author:** Man Chen, c_man927@163.com

Abstract: In this study, the paper "China's Energy Development in the New Era" was selected as the research object, and three representative machine translation tools, Baidu Translate, Google Translate and ChatGPT, were used. Through comparative analysis, this study aims to reveal the differences in the number and types of errors shown by these three translation tools in processing specific texts, and provide implications for post-editing.

Keywords: Machine translation; Types of translation errors; Energy field; Post-editing

Online publication: February 10, 2025

1. Introduction

As a key technology in the field of natural language processing, machine translation not only facilitates cross-language communication but also promotes the process of globalization and shows great potential in many fields, including business, scientific research and government services. However, it still faces many challenges when processing specialized texts in specific fields. Problems such as linguistic ambiguity, cultural differences and the accuracy of professional terms limit the translation quality of machine translation [1]. In this study, the paper "Energy Development in China in the New Era" was selected as the research object, and through a comparative analysis of three mainstream machine translations (Google, Baidu and ChatGPT), the possible types of errors in machine translation in the field of energy propaganda text translation were discussed and the impact on post-editing was assessed.

2. Literature review

Since Jing (1959) first published his research on machine translation in China in 1959 [2], China's research in the field of machine translation has been continuously developed and prospers. Zhang *et al.* (2011) carried out an in-depth analysis of errors in machine translation practice [3]. They analyzed the causes of machine translation errors in five steps from the four levels of technology, language, thinking and management and put forward three-

dimensional solutions [3]. Luo *et al.* (2012) conducted a parallel and comparative study of machine translation and manual translation and described in detail the typical error forms displayed by machines in the translation processing of vocabulary, syntax and symbols [4]. Li *et al.* (2013) conducted a comparative analysis of 100,000 English-Chinese translation sentences between machine translation and human translation and found that the error rate of machine translation was more than 80%, of which lexical errors accounted for 70% [5]. Luo (2014) further labeled and counted the mistranslations of syntactic structures such as noun phrases and phrasal verbs in machine translation, and summarized and analyzed the typical forms of various syntactic mistranslations [6]. In the early stage of the research and development of machine translation errors, scholars mainly focused on the applicability of machine translation in the general field.

In the practice of vertical domains, machine translation errors have become the focus of many scholars. Scholars have studied the types of machine translation errors in different fields such as politics, economy, literature, natural science, the petrochemical industry, medicine, journalism and communication, management, e-commerce, and other key technologies. For example, in the political field, Cai *et al.* (2021) found the commonality of machine translation at the word and sentence levels by parallel comparing the Chinese-English translation cases of speeches on the website of China Daily [7]. In the economic field, Zhang (2021) used a case study to analyze the types of common errors in the Economic Report of 2018 [8]. In the field of literature, Liu *et al.* (2022) conducted an in-depth comparative analysis of Chinese-English machine translation of online literature [9]. Yu *et al.* (2019) analyzed errors in online machine translation of scientific and technological texts [10]. Ren *et al.* (2023) analyzed the errors in machine translation of petrochemical texts [11]. Tang *et al.* (2020) evaluated the translation of medical texts by online machine translation software [12]. Fang (2018) analyzed the Chinese translation errors of machine translation of business texts [13]. Zou (2015) analyzed translation errors in machine translation of management texts [14].

3. Research methods

Taking the paper "China's Energy Development in the New Era" as the research object, this study targeted three popular online machine translation interactive platforms at the present stage, taking the official translation as the reference, and used two methods of machine evaluation and manual evaluation to analyze machine translation from different dimensions, to reflect the situation of machine translation in the Chinese-English translation of energy political propaganda text.

The two main questions of this study are: (a) What are the errors and characteristics of the Chinese-English translation of the energy policy propaganda text under machine translation? (b) What are the differences in terms of error types and percentages among the three machine translation tools?

4. Results and discussion

In this study, three kinds of machine translation errors were marked manually by using manual labeling software, compared with official translation. Based on the internationally accepted translation quality assessment system MQM, machine translation errors are classified and analyzed statistically.

4.1. Terminology errors

There are a large number of energy terms in the political propaganda text in the field of energy, which are

highly professional and require accuracy. The error rates of the three machine translation engines are quite different in terms, of which Google Translate has the highest error rate, Baidu Translate has the lowest error rate, and ChatGPT has the middle error rate, which indicates that there is still much room for improvement in terms of machine translation. For specific texts, terminology errors are mainly reflected in the popularization of specialized terms, such as "cut-and-fill mining," and "water-preserved mining," Google Translate translated it as "backfill mining" and "water conservation mining", which directly adopts the word-to-word translation strategy, which lacks professionalism and accuracy.

4.2. Semantic failure

Accuracy is the first and foremost evaluation index of a translation. Fidelity to the original determines the quality of the translation. Errors in the accuracy of the translation mainly include mistranslation, addition and omission. The statistics found that all three machine engines had a high mistranslation rate, while Baidu had the lowest accuracy, Google and ChatGPT were about the same.

4.2.1. Mistranslation

Mistranslation refers to errors in the understanding and expression of the sentence, and the original meaning of the translation is distorted. Among them, Baidu's error rate is the highest, nearly twice that of the other two, and Google and ChatGPT's error rate is not much different. The data analysis found that the three most common mistranslations in machine translation were "inanimate subjects with animate predicates." For example, Google Translate Engine: "It is making strong headway in the fields of smart grid and large-scale power system control" is translated to" Smart grid, large power grid control, and other technologies have made significant progress." In this sentence, "Smart grid, large grid control" is an inanimate subject, and the subject of "make progress" should be spiritual, or it can be understood that "China" has made progress in these two aspects.

4.2.2. Addition

According to the data analysis, the three machine translation engines, Google and ChatGPT, rarely fail in terms of addition. Baidu translation only had 13 errors. In addition, most of the errors in Baidu Translate were the same translation of synonyms, resulting in the repetition of synonyms, such as "oppose extravagance and waste and unreasonable consumption." Baidu translated it as "guide the establishment of a consumption concept of diligence, and thrift," and "extravagance and waste" and "unreasonable consumption" were translated as "thrift". Semantic duplication appears.

4.2.3. Untranslated

Due to the lack of an "alignment" module in the neural machine translation model, untranslated errors often appear in the translation. Untranslated, that is, "the translation is empty," finally leads to incomplete meaning [15]. For untranslated, the error rates of the three translation engines are not much different.

4.3. Grammatical and verbal errors

There are various types of grammatical collocation errors, but the problems in machine translation mainly include fragments, run-on sentences, verb errors, punctuation errors, etc. The verbal errors of machine translation are concentrated in the two aspects of verbs and nouns, and other types of errors are very rare.

4.3.1. Fragments

An incomplete sentence is when a sentence is missing key components, such as a subject, predicate, and so on. There is little difference between the three translation engines in terms of sentence incomplete errors, mainly reflected in the absence of subjects or predicates in the translation. For the common non-subject sentences in Chinese propaganda texts, the three translation engines did not pay attention to adding logical subjects but translated them literally into verbs and objects. For example, ChatGPT translated "implement a special action to upgrade the quality of refined oil" to "implement a special action to upgrade the quality of refined oil." In addition, the machine translation engine also suffers from the error of missing predicates, which is particularly prominent in Baidu Translate, for example, the translation of "Preliminary formation of a flexible, safe and reliable natural gas transportation system for scheduling."

4.3.2. Verb errors

Verb errors mainly refer to verb tenses and errors in single and plural forms. In this regard, Baidu translation has the highest error rate, followed by ChatGPT translation, and Google Translate Engine has a lower error rate. The errors of the three engines in this aspect focused on tenses, and there were fewer errors in the single and plural verbs. Baidu translation translates some past and even present tenses into future tenses, mainly because the machine has no specific context input, in addition, because the propaganda text is more formal, Chinese verbs themselves do not have tense changes, so the translation engine is difficult to determine whether the event has happened or will happen, so there will be errors.

4.3.3. Punctuation errors

In machine translation, some punctuation marks are not used following Chinese expression habits, and most of them are formatting errors. In contrast, Google and Baidu have more errors, and ChatGPT has only one error. Punctuation errors in Google Translate focused on adding a period before the end of a sentence, or the lack of one, as in "and achieves its development at the same time". In Baidu Translate, the errors were mainly caused by adding unnecessary commas.

4.4. Localization specification errors

Localization specification refers to errors in writing specifications for numbers, currency, time, etc. According to the statistics, the Baidu translation error rate is the highest, the Google error rate is in the middle, and ChatGPT only made six errors. Baidu's errors were mainly due to the absence of commas as separators for numeric expressions, such as "78,000" and "100,000." The Google Translate error is mainly a misrepresentation of the unit of measure.

5. Conclusion

To sum up, different machine translation engines have their advantages and disadvantages, but the errors are mainly reflected in mistranslations, terminology errors, localization norms, verb errors, and noun errors. Among them, mistranslation accounts for the largest proportion, indicating that the machine translation engine also needs to improve the accurate understanding of the original text. In addition, it also needs to strengthen the accuracy and standardized expression of professional terms.

Through the analysis of the types of machine translation errors in energy political propaganda texts, some

references can be provided for post-editing. In order to improve the translation quality of government propaganda texts in the field of energy, it can be started from the following points:

(1) We should establish a sound professional terminology database in the field of energy political propaganda. The research finds that terminology errors are common errors in machine translation. If we can solve this problem, the translation quality of energy political propaganda texts will be greatly improved.

(2) Post-editors should pay attention to the coherence and understanding of context when checking long and complex sentences to avoid mistranslations and omissions in the process of machine translation. In addition, AI-assisted tools such as translation memory, language style and context detection tools can be used to improve the efficiency and quality of translation.

Funding

2024 Undergraduate Education, Teaching Reform and Research Project of China University of Mining and Technology (Beijing), "College English Digital Assessment Reform" (Project No.: J240803)

Disclosure statement

The author declares no conflict of interest.

References

[1] Cui Q, Li W, 2015, Research on Types of Post-Editing Errors: Based on E-C Machine Translation of Scientific and Technological Texts. China Science and Technology Translation, 28(4): 19–22.

[2] Jing S, 1959, Special Issue on Machine Translation in Linguistic Translation Cluster, 2nd Issue 1959. Journal of Western Languages, 1959(3): 190–156.

[3] Zhang X, Zhao Y, 2011, The Construction of "Human-Machine Thinking Rendering" from "Machine Translation Error." Chinese Journal (Foreign Language Education and Teaching), 2011(8): 60–62.

[4] Luo J, Li M, 2012, Analysis of Translation Errors in Machine Translation. Chinese Journal of Translation, 33(5): 84–89.

[5] Li M, Zhu X, 2013, Classification and Statistical Analysis of Errors in E-C Machine Translation. Journal of University of Shanghai for Science and Technology (Social Science Edition), 35(3): 201–207.

[6] Luo J, 2014, Syntactic Error Analysis in Machine Translation. Journal of Tongji University (Social Science Edition), 25(1): 111–118 + 124.

[7] Cai X, Wen B, 2021, Statistical Analysis of Errors in Chinese-to-English Machine Translation: A Case Study of the Translation of Publicly-Published Texts. Journal of Zhejiang Sci-Tech University (Social Science Edition), 46(2): 162–169.

[8] Zhang J, 2021, Analysis on Types of Errors in English-Chinese Machine Translation and Their Post-Translation Editing Strategies, thesis, Beijing Foreign Studies University.

[9] Liu X, Du Y, 2022, Research on Error Types in Chinese-English Machine Translation of Online Literary Works. Overseas English, 2022(9): 15–17.

[10] Yu Y, Nie B, 2019, Analysis of Errors in Online Machine Translation Based on Scientific and Technological Texts.

Materials for Information Recording, 20(5): 130–132.

[11] Ren J, Yu Y, 2023, Research on Translation Errors in Machine Translation of Petrochemical Texts: A Case Study of Youdao Translation from Chinese to English. Modern Commerce and Trade Industry, 44(5): 67–69.

[12] Tang B, Chen S, 2019, A Review of Medical Text Translation Using Online Machine Translation Software. Journal of Sci-Tech Translation, 33(3): 23–26 + 49.

[13] Fang F, 2018, Analysis of Errors in Chinese Translation of Business English Machine Translation, thesis, Guangdong University of Foreign Studies.

[14] Zou Q, 2015, Translation Error Analysis of Management Text Machine Translation, thesis, Lanzhou University.

[15] Guo W, Hu F, 2021, Study on Translation Evaluation and Post-Translation Editing of Neural Machine Translation. Journal of Beijing International Studies University, 43(5): 66–82.