Journal of Social Science and Humanities ISSN: 1811-1564
DOI: 10.53469/jssh.2024.6(08).18

Analysis of Machine Translation Post-editing of IT Texts under the Guidance of TAUS Guidelines

Haiwei Liao

School of Foreign Languages, Jiangxi University of Technology, Nanchang, Jiangxi, China

Abstract: ***Machine Translation Post-Editing (MTPE) is essential for enhancing the accuracy and proficiency of machine-generated translations across diverse sectors, such as scientific, technical, and translation services industries. The Translation Automation User Society (TAUS) standards are used in this study to investigate the applicability of MTPE to IT texts. Even though neural machine translation (NMT) has made great strides and is now able to provide translations that are more accurate, these results frequently still need to be refined by humans in order to satisfy professional and industry standards. This thesis explains how TAUS principles, which address common problems such as terminology accuracy, fluency, and general readability, can maximize MTPE.***

Keywords: MTPE, IT Texts, TAUS, Human Translation Quality**.**

1. Introduction

In the era of globalization, the demand for digital and intelligent translation services has increased exponentially. Furthermore, an increasing number of translation projects have embraced the use of machine translation in conjunction with post-editing, highlighting the significance of this technique for improving translation accuracy and efficiency. This is particularly true when it comes to translating IT texts, legal documents, instruction manuals, industrial and agricultural texts, medical information, industrial patents, brochures, and other texts. For information to be communicated across language and cultural barriers, IT texts must be translated accurately and swiftly. Even though traditional human translation is quite accurate, it can be costly and time-consuming. Outputs and quality, the two translation cores, cannot be satisfied by using only human translators. Machine translation (MT), therefore, has become more well-known as a quicker and more affordable substitute. Through machine translation, we must assist translators in achieving both productivity and quality. Despite its advantages, MT frequently produces translations that are too poor quality for use in academic and professional settings. This discrepancy has led to the development of the Machine Translation Post-Editing (MTPE) technique, which combines the precision and expertise of human editing with the effectiveness of machine translation [1].

The Translation Automation User Society (TAUS) has been a driving force behind the advancement of MT and MTPE standard practices. Comprehensive instructions are offered by TAUS with the aim of improving the effectiveness and level of post-editing procedures. These guidelines provide organized methods for locating and fixing typical mistakes in texts that have been machine translated, enhancing the overall quality of translations. The objective of this research is to investigate the efficacy of TAUS recommendations in augmenting the quality and productivity of MTPE for IT texts, consequently advancing the wider objective of bolstering worldwide scientific communication.

2. Machine Translation and Post-editing

2.1 Machine Translation

The process of automatically translating text from one language to another with computer programs is known as machine translation, or MT. There have been three phases to MT's growth thus far. The first phase began in the 1950s, when rule-based and grammar translation models were the main focus of early research. The early MT systems, like IBM's Georgetown project, used dictionaries and intricate rules to translate text. However, it didn't take long for this Rule-based Machine Translation (RMT) to show its limitations in handling complex linguistic contexts and structures.

With advances in computational power and linguistic study, MT evolved into a statistical model-based era. Statistical Machine Translation (SMT) gained popularity beginning in the 1990s. After being trained on large bilingual corpora, SMT systems generated translations by applying statistical methods. However, SMT was not without its limitations, particularly when it comes to analyzing long sentences and unusual vocabulary.

Neural Machine Translation (NMT) is a new phase of Machine Translation (MT) that emerged in the 21st century with the advent of sophisticated deep learning technologies. In order to better capture the contextual and semantic aspects of language, neural network models—specifically, recurrent neural networks (RNNs) and convolutional neural networks (CNNs)—are used in neural machine translation (NMT). Recent advances in quality and fluency of machine translation can be attributed to models based on the Transformer architecture, such as Google's BERT and OpenAI's GPT.

2.2 Post-editing

Post-editing (PE) is the process by which human translators improve the quality of a translation by making changes to the initial drafts generated by MT. As MT technology continues to progress, particularly with the widespread use of NMT, PE has progressively grown in importance within the translation sector [2]. In MTPE, the output generated by an MT system is reviewed and corrected by a human translator. Through error correction, improved fluency, and verification that the translation satisfies the necessary requirements for accuracy and readability, this technique tackles the drawbacks of machine translation.

PE is divided into two categories by the worldwide MTPE standard ISO 18587:2017: Light Post-editing (LPE) and Full Post-editing (FPE) [3]. Early PE is mostly concerned with LPE. Resolving only the most significant errors that affect understanding is required. This phase was about increasing translation speed and efficiency, not about striving for perfection.

As MT quality increased, PE progressed into the FPE stage. FPE entails thorough optimization of the translation's style, consistency, and fluency in addition to error correction. Currently, the aim of PE is to reach a level of quality that is on par with human translation [4].

3. TAUS MT Post-editing Guidelines

The Translation Automation User Society (TAUS), a language data network and think tank, serves as a hub for resources for the international language and translation sectors. TAUS, which was established in 2004, offers the translation industry data, insights, analytics, benchmarking, and resources via its Academy, Data Cloud, and Quality Dashboard.

Even though PE is divided into LPE and FPE, nobody will simply pay for LPE, hence FPE is essentially what is done in some certain practices. Thus, in order to give translators a methodical approach for PE and enable them to more accurately detect and fix machine translation errors, TAUS released the "MT Post-Editing Guidelines" in 2016. This greatly improved translation efficiency and quality. Additionally, the book describes how to identify the appropriate level of PE and how the final quality criteria of the target readers or translation buyers are correlated with the beginning quality of the MT in order to measure the quality of PE.

According to the guidelines, there are two basic criteria by which post-editing quality is judged: 1) The quality of the MT raw output; 2) The expected end quality of the content [5]. Because of these two factors, the execution of the PE task is contingent upon both the MT's competence and the translation buyer's requirements for quality. Consequently, "Good enough" quality and "Human translation quality" are the categories into which TAUS divides the anticipated final PE quality. The detailed guidelines are as follows:

"Good enough" quality:

(a) Aim for semantically correct translation.
(b) Ensure that no information has been accidentally added or omitted.
(c) Edit any offensive, inappropriate or culturally unacceptable content.
(d) Use as much of the raw MT output as possible.
(e) Basic rules regarding spelling apply.
(f) No need to implement corrections that are of a stylistic nature only.
(g) No need to restructure sentences solely to improve the natural flow of the text.

Human translation quality:

(a) Aim for grammatically, syntactically and semantically correct translation.
(b) Ensure that key terminology is correctly translated and that untranslated terms belong to the client's list of "Do Not Translate" terms.
(c) Ensure that no information has been accidentally added or omitted.
(d) Edit any offensive, inappropriate or culturally unacceptable content.
(e) Use as much of the raw MT output as possible.
(f) Basic rules regarding spelling, punctuation and hyphenation apply.
(g) Ensure that formatting is correct.

4. Analysis of Post-editing Practices in IT Texts Based on TAUS Guidelines

4.1 Basic Features of the Texts

IT texts include special vocabulary, intricate sentence structures, and require clear and accurate communication, which makes them particularly difficult to translate. It is rich in terminology, is clear and rationally ordered, objective and accurate, succinct, standardized, useful, and instructional. Misunderstandings, incorrect information spreading, and miscommunication can all be caused by translation errors in IT texts. This is why precise translation is essential in this industry. When machine-translated IT materials are post-edited, TAUS rules offer an appropriate approach to addressing these problems. By following these guidelines, post-editors can ensure that translated materials are accurate, comprehensible, and suitable for the intended usage.

Based on the author's daily translation work, the Installation Guide Manual of the PowerChute Business Edition v9.5 from Schneider Electric IT Corporation was selected to practice the "MT+ PE" model. Data communication professionals are the targeted readers since they place a premium on correctness in technical details that ensure proper operation of control systems. The MTPE model is also better suited for the product installation instructions as an IT text type. In order to ensure that the output meets the requirement of "Human Translation Quality", we choose to handle the original text utilizing the effective and time-saving LPE mode after machine translation. This translation process's MT is done using DeepL. About 9,000 words make up the Guide, which was published in March 2018. This content is intended to be used as a tool for evaluating the design and construction of data centers.

Undoubtedly, MT offers benefits in certain IT text processing scenarios, but it is not error-free. Errors such as terminology translation errors, repetitions, omissions, additions, mistranslations, logical comprehension errors, etc. were frequently discovered in IT texts following machine translation once NMT became popular [6].

4.2 Errors in Machine Translation

4.2.1 Terminology Translation and Semantic Errors

Example 1:

ST: A compromised or malicious website may run "exploits"

that automatically trigger the download of malware onto your computer.
MT: 被入侵或恶意的网站可能会运行“运用”, 自动触发恶意软件下载到你的电脑上.
PE: 被入侵网站或恶意网站可能会运行“漏洞利用”, 自动触发将恶意软件下载到您的计算机.

The word “exploits” in the original text is translated as “运用” which is the common meaning of the word. However, in the IT industry, “exploit” is a terminology used in the computer field, and according to the item (b) of “Human Translation Quality” requirement of the TAUS MTPE guidelines, the PE should be carried out by changing the MT of the word to the customary or conventional term in the IT industry. When PE is performed, the word should be translated into the customed or conventional terminology of the IT industry, and the terminology should be translated consistently throughout the text. Therefore, it should be translated as “漏洞利用”.

Example 2:

ST: An Advanced Persistent Threat, or APT, is an attacker that gains access to your network and remains undetected for an extended period of time… Bots allow an attacker to remotely control your computer and make it do something nefarious.
MT: 高级持续性威胁（APT）是指能够进入您的网络并在较长时间内不被发现的攻击者. ……机器人允许攻击者远程控制你的电脑, 让它做一些邪恶的事情.
PE: 高级持续性威胁或 APT 是一种攻击程序, 可访问您的网络, 且长期不被察觉. ……僵尸病毒让攻击程序能够远程控制您的计算机, 并使其做一些违法事情.

After a quick review at the output of the MT, we can discover that there are errors in the translation of “attacker”, “Bots” and “nefarious” in the original text, resulting to a deviation in the meaning of the sentence. In order to fulfill the requirement of “semantically correct translation” in item (a) and the requirement of item (b), all the translation of three words need to be edited. “Bots” is literally translated as “机器人”, but in the IT cybersecurity field, “Bots”, which is a technical term, usually refers to the viruses or the programs in botnet. Botnet means “僵尸网络” in Chinese. Meanwhile, MT defines “attacker” as a human being, but in this context, “APT” refers to a type of cyberattack executed by a highly specialized human attacker. The original meaning of “nefarious” is wicked and immoral, but it also has the meaning of criminal, and the purpose of “Bots” in this sentence is to steal sensitive data or damage systems, which is a criminal act. Therefore, when doing PE, the three terms would be more accurate to be translated into “攻击程序”, “僵尸病毒” and “违法”.

4.2.2 Syntactical Errors

Example 3:

ST: PowerChute Business Edition version 9.5+ uses SNMP to integrate with your chosen Network Management System (NMS) for the discovery, monitoring and control previously provided by the PowerChute Business Edition console and server.
MT: PowerChute Business Edition 9.5+版本使用SNMP与您选择的网络管理系统 (NMS) 集成, 以便发现、监视和控制以前由 PowerChute Business Edition 控制台和服务器提供的功能.
PE: PowerChute Business Edition 版本 9.5+ 采用 SNMP 与您所选的网络管理系统 (NMS) 进行集成, 以实现之前由 PowerChute Business Edition 控制台和服务器提供的发现、监控和控制功能.

In this example, “previously provided” is used as the postpositive attributive of “the discovery, monitoring and control”. However, the machine translation output failed to modify the word order, instead producing a translation that was illogical and rife with semantic errors based on the original text’s lexical properties and order. This fails to meet the requirement of item (a) in the TAUS guidelines. Sentence is improved in fluency and logic by rearranging the word order after editing. Additionally, it is also more appropriate to use “的” before “发现、监控和控制功能” in Chinese. In order to better align the language with Chinese terminology, the initial “版本” has also been moved to come before the phrase “9.5+” after completing PE. In the meantime, according to the item (e) Use as much of the raw MT output as possible, there is no other editing in the rest translation of the sentence.

4.2.3 Logical Errors

Example 4:

ST: Enter 1 into the ‘Calib Value’, This value could change when a calibration is done if the calibrated value is outside the 5% tolerance, and the gauge is reset.
MT: 在“校准值”中输入1, 如果校准值超出 5%的误差范围, 则在进行校准和重置仪器时, 该值可能会发生变化.
PE: 在“校准数值”中输入1, 如果校准的数值超出5%的误差, 当校准完成时该值可能会发生更改, 且测量仪复位.

Since IT personnel are the target audience for this IT text and value technical precision, the translation should be grammatically, syntactically, semantically and logically correct so as to meet the requirements of “Human Translation Quality” items (a), (b), and (f). The structure of MT’s output makes the logic of the sentence problematic. The sentence’s logic is flawed by the way that MT’s output is structured. The word “and” in the original text bridges two simultaneous elements: “the gauge is reset” happens after “calibration” and concurrently with “value could change”. Therefore, sentence structure should be adjusted while doing PE.

We discovered that items (a), (b), (e), and (g) in “Human Translation Quality” are more instructive through the application of the TAUS guidelines. This practice uses an IT text, which is characterized by its clear language and lack of long, intricate sentence structures. As a result, as stated in item (c), MT typically does not have the situation where translations are added or removed without cause. Furthermore, the practice in question pertains to the translation of IT texts from English to Chinese and does not entail the culturally inappropriate content in item (d) or the spelling mistakes in item (f). The requirements of item (e) are essentially met in PE,

as we also discovered through practice that many MTs already have very high-quality raw output, and in some cases, only the terminological expressions need to be modified. Of course, the MT Post-Editing Guidelines of TAUS applied in this essay is not a perfect standard, and there is still room for refinement in the future practice.

5. Conclusion

By using the "MT+PE" model in practice, we find that the quality of machine translation has significantly improved with the development of NMT, making it a powerful tool for IT text translation. In this paper, we analyze the common error types of MT in IT text translation and show how important PE is to the quality improvement of translation and how important it is to quality assessment. The language service industry needs to create uniform standards to help evaluate the quality and pricing of work when it adopts the MTPE model to improve the efficiency and quality of translation.

References

[1] TRANSPERFECT, SPAIN, ZARETSKAYA A. Optimising the Machine Translation Post-editing Workflow [C]//Proceedings of the Second Workshop Human-Informed Translation and Interpreting Technology associated with RANLP 2019. Incoma Ltd., Shoumen, Bulgaria, 2019: 136-139.

[2] LIU YANG. Frontier Advances in Neural Machine Translation[J]. Journal of Computer Research and Development, 2017, 54(6): 1144-1149.

[3] ISO 18587: 2017 Translation services —Postediting of machine translation output —Requirements[S]. 2017.

[4] YANG Y, WANG X. Predicting student translators' performance in machine translation post-editing: interplay of self-regulation, critical thinking, and motivation[J]. Interactive Learning Environments, 2023, 31(1): 340-354.

[5] TAUS. TAUS MT Post-editing guidelines [EB/OL]. (2016-01-20) [2024-07-06]. https://www.taus.net/resources/reports/mt-post-editing-guidelines.

[6] LUO JIMEI, LI MEI. Machine Translation Error Analysis [J]. Chinese Translators Journal, 2012, 33(5): 84-89.