

# Discussion on Machine Translation Error Types and Correction Methods

Xiaohong Yu

*Department of Foreign Languages, Eastern Liaoning University, Dandong, Liaoning, China*

**Abstract:** This study comprehensively examines the challenges encountered in the practical application of machine translation technology, focusing on the core error types such as semantic understanding bias, grammatical errors and poor choice of words. In order to solve these errors, this paper discusses many correction strategies, such as post-editing and manual audit, model optimization and data enhancement, interactive translation and personalized customization. Post-editing and manual review ensure accurate control of translation quality; Model optimization and data enhancement enhance the generalization ability and accuracy of the translation system through technical means. Interactive translation and personalized customization meet the diversified translation needs of users and enhance the user experience. These modification methods not only provide strong support for the wide application of machine translation technology, but also inject new vitality into the continuous development of natural language processing. In the future, with the continuous progress of technology, machine translation technology is expected to achieve more intelligent and personalized translation services, and build a more convenient bridge for cross-language communication.

**Keywords:** Machine Translation; Semantic Deviation; Grammar Errors; Vocabulary Selection; Amendment Method; Model Optimization

## 1. Introduction

Under the tide of globalization, cross-language communication has become an important bridge to connect different cultures and promote international cooperation. Machine translation technology, a key underpinning of this bridge, has made significant progress in

recent years, offering the possibility of fast and easy language conversion. However, despite the increasing maturity of machine translation technology, it still faces many challenges in practical application. The deviation of semantic understanding, grammatical errors and improper choice of words not only affect the accuracy of translation, but also limit the application of machine translation in some professional fields or high-demand scenarios. Therefore, it is of great significance to explore the main error types of machine translation and find effective correction methods for improving the overall level of machine translation technology and promoting its wide application. The purpose of this study is to systematically analyze the error types of machine translation technology and propose corresponding correction strategies, in order to provide reference for the future development of machine translation technology.

## 2. Current Status and Development Characteristics of Machine Translation Technology

### 2.1 Technology Integration and Innovation

In recent years, machine translation technology has ushered in an unprecedented leap, and its development history profoundly reflects the comprehensive innovation from basic algorithm to cutting-edge technology. Technological integration and innovation are undoubtedly the core driving force for the development of this field. At the algorithmic level, the introduction of deep learning technology has brought disruptive changes to machine translation. The continuous optimization of neural network architecture enables machine translation system to simulate the complex operation mechanism of human brain neurons, and realize deep mining and efficient processing of language features. This advance significantly improves the accuracy and naturalness of translation, bringing the

output of machine translation closer to human expression habits. At the same time, the vigorous development of natural language processing technology has also injected new vitality into machine translation. With the integration of advanced technologies such as semantic understanding, syntactic analysis and emotion recognition, machine translation systems can more accurately grasp the intention and emotional color of the original text, so as to generate a more appropriate and vivid translation. This intelligent processing method undoubtedly adds more humanistic color to machine translation. The advent of the era of big data has provided unprecedented data resources for machine translation. Massive language data provide rich materials for model training, which enables machine translation systems to dig out more language rules and knowledge. This not only enhances the generalization ability and adaptability of the system, but also provides the possibility for its application in more fields. Driven by technology integration and innovation, machine translation technology is gradually transforming into an intelligent language service system. It not only plays an irreplaceable role in high-end fields such as academic research and cultural exchanges, but also is widely used in many daily scenes such as business translation, educational guidance, and tourism navigation. The continuous progress of machine translation technology is building a more solid and convenient bridge for cross-cultural communication in human society[1].

## **2.2 Multi-Language Processing Ability**

In today's increasingly globalized world, the multilingual processing capability of machine translation technology has become a key bridge connecting different countries and cultures. With the rapid progress of technology, modern machine translation systems have been able to overcome language barriers and achieve seamless and accurate conversion between multiple languages. The multi-language processing capability of machine translation is first reflected in the diversity of language pairs it supports. Whether it is common English, Chinese, French, etc., or a relatively small number of languages, modern machine translation systems can provide high-quality translation services. This wide

language coverage makes it easier to exchange information on a global scale. More importantly, the machine translation system can accurately grasp and flexibly handle the characteristics of different languages. By introducing multi-language sharing encoders and decoders, the system can realize cross-language knowledge transfer and sharing, thus improving the efficiency and accuracy of translation. At the same time, according to the grammatical structure, vocabulary characteristics and expression of different languages, the machine translation system can also carry out personalized adjustment and optimization. This kind of meticulous processing ensures the accuracy and readability of the translation results, and makes the communication between different languages smoother. In addition, the continuous development of big data and artificial intelligence technology has also provided strong support for the multi-language processing capabilities of machine translation systems. By mining and utilizing the linguistic rules and knowledge in the massive multilingual data, the system can better understand the semantic information of the original text and generate a more natural and smooth translation. The improvement of this ability not only makes machine translation play a greater role in cross-cultural communication, but also injects new vitality into the diversified development of global culture. The multi-language processing capability of machine translation technology is constantly improving, providing a more convenient and efficient tool for cross-cultural communication in the context of globalization[2].

## **2.3 User Interaction and Feedback Mechanism**

In the continuous evolution of machine translation technology, user interaction and feedback mechanism has become increasingly important, and has become a key element to improve translation quality and optimize user experience[3]. The introduction of this mechanism not only allows users to directly participate in the translation process, but also realizes the dynamic adjustment and continuous improvement of the translation system through the closed-loop of real-time feedback and model optimization. At the user interaction level, modern machine translation

systems are committed to providing a rich variety of interface design and functional options to meet the diverse needs of users. Users can refine and optimize the machine-generated translation according to their own or specific field of expression habits, through highlighting, vocabulary replacement, sentence structure adjustment and other means. This flexible interaction not only improves the accuracy and personalization of the translation, but also enhances the user's sense of participation and satisfaction. At the same time, some advanced machine translation systems also introduce intelligent recommendation functions. Based on the user's editing history and preferences, the system can automatically recommend possible translation options, further simplifying user operations and improving translation efficiency. This intelligent interaction design not only improves the user experience, but also promotes the development of machine translation technology in a more intelligent and personalized direction. In terms of feedback mechanism, feedback data such as user satisfaction evaluation and error correction are regarded as valuable training resources to guide the further optimization of machine translation models. These feedback data not only help the model identify and correct common errors in translation, but also encourage the model to learn and adapt to more complex and subtle linguistic expressions and cultural differences. Through continuous iterative training, the translation quality of the machine translation system has been significantly improved, and the user satisfaction has also been improved. The introduction of user interaction and feedback mechanism not only promotes the continuous progress of machine translation technology, but also provides new possibilities for cooperation between humans and machines. In the future, with the continuous development of technology, the user interaction and feedback mechanism will be more perfect, which will bring broader development prospects for machine translation technology.

### **3. Main Error Types in Machine Translation**

#### **3.1 Semantic Deviation**

In the extensive exploration of machine

translation, semantic deviation is an invisible barrier that often hinders the accurate transmission of translation results, and becomes a key problem to be solved urgently[4]. Semantic bias, in short, refers to the machine translation system fails to accurately capture and faithfully convey the semantic essence of the source language during the conversion process from the source language to the target language, resulting in significant deviation between the translation result and the original text in meaning. The root causes of this deviation are complex and diverse, among which the polysemy of vocabulary, the subtlety of context and the difference of culture are the three main factors. At the lexical level, the words of a language often have multiple meanings, and their exact meaning often depends on the specific context. However, when the machine translation system deals with these polysemous words, it is often difficult to accurately judge their specific meaning in this context, which leads to translation bias. At the context level, the context of language is intricate, and the meaning of a word or phrase is often closely related to the context of the whole sentence. When analyzing this complex context, machine translation system is easy to fall into the dilemma of isolated analysis, ignoring the restricting effect of context on the meaning of words, resulting in the translation result deviating from the real intention of the original text. In addition, cultural difference is also an important cause of semantic deviation. There are significant differences in the ways and habits of language expression in different cultural backgrounds, which are often reflected in many aspects such as vocabulary choice, sentence structure and rhetorical devices. In the face of these cultural differences, machine translation systems are often difficult to fully understand and adapt, resulting in misunderstandings or ambiguities in the translation process, further exacerbating the problem of semantic bias. The existence of semantic bias not only seriously affects the accuracy of machine translation, but also limits its wide application in professional fields or high-demand scenarios. Therefore, how to effectively reduce and eliminate semantic bias has become an important issue in the field of machine translation.

### **3.2 Grammar Errors**

In the vast field of machine translation, grammatical errors are a challenge that cannot be ignored, which often troubles the fluency and accuracy of translation results[5]. As a common phenomenon in the process of machine translation, this kind of error is manifested in various forms, involving not only the confusion of sentence structure, but also the wrong application of basic grammatical elements such as tense, voice and part of speech. At the sentence structure level, it is often difficult for machine translation systems to accurately identify and effectively reconstruct complex sentence patterns in the source language, resulting in structural disorders such as incomplete sentence components, redundancy, or improper collocation of subject-verb-object in the translation results. These structural defects not only destroy the smoothness of the translation result, but also weaken the accuracy of its information transmission, making the reader feel confused when reading. In terms of tense and voice, the machine translation system is often unable to grasp the tense change and voice conversion of the original text, resulting in tense confusion and voice errors in the translation results. Such misuse of tense and voice not only reduces the readability of translation, but also misleads readers to understand the intention of the original text, resulting in the deviation of information transmission. In addition, part of speech error is also a grammatical problem that cannot be ignored in machine translation. Due to the complexity of the rules of parts-of-speech conversion between different languages, machine translation systems are prone to errors when dealing with such conversions, resulting in improper use of parts-of-speech in translation results. This kind of errors not only affect the whole grammatical structure of the sentence, but also damage the accurate expression of semantics, making the translation result appear stiff and unnatural. The existence of grammatical errors reveals the limitations and challenges of machine translation systems in grammar processing. In the pursuit of translation efficiency, we must strictly control the accuracy of grammar to improve the quality and readability of translation results.

### **3.3 Improper Choice of Vocabulary**

In the precise exploration of machine translation, the improper choice of words is like an elusive semantic fog, which often puzzles the accuracy and authenticity of the translation results. Vocabulary, as the cornerstone of language, not only determines whether the grammatical structure of the sentence is smooth, but also profoundly affects the accurate transmission of information and the delicate expression of emotions. In the process of machine translation, although the algorithm is constantly optimized and improved, due to the limited understanding of the complex relationship between the source language and the target language vocabulary, the problem of improper vocabulary selection often occurs in the translation results. This kind of improper choice may be reflected in the excessive literal translation of words, that is, the mechanical translation according to the literal meaning, while ignoring the cultural differences and expression habits between languages, making the translation result appear stiff, unnatural, and even ambiguous. It may also be reflected in the generalized use of words, that is, the failure to accurately capture the specific meaning and context of words in the original text, resulting in misunderstanding of information or loss of key information, making the translation result unable to faithfully convey the intention of the original text. To make matters more complicated, poor choice of words may also be accompanied by confusion about the hierarchy of words. The same word may have multiple meanings and usages in different contexts, and the hierarchy and nuance are often elusive. In the face of this complex situation, the machine translation system is often difficult to accurately judge the appropriate level of words in this specific context, so as to make wrong choices, resulting in the translation results compared with the original text deviation. In addition, lexical collocation errors are also one of the common problems of improper lexical selection in machine translation. The collocation rules of different languages are very different, and what is reasonable in one language may not be applicable in another language. When machine translation system deals with such complicated lexical collocation, it is easy to make mistakes due to the lack of sufficient context understanding and language perception ability, which leads to the awkward situation of

improper lexical collocation in the translation result, which affects the overall fluency and readability. Improper lexical selection reveals the complexity and challenge of lexical processing in machine translation systems. While using machine translation technology, we should maintain a cautious and critical attitude, combine manual proofreading and language expertise, and jointly promote the continuous progress and improvement of machine translation technology, in order to achieve more accurate and authentic translation results.

#### **4. Correction Methods for Machine Translation Errors**

##### **4.1 Post Editing and Manual Review**

In the context of the wide application of machine translation technology, post-editing and manual review have become the two core links to ensure the accuracy of translation. Post-editing plays an important role in the machine translation process. It is not a simple proofreading and modification of the machine translation results, but a process of deep digging and reshaping the quality of translation. In the post-editing process, professional translators will conduct a comprehensive and detailed review of the machine translation results based on the grammar rules, vocabulary collocation principles and context understanding of the source and target languages. With their keen language perception and deep translation experience, they can accurately catch and correct the possible problems such as grammatical errors and improper word choice in machine translation. At the same time, on the basis of maintaining the meaning of the original text, they will also carry out the necessary polishing and optimization of the translation result, so that it is more suitable for the expression habits and cultural background of the target language, so as to improve the overall quality of the translation. At the same time, manual review, as a supplement and sublimation of the post-editing process, further strengthens the accuracy and reliability of the translation results. Manual reviews are usually conducted by experts with deep language skills, extensive translation experience, and keen cultural insight. They will conduct a final review and evaluation of the translation results after

editing, taking into account the accuracy, fluency, professionalism and cultural adaptability of the translation. In this process, experts will use their professional knowledge and experience to fine-polish and improve the translation results to ensure that the translation results are not only faithful to the original text, but also meet the reading expectations of the target language readers. Through the dual guarantee of post-editing and manual review, machine translation technology can continuously improve its translation quality and accuracy while maintaining efficiency. This revised method not only provides a solid support for the wide application of machine translation technology, but also injects new vitality into the continuous development and progress of the translation industry.

##### **4.2 Model Optimization and Data Enhancement**

In the continuous evolution of machine translation technology, model optimization and data enhancement have become the key means to improve translation quality and reduce translation errors. These two aspects complement each other and jointly promote the development of machine translation technology to a higher level. Model optimization aims to improve the translation capability and accuracy by adjusting and optimizing the structure, parameters and algorithms of the translation model. This includes the introduction of more advanced neural network architectures such as Transformer to capture more complex language features and semantic relationships; More refined parameter tuning strategies, such as learning rate adjustment and regularization methods, are used to improve the generalization ability and stability of the model. And develop more efficient training algorithms, such as distributed training, mixed precision training, etc., to speed up model training and reduce computing costs. Through model optimization, the machine translation system can understand the source language text more accurately and generate a more natural and smooth target language text. Data enhancement improves the performance of the translation model by increasing the diversity, richness, and scale of the training data. This includes collecting and integrating text data from different domains, different styles, and

different languages to extend the training scope of the model; Using data enhancement techniques, such as synonym substitution, sentence recombination, noise addition, etc., to increase the diversity of data; As well as building high quality manually annotated datasets to provide accurate translation references. Through data enhancement, machine translation systems can learn richer language knowledge and translation rules, thereby improving the accuracy and fluency of translation. Model optimization and data enhancement play a crucial role in machine translation. They can not only help machine translation systems reduce translation errors and improve translation quality, but also enhance the robustness and adaptability of the system, making it better able to cope with a variety of complex translation scenarios. At the same time, these two aspects are mutually reinforcing: better models can make more efficient use of data, and richer data can further improve the performance of models. In the future, with the continuous progress of technology and the deepening of application, model optimization and data enhancement will play a more important role in the field of machine translation. We will continue to explore more efficient and intelligent model optimization methods and data enhancement techniques to drive the continuous advancement and development of machine translation technology.

#### **4.3 Interactive Translation and Personalized Customization**

With the rapid development of machine translation, interactive translation and personalized customization have gradually become an important means to improve translation quality and user experience. Combined, the two not only inject new vitality into machine translation technology, but also provide users with a more flexible and efficient translation solution. Interactive translation is an innovative translation model that allows users to make real-time interventions and adjustments in the process of machine translation. In the interactive translation environment, users can instantly modify and optimize the results of machine translation by simply clicking, dragging or typing. This model breaks the passivity of traditional machine translation and enables users to

actively participate in the translation process according to their needs and preferences, thus ensuring the accuracy and fluency of the translation results. Interactive translation not only improves the efficiency of translation, but also enhances the user's sense of participation and satisfaction. Personalized customization is the process of making customized adjustments to the machine translation system for specific users or application scenarios. Through in-depth understanding of users' translation needs, language habits and professional domain knowledge, personalized customization can provide users with more suitable translation services to their actual needs. For example, for translation in professional fields such as law and medicine, personalized customization can ensure the professionalism and accuracy of translation results. For the translation of tourism, entertainment and other scenes, personalized customization can pay more attention to the vividness and interest of language. Personalized customization not only improves the quality of translation, but also enhances the adaptability and flexibility of machine translation systems. The combination of interactive translation and personalized customization opens up a new way for the development of machine translation technology. Through interactive translation, users can adjust and optimize the translation results in real time to ensure the accuracy and fluency of the translation. Personalized customization can meet the specific needs of different users or application scenarios, and improve the quality and efficiency of translation services. This combination not only improves the intelligence level of machine translation technology, but also provides users with a more convenient and efficient translation experience. In the future, with the continuous progress of technology and the changing needs of users, interactive translation and personalized customization will play a more important role in the field of machine translation.

#### **5. Conclusions**

This study reveals the current challenges and problems of machine translation through an in-depth analysis of the main error types of machine translation technology. In response to these challenges, this study proposes a variety of correction methods, including post-editing

and manual review, model optimization and data enhancement, interactive translation and personalized customization, to improve the accuracy and applicability of machine translation. These modification methods not only provide a strong support for the wide application of machine translation, but also inject new vitality into the continuous development of natural language processing. In the future, with the continuous progress of technology and changes in user needs, machine translation technology will continue to develop in a more intelligent and personalized direction. Through continuous technological innovation and optimization, machine translation will contribute more to the realization of barrier-free cross-language communication.

#### References

[1] Chen Jiale. Analysis of errors in Chinese-English machine translation of business contract text [J]. Overseas

English,2023(17):10-12.

- [2] Qiu Bailian, Wang Mingwen, Li Maoxi, et al. Construction and thinking of a Fine-grained English-Chinese Machine Translation Error Analysis Corpus [J]. Journal of Chinese Information Technology, 2002,36(1):47-55.
- [3] Mo Ruyu, Jiang Nan. Error analysis and post-translation editing strategies in machine translation of scientific and technological texts [J]. Modern Linguistics, 2019,12(10):7.
- [4] Lu Chen, Luo Guihua. Comparative Analysis and Statistics of machine translation error types in prose [J]. Journal of Luoyang Normal University,2022,41(12):73-78.
- [5] QIU Bailian. Review on Machine Translation error Analysis [J]. Journal of Hubei Second Normal University,2023,40(11):102-108.