Development of post-editing skills in machine translation

Qianqian SHI

Xi'an University of Posts &Telecommunications, Xi'an 710121, China

Abstract: As a necessary complement to machine translation, post-editing plays an increasingly vital role in the language service industry. It represents a crucial pathway for enhancing machine translation quality and serves as a significant hallmark of human-machine interactive translation. With the advancement of machine translation technology and its expanding application scope, the cultivation of post-editing skills has become a crucial component of translation education. Through well-designed training strategies, learners can enhance their post-editing abilities to meet the translation market demands of the machine translation era.

Key words: post-editing; machine translation; editing skills

1 Introduction

With the continuous development and application of machine translation (MT) technology, the post-editing of machine translation (MTPE) has become an indispensable component of the language service industry and an essential aspect of translation education and training. However, MTPE is not simply about correcting the output of MT; it requires specific abilities and skills, including service capabilities, risk assessment abilities, tool proficiency, language proficiency, translation skills, and editing skills. Effectively cultivating MTPE capabilities is a significant and pressing issue facing us today. This article aims to explore the theory and practice of MTPE skill development, providing insights and references for MTPE education and training.

2 Overview of machine translation

2.1 Current state of machine translation

Machine translation is a vital branch of artificial intelligence and an applied field of computational linguistics. It represents a form of translation activity that is "fully automated" and requires no human intervention during the translation process. MT trainers need to design a series of structured languages internally to code the text to be translated and then output the translation intuitively. The development of machine translation has gone through several stages, from the initial rule-based methods to statistical methods, and more recently, neural network-based methods. The goal of machine translation is to provide high-quality, efficient, and intelligent language services to meet the growing demand for cross-lingual communication [1].

Currently, neural network-based machine translation has become the mainstream and cutting-edge approach in the field of machine translation, achieving remarkable performance on various public test sets. However, machine translation also faces new challenges and opportunities, such as addressing issues related to low resources, multimodality, multiple

domains, and multitasking. Additionally, improving the interpretability, reliability, and interactivity of machine translation poses significant issues. In the future, machine translation will continue to intersect with fields like artificial intelligence, natural language processing, and computational linguistics to explore more efficient, intelligent, and user-friendly language service solutions [2].

2.2 Applications of machine translation

Machine translation finds extensive applications across various domains and contexts. For instance, in the field of education, machine translation can assist learners in acquiring and comprehending knowledge and information in different languages. It can also help learners improve their language skills by practicing listening, speaking, and writing. In the realm of tourism, machine translation is popular among travel enthusiasts as it facilitates convenient and enjoyable travel experiences by enabling communication with locals and providing insights into local culture and customs [3]. In the domain of business and trade, machine translation can effectively assist communication between business people from different countries, eliminating language barriers that could hinder commercial activities. It aids business professionals in conducting efficient and successful global market activities, such as market research and analysis, devising and executing marketing strategies, negotiation, and contract signing with clients and partners [4].

In the above contexts, machine translation is performed purely by machines and does not involve post-editing. However, in many real-life situations, pure machine translation may not meet user requirements due to limitations in machine capabilities. The inflexibility of machine translation underlines the importance of the post-editing process. For instance, in political and diplomatic contexts, as globalization progresses, countries need to engage in political and diplomatic communication with others. The advent of machine translation can enhance the efficiency of interlingual communication and empower politicians to take influential actions on the international stage, such as delivering speeches, issuing statements, participating in international organizations, conferences, agreements, dialogues, and collaborations with other countries and regions. However, challenges can arise during machine translation, and errors in translation may have implications for a country's image in these formal contexts. Therefore, there is considerable space for improvement in the translated text identified through machine translation [5].

3 The importance of post-editing of machine translation

Machine translation (MT) can only undertake preliminary processing during the translation process, and pure machine translation often contains issues that render the translated content unusable. To ensure translation accuracy and high quality, an essential step is to apply post-editing to the output. Post-editing of machine translation (MTPE) involves modifying and refining the text generated by machine translation (MT) to enhance its quality and adaptability. MTPE can be categorized into two types: rapid post-editing and comprehensive post-editing. Rapid post-editing addresses serious and obvious errors while allowing minor issues to remain, suitable for internal use or as a reference. Comprehensive post-editing, on the other hand, corrects all errors to bring the translation closer to the quality of human translation, suitable for publication or public use. The importance of MTPE in the language service industry can be elucidated in several ways.

Firstly, MTPE enhances the efficiency and scale of language services. With the development of globalization and informatization, the demand for language services is growing rapidly, but the cost and time constraints of human translation limit its capacity and scope [6]. MTPE can leverage the speed and breadth of MT to quickly generate a large volume of initial translations, which can then be refined and optimized through human intervention to meet customers' needs and adhere to target language standards. Thus, MTPE saves time and manpower, improving the efficiency and scale of language services.

Secondly, MTPE improves the quality and flexibility of language services. Due to the imperfections and instability of

MT systems, MT output often contains errors and deficiencies in aspects like spelling, grammar, vocabulary, style, and cultural nuances. MTPE can eliminate or reduce these issues through manual review and correction, enhancing translation accuracy and fluency. Moreover, MTPE allows for adjustments and improvements to the MT output based on factors such as different clients, contexts, and purposes, making it more suitable for specific language service requirements and standards. Thus, MTPE enhances the quality and flexibility of language services [7].

Lastly, MTPE fosters innovation and development in language services. MTPE is not just a means of providing language services but also an opportunity for research and learning in the field. Through MTPE, translators can learn about and master different MT systems and technologies, identify and address issues and challenges in MT output, and propose and implement methods and strategies to improve its output. As a result, MTPE promotes the enhancement of professional skills and competencies among translators and contributes to the optimization and advancement of MT systems and technologies. Additionally, MTPE can provide data and insights to other related fields such as machine learning, natural language processing, and computational linguistics, thereby driving innovation and development in language services.

In conclusion, MTPE plays a crucial and positive role in the language service industry and is a language service model worthy of attention and promotion. To enhance the effectiveness and efficiency of post-editing, this paper will analyze common errors that occur during machine translation and propose relevant strategies for improving post-editing skills. Through well-designed training strategies, learners can enhance their post-editing abilities to meet the translation market demands of the machine translation era.

4 Analysis of post-editing error types

4.1 Addition or omission errors

These errors refer to the presence of extra or missing words or phrases in the machine translation output, requiring deletion or addition during post-editing. Translation between different languages is influenced by various factors such as grammar, vocabulary, style, and culture, which can lead to different expressions or the inability to directly correlate some words or phrases between languages. Additionally, varying translation requirements and goals may diverge from the default settings or training data of machine translation systems, resulting in the unnecessary addition or omission of words or phrases. For example, some specialized terms, names, or locations may need to be preserved verbatim in certain contexts but require appropriate translation or transcription in others. Additionally, some function words, articles, or particles may be mandatory in one language but optional in another.

4.2 Word order errors

Word order errors are common in machine translation and result from differences in sentence structures between languages. These errors occur when the word order or structure of words or phrases in the machine translation output does not conform to the conventions or norms of the target language. Such errors can affect the fluency and readability of the translation, reducing its quality and comprehensibility. Due to the complexity and variation in languages, different languages have distinct grammar, vocabulary, style, and cultural features, leading to different expressions or orders for words or phrases. For example, English generally follows the Subject-Verb-Object (SVO) sentence structure, while Chinese can have various variations like Subject-Object-Verb (SOV), Subject-Verb-Object (SVO), or Verb-Subject-Object (VSO) based on context and emphasis. If the machine translation system cannot handle these differences and complexities accurately, word order errors may occur. Additionally, the imperfections and instability of machine translation systems may lead to errors as they struggle to capture and process the correspondences between source and target languages accurately. For example, English generally uses adjectives to modify nouns, while Chinese often uses nouns to modify nouns. If the machine translation system cannot handle such modification relationships properly, word order errors may

occur.

4.3 Untranslated errors

Untranslated errors refer to source language words or phrases in the machine translation output that have not undergone any translation and require to be translated during post-editing. Due to the internal mechanisms and external data limitations of machine translation, the system may not accurately recognize and process some specific or complex words or phrases in the source language, resulting in them being directly copied into the target language without translation. Language differences and complexities can also lead to situations where some words or phrases do not have direct or unique counterparts in different languages or require flexible conversion based on context or purpose. If the machine translation system cannot handle these differences and complexities properly, some words or phrases may be ignored or retained without appropriate translation. For example, idiomatic expressions, proverbs, puns, or some common nouns may remain untranslated. Additionally, proper nouns, abbreviations, or internet slang may also remain untranslated.

4.4 Named entity translation errors

Named entity translation errors refer to incorrect or inconsistent translations of named entities such as personal names, place names, or organization names in the machine translation output, requiring correction or standardization during post-editing. Machine translation systems may not accurately identify and process named entities in the source language, leading to incorrect translations or omissions. Some rare or newly emerged named entities may not be included in the machine translation system's dictionaries or corpora or may be mistakenly treated as ordinary words or phrases. Additionally, some personal names or place names may need to be retained unchanged in certain languages but require appropriate translation or transcription in others. If the machine translation system cannot accurately recognize the need for translation of these terms, it may result in inappropriate or inconsistent translations of named entities. To address these issues, optimization and updates to machine translation systems should be implemented to improve their ability to learn and process correspondences between source and target languages, as well as to enhance their adaptability and flexibility in different domains and styles, thus addressing these translation errors.

5 Strategies for developing post-editing skills

5.1 Participate in professional post-editing training courses

Post-editing skills are crucial in the era of machine translation and require continuous learning and improvement. One effective way to enhance these skills is by enrolling in professional post-editing training courses. These courses cover both the theory and practice of post-editing, including its definition, standards, processes, tools, and techniques. Participants learn how to formulate content strategies, select appropriate machine translation systems, and choose post-editing types based on various translation scenarios and requirements, such as rapid post-editing or comprehensive post-editing. Additionally, they become familiar with different levels of quality assessment standards, such as error analysis or functional analysis methods. This knowledge enables learners to adapt flexibly to diverse translation projects and client demands, ultimately improving translation efficiency and client satisfaction. Participants also learn how to address specific issues that may arise in machine translation, such as gender bias, cultural differences, and linguistic variations. Furthermore, they gain insights into cutting-edge machine translation technologies and trends, including neural machine translation and language artificial intelligence, which enhance learners' sensitivity to machine translation and their ability to adapt to the rapid evolution and updates in machine translation technology. Formal training courses significantly reduce the time and effort required for self-study, enabling learners to acquire professional post-editing skills more quickly.

5.2 Establish a post-editing style

Defining a post-editing style is another key aspect of improving post-editing skills. While machine-generated

language may have a consistent style, the writing style of different authors and translators can vary significantly. Therefore, during translation, it is important to ensure the accuracy of the translated content while also preserving the linguistic style of the original author. After understanding the objectives and requirements of different types of post-editing, select the appropriate post-editing style based on various translation scenarios and needs. Styles may include rapid post-editing or comprehensive post-editing, each adhering to corresponding quality assessment standards, such as error analysis or functional analysis methods. The goal of rapid post-editing is to ensure translation accuracy and comprehensibility while minimizing changes to style and personal preferences. Comprehensive post-editing aims to bring the translation closer to the quality of human translation by correcting all errors and ensuring grammatical correctness, appropriate style, and terminological consistency. Machine translation offers advantages such as rapid processing of large volumes of text, maintaining consistency and accuracy, and handling complex or specialized terminology. However, it has limitations, including difficulty in handling ambiguity, metaphors, cultural differences, linguistic variations, and more. When performing post-editing, it is essential to trust machine translation in its areas of strength and address potential issues that may arise in areas where it is less proficient.

5.3 Utilize effective tools and resources

While post-editing involves human refinement of machine-generated translations, it is a substantial task. Therefore, optimizing the use of relevant tools and resources can significantly enhance the capabilities of post-editors. Depending on the training content and methods, select suitable tools and resources, such as computer-assisted translation software, machine translation platforms, quality assessment tools, online corpora, and more. Computer-assisted translation software combines machine translation and human translation, facilitating a comparison, review, and modification of machine-generated translations. It also provides helpful features such as terminology management, translation memory, and quality checks. Machine translation platforms offer different types and qualities of machine translation services, allowing learners to explore various machine translation systems' characteristics and performance, enabling them to choose the most suitable system for testing or use. Quality assessment tools assist learners in objectively and systematically evaluating machine translation or post-editing results, providing statistics and reports. Online corpora provide learners with extensive bilingual or multilingual text resources for reference and inspiration. These tools significantly enhance the efficiency and quality of post-editing, saving time and costs and improving work productivity.

5.4 Cultivate self-learning and reflective skills

In addition to guidance and assistance from instructors, learners should develop self-learning and reflective skills to adapt to the rapid changes and continuous updates in machine translation technology. Learners should proactively stay informed about the latest developments and trends in machine translation and explore new machine translation systems or tools. Through self-guided learning and practice, utilize online resources and platforms, such as online translation tools, post-editing assessment systems, and post-editing communities, to continually experiment with and improve post-editing skills. This approach allows learners to adapt to different machine translation systems and text types, develop their judgment and risk awareness, and stay up-to-date with the evolving landscape of machine translation technology. Learners should also regularly reflect on their post-editing processes and results, and identify their strengths and areas for improvement. By creating improvement plans and seeking feedback and advice from others when needed, they can refine their post-editing skills. Collaborating and sharing experiences and insights with fellow post-editors or experts can also lead to mutual evaluation and feedback.

6 Conclusion

With the advancement of machine translation technology and its expanding application scope, the cultivation of post-

editing skills has become a pivotal component of translation education. This article has explored strategies for developing post-editing skills, including the determination of training objectives and standards, the design of training content and methods, the effective utilization of tools and resources, and the fostering of self-learning and reflective abilities. It is emphasized in this article that post-editing skills encompass the comprehensive ability to effectively perform post-editing tasks, including linguistic proficiency, translation competence, machine translation knowledge, quality assessment proficiency, tool utilization skills, and project management capabilities.

Acknowledgments

Conflicts of interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Cui QL. 2014. On post-editing in machine translation. *Chinese Translators Journal*, 6: 68-73.

[2] Cui QL, Li W. 2015. Research on types of post-editing errors--based on English-Chinese machine translation of scientific and technological texts. *Chinese Science & Technology Translators Journal*, 4:19-22.

[3] Feng QG, Zhang HY. 2015. Research on the cultivation of post-editors in the context of the global language service industry. *Foreign Language World*, 1: 65-72.

[4] Feng QG, Cui QL. 2016. Post-editing research: focusing and development trends. *Shanghai Journal of Translators*, 6: 67-74+89.

[5] He Y. 2018. The application of machine translation and post-editing strategies for college students in the internet age. *Journal of Shenyang Agricultural University (Social Sciences Edition)*, 20(6): 756-760.

[6] Luo JM, Li M. 2012. Analysis of errors in machine translation output. *Chinese Translators Journal*, 5.

[7] Li M, Zhu XM. 2013. Exploring the automation of post-editing in English-Chinese machine translation. *Chinese Translators Journal*, 4:83-87.