Research on Computer-Assisted Translation of Abstracts of Railway Engineering Papers

Jun Wang[1,*] Chaowu Liu[1]

[1]East China Jiaotong University, Nanchang, Jiangxi Province, China

*Corresponding author

Keywords: Railway engineering, Abstracts traslation, Computer-assisted translation(cat), Pre- and post-editing

Abstract: The article analyzes the translation quality and stability of common online translation platforms and compares machine translation and manual translation. It summerizes the problems existing in machine translation and puts forward pre- and post-editing countermeasures to improve the quality of computer-assisted translation of Railway engineering papers, so as to help the Chinese Abstracttranslation based on CAT technology recognized by the international academy community.

1. Introduction

The development of China's high-speed railway has attracted worldwide attention, and academic exchanges related to railway engineering are becoming more frequent. The abstract of relevant academic papers has also attracted more attention because it can briefly explain the core content of the paper, so that international scholars can quickly master the content of the paper [1]. Whether the translation of the abstract of the paper is accurate, conforms to the English expression and thinking habits, etc., are all related to whether the paper can be included in an authoritative database or platform. The English translation of abstracts of Chinese academic papers has become an important means for Chinese academic papers to go abroad and move towards internationalization. Therefore, the rational use of computer-assisted translation is important to improve the translation quality of abstracts and make them better meet the collection standards of international retrieval institutions.

2. Preliminary Evaluation of Common Online Translation Platforms

The author and two railway engineering experts made a preliminary evaluation of the translations of six common online translation platforms on the percentage system. The translation score in Table 1 is the average of the four.

Table 1 Statistical Data Of Translation Scores of Online Platforms

Platform Original	1	2	3	4	5	6	7	Average	Standard deviation
CNKI	86.25	87.00	87.25	83.00	85.50	79.75	85.25	84.86	2.46
Bing	78.25	76.50	77.25	81.25	84.25	76.00	78.25	78.82	2.72
Baidu	86.25	85.75	87.25	86.00	85.75	83.25	85.75	85.71	1.12
Google	85.25	82.50	82.00	81.00	78.75	75.75	84.75	81.43	3.09
DeepL	83.75	77.75	79.25	79.25	77.50	74.75	82.75	79.29	2.88
Sogou	85.50	81.50	77.00	82.75	81.25	80.25	86.00	82.04	2.87

The accuracy of the original text can be preliminarily judged from the average translation score of each translation platform. As shown in Table 1, the average scores from high to low are Baidu, CNKI, Sogou, Google, DeepL and Bing. Therefore, for the summary text of railway engineering specialty, the translation accuracy of Baidu and CNKI is relatively the highest, while that of DeepL and Bing is the lowest. The standard deviation can reflect the stability of the translation quality of

the translation platform. From low to high, the standard deviations of each platform in Table 1 are Baidu, CNKI, Bing, Sogou, DeepL and Google. It shows that the translation quality stability of Baidu and CNKI is better, while the translation quality stability of Google and DeepL is relatively poor. Based on the comprehensive accuracy and stability results, it can be preliminarily judged that Baidu and CNKI rank in the top two in terms of translation quality and stability, so they are relatively more suitable as online translation platforms for abstract translation of railway engineering papers. Due to the limited number of texts, the number of evaluators and evaluation indicators, the evaluation results of this part are only for reference.

3. Error Analysis of Online Translation Platform for Abstracts of Railway Engineering Papers

3.1 Corpus Source

With the increasingly frequent academic exchanges between China and foreign countries in railway engineering, the number of railway engineering papers is increasing. In order to make the corpus selection more objective and accurate, the author searched the abstracts of all papers issued in 20 journals in the field of railway engineering, such as *China Railway and science* and *Journal of Railway Science and Engineering* from 2011 to 2021, With the help of relevant experts, 10 papers were selected according to the three principles of representativeness, authority and availability put forward by Nwogu[2], including applied research on railway transportation, facilities and equipment, transportation organization and safety, high-speed railway and heavy haul transportation, urban rail transit, etc. The English abstracts of these 10 papers are used as the artificial translation corpus (ATC) of this study. Through the online translation platform, the Chinese abstracts of the above papers are translated into English also as the machine translation corpus (MTC) of this study.

3.2 Analysis of Research Methods and Results

The translation error characteristics of specific machine translation output are obvious. Understanding and identifying the error characteristics can improve the efficiency and quality of post-translation editing [3]. According to the requirements of international well-known retrieval institutions for the writing of terms in English abstracts and the language characteristics of abstracts of railway engineering papers, the author evaluates and analyzes the quality of the two translations through manual translation practice and CAT practice, drawing the following conclusions: grammatical markers error, terminology errors, translation omission, sentence structure error, tense and voice errors, translation redundancy, word order error, etc.

3.2.1 Grammatical Markers Error

English uses more predicate verbs to express behavior events, so tense markers are rich and diverse, which is a typical embodiment of strong timeliness.[4] In contrast, although verbs are also used frequently in Chinese, grammatical markers are much less than those in English. Chinese grammatical markers are more reflected in some specific time words or quantifiers. There are many noun grammatical markers in English translation of abstracts of railway engineering papers. English countable nouns should have grammatical markers, which can indicate the class, general and special reference of nouns [5]. However, machine translation is not perfect enough, so it is easy to ignore the grammatical markers of the source language in the translation process, resulting in translation errors. Therefore, when using machine translation, the translator should check whether the translation is marked with correct grammar, and then make appropriate post-editing.

3.2.2 Terminology Error

Terminology translation standards should include accuracy, readability, transparency and conventionality[6]. Terminology translation in abstracts of scientific and technological papers must be accurate and avoid ambiguity, because readers who read abstracts of scientific and technological papers basically have relevant knowledge background. Inaccurate terminology translation will make them question the academic level of the whole article. On the contrary, if the terminology translation of the abstract is accurate enough, even if the fluency has some defects, professional

readers can also understand the main content of the translation with those keywords [7]. However, machine translation often translates some technical terminologies literally because of the lack of relevant professional knowledge, resulting in a great difference between the meaning of the translated terminologies and the source language. Therefore, terminology translation needs to be verified in many ways. Translators can also make full use of existing literature, electronics, network resources and agreed terminology translation standards to make their translation more accurate.

3.2.3 Sentence Structure Errors

English sentences attach importance to structure and grammar, such as attributive clauses and nominal clauses, which appear frequently in English [6], while Chinese syntactic structures attach importance to parataxis, pay attention to the unity of artistic conception and weaken the role of grammar. Therefore, translators should fully consider the different characteristics of the two languages and choose appropriate sentence structures. Due to the segmentation, recognition error or relationship judgment confusion of morphemes or lexical chunks in short sentences, machine translation is easy to lead to structural confusion, semantic errors or contradictions and logical confusion. Therefore, when using machine translation to translate such sentences, translators have to carry out appropriate post-editing to modify the grammatical and semantic errors of the translation, so as to improve the quality of the translation.

3.2.4 Tense Errors

The present tense, the past tense and the present perfect tense are in the majority in the translation of abstracts of railway engineering papers. The present tense is generally used to describe the purpose, method and conclusion of the current research; The general past tense describes past or previous research; The present perfect tense is used to link past research with current research, emphasizing the impact and role on the present or future [1]. Since the abstract is taken as a whole, the use of tenses should take into account the previous and subsequent texts and the specific context. Machine translation cannot choose the appropriate tenses according to different contexts, resulting in translation tenses errors. Therefore, when translators use machine-assisted translation, they should determine the context of the source text, so as to edit the translation according to the context.

3.2.5 Voice Errors

English voice includes active voice and passive voice. English has an obvious tendency of object name, so passive voice is widely used; The passive voice is rarely used in Chinese, because the passive voice often shows a mandatory meaning in Chinese, so the multi-use of this voice will make the expression of the article appear unnatural[6]. Railway engineering text emphasizes the objectivity of content and avoids subjective color. Passive voice has the objectivity of statement, which is widely used in the translation of railway engineering paper abstracts. In the process of Chinese-English translation of abstracts of railway engineering papers, machine translation is easy to translate the voice of the original sentence directly, with less voice conversion, which is easy to make the translation not meet the characteristics and requirements of railway engineering English translation. Therefore, the translator can properly edit the sentence patterns suitable for passive voice, so as to make the translation more in line with the translation standards of railway engineering translation.

3.2.6 Translation Redundancy

Railway engineering texts are oriented by transmitting relevant information of railway engineering, which are objective, rigorous and concise. Translation redundancy refers to excessive translation, which translates the content that does not need to be translated, resulting in the redundancy of the translation[7], making it difficult for the translation to meet the inclusion standards of international retrieval institutions. Furthermore machine translation is more mechanical and rigid in word selection. When dealing with common words or continuous synonyms in the text, the selected translation words are often repeated. Although this does not affect the

semantic expression of the translation, it reduces the readability of the translation and affects the natural and smooth expression of the translation. Therefore, when the machine cannot recognize semantic repetition, the translator should carry out post-editing to delete unnecessary repetition.

3.2.7 Word Order Error

English and Chinese have different logical thinking modes, there are differences in the order of some component structures, such as inverted word order and natural word order, differences in the position of modifiers, adjustment of focus, etc. From the logical level, due to the different ways of thinking in sentence construction between the two languages, Chinese tends to put the focus of expression in the front, while English tends to put it in the back. The specific differences are reflected in the different positions of attributives, adverbials and subjects. There are often too many juxtaposed prepositions in machine translation. Although the number of parallel premodifier is theoretically unlimited, more than 3,4 are not common, Therefore, in practice, the translator can convert some prepositional attributive into prepositional phrase structure or of structure to make the translation more in line with the target language expression.

4. Conclusion

For professional texts such as railway engineering paper abstracts, there are still many deficiencies in computer-assisted translation. Therefore, when selecting the online translation platform, we should scientifically evaluate the translation quality of each platform. By comparing and analyzing the manual translation and machine translation of English abstracts of railway engineering papers, this paper summarizes that machine translated English abstracts are prone to errors such as grammatical markers error, terminology errors, translation omission, sentence structure error, tense and voice errors, translation redundancy, word order error, etc. The main causes of errors include intensive terminology, differences in sentence structure and language thinking between English and Chinese, long and complex sentences, limited context recognition and analysis ability of machine translation platform. Therefore, the translator can carry out pre-editing including sentence component supplement, context determination of the original text, sentence structure transformation, and post-translation editing including term verification, grammatical and semantic error modification, voice conversion, and word order adjustment, so as to promote the translation of English abstracts by CAT technology to meet the inclusion requirements of international retrieval agencies.

References

[1] Yahui Wang. Research on translation of abstracts of academic papers on railway transportation. Journal of Lanzhou Jiaotong University, 2020,39 (06):146-150.

[2] Nwogu K N.The Medical Research Paper: Structure and Function. English for Specific Purposes,1997,16 (2):119-138.

[3] Qiliang Cui. On post editing of machine translation. China Translation, 2014 (6):68-73.

[4] Yamei Ou, Zhengguang Liu. The relevance model between English and Chinese timeliness and grammatical markers. Foreign Language Teaching and Research, 2021,53 (01): 16-28 + 159.

[5] Sheng Chen, Chuanmao Tian. Problems in Chinese-English Translation of Online Translation Platform and Post-translation Editing: Taking Petroleum Geological Documents as an Example. China Science and Technology Translation, 2021,34 (01): 31-34 + 49.

[6] Linglin Feng. English Chinese Language Differences in English Translation and Translation Countermeasures. Journal of Yanbian Institute of Education, 2021,35 (03): 16-18.

[7] Yiqun Sun. Error Analysis of Machine-aided Translation of Material Abstracts.China Science and Technology Translation, 2018,31 (03): 15-18.