A Study of Pre-editing Methods at the Lexical Level in the Process of Machine Translation

Fan Feifei
School of Humanities and Foreign Languages
Xi’an University of Posts and Telecommunications, Xi’an, China

Chen Rong
School of Humanities and Foreign Languages
Xi’an University of Posts and Telecommunications, Xi’an, China
Correspondent Author: chenrong@xupt.edu.cn

Wang Xiao
School of Humanities and Foreign Languages
Xi’an University of Posts and Telecommunications, Xi’an, China

Received: 1/7/2022 Accepted: 5/4/2022 Published: 5/24/2022

Abstract
As the demand for translation increases, the role of machine translation in improving the efficiency of translation is increasingly prominent, but due to its inherent limitations, the translation quality of machine translation is not guaranteed. Pre-editing is one of the ways to enhance the quality of machine translated texts. With Google online translation as a tool, the effect of pre-editing on the quality of machine translation is studied by adopting the Bilingual Evaluation Understudy (BLEU) approach. The pre-editing methods of additions, omissions, replacing and terminology preprocessing based on the lexical level are proposed. The results show these methods can play a role in improving the quality of English-to-Chinese machine translations. In most cases, pre-editing at the lexical level is sufficient to generate high-quality machine translation output, but the effectiveness of pre-editing can be further improved when perspective-shifting is involved or when the target language involves contextually consistent verb or person usage.

Keywords: Bilingual Evaluation Understudy, lexical level, machine translation, pre-editing

Introduction
With the continuous progress of machine translation technology, an increasing number of translators have chosen to use machine translation to do their work. Machine translation is a convenient and efficient translation aid, but it is not possible to directly adopt the version of translation achieved through it. Translators usually revise the translated texts through MT in two ways: pre-editing and post-editing. Pre-editing refers to adjusting and modifying the source language before inputting it into the machine so that it is more in line with the characteristics of the machine translation software. It is expected to improve the quality of the machine translation through pre-editing (Wei, 2008). In contrast to pre-editing, post-editing is widely used. Based on this, this paper focuses on the lexical level, with the help of Bilingual Evaluation Understudy (BLEU), exploring the pre-editing method of machine translation and analyzing the application of additions, omissions, replacing and preprocessing terminology so as to improve the quality of machine translation.

Literature Review
The idea of using machines for translation first appeared in the early 1930s, scientists wanted to design a machine that could be used for translation, but due to the limitations of technological development and conditions at that time, the translation machine was not successfully developed. In 1949, the American scientist Weaver published the translation memorandum, formally presenting the idea of machine translation (Feng, 2018). Neural machine translation developed rapidly in 2016, which made major breakthroughs. The performance of machine translation and its quality were significantly improved, replacing the dominant position of statistical machine translation. In the same year, Google launched a neural machine translation system, using sentences rather than phrases as translation units, greatly improving the accuracy and fluency of translation (Zhao et al., 2019). However, the work of machine translation has not yet reached the perfect level which requests no manual adjustment. Feng and Gao (2017) pointed: “Machine translation is a useful tool in the language service industry, and pre-editing and post-editing are important ways to improve the quality of machine translation. In order to produce high-quality translations, only pre-editing is not enough, but post-editing is also needed”. Post-editing has always been given much attention in improving the quality of machine translation while research on pre-editing is of little notice. In the field of computer science, the development of pre-editing system is the focus while in the field of translation, the focus of the research is on pre-editing strategy. In terms of foreign research, Pym (1990) put forward the idea of editing the source text before using machine translation, and summarized the ten rules for simplifying writing. But most of the advice is based on sentence-level considerations. Yoshimi (2001) studied the machine translation of English News titles, proposed to add pre-editing module in the machine translation system, the English News titles were automatically rewritten into ordinary text and then MT was used, in order to improve the quality of translation, but this approach of pre-editing is chiefly to rewrite the text, which makes the operation more complex. Babych (2009) proposed a method of using structural-level manual evaluation to detect translation errors and create automatic pre-editing modes. This method can enhance the automatic pre-editing function of machine translation. Additionally, Ding (2016) examined the pre-editing system of machine translation based on Japanese and English. Arenas (2019) described pre- and post-editing of source text machine translations from an industry and academic research perspective, helping researchers and the
industry to understand the impact of machine translation technology on the translator’s output and working environment, but didn’t make specific recommendations for pre-editing strategies/methods.

In contrast, studies in China focus more on pre-editing strategies. Zhong (2004) conducted one of the earlier studies on pre-editing in China, but only from a theoretical point of view to analyze the importance of human intervention for machine translation. The pre-editing methods summarized by Wei (2008) include format conversion, the use of controlled languages, the establishment of translation memories, etc., but are not verified by specific examples. Zhong (2004), Lin (2009), Wu (2012) and Li (2021) summarized pre-editing strategies by case studies. All the four studies concluded that pre-editing is helpful in improving the quality of translations, but there are limitations to their samples of case study, and they mostly use subjective analysis to evaluate the effectiveness of translations. In summary, the importance of pre-editing research is self-evident. Vocabulary is the constituent unit of sentences, based on the vocabulary level of research is more targeted, can assist the translation machine more smoothly converted to translation, thereby reducing the effort required for post-editing, improve translation efficiency. In addition, the use of machine translation automatic evaluation method is also essential to increase the objectivity of the study.

**Research Methodology**

Pre-editing methods have been proven in practice and shown to make machine translation result in higher quality translations. Specifically, common pre-editing methods proposed in Wei (2008) include formatting conversion of the original text, spelling and punctuation checking, grammar checking, sentence adjustment, and using controlled natural language; Cui (2016) mentioned that pre-translation editing usually involves two parts: format processing and language processing; Huang (2017) summarized pre-editing methods including restructuring the original text, adding and subtracting, translating terms in advance, simplifying the original text, and adding punctuation; and the main pre-editing guidance of Guo (2017) included lexical and syntactic levels of segmentation, semantic integration of discrete words, semantic segmentation of inflectional words, meaning group segmentation and sentence breaking of long sentences. After a comparative study, it was concluded that lexical additions, omissions, replacing and preprocessing terminologies are ideal methods for pre-editing at the lexical level.

**Research tools**

The tool for quality evaluation of translations is the Bilingual Evaluation Understudy (BLEU). Currently, the mainstream translation evaluation methods are mainly manual evaluation and automatic machine evaluation. Since there is no absolute standard for the expression of the translation, when the evaluation is done manually, the evaluation results will be highly subjective due to the differences in knowledge, experience and thinking styles of different people, etc. To avoid this effect, the automatic machine evaluation method is used.

BLEU is an automatic evaluation method of machine translation based on N-tuple grammar, which was proposed by IBM (Papineni et al., 2002). Its basic idea is to calculate the ratio of the number of matching N-tuples to the total number of N-tuples in machine translation by comparing
N-tuples appearing in translation with reference translation, that is, comparing the translation produced by machine translation with reference translation provided by human translators, the closer it is, the higher the accuracy of machine translation.

The overall evaluation formula of BLEU is as follows.

$$BLEU = BP \exp(\sum_{n=1}^{N} \omega_n \log p_n)$$

where $P_n$ is the proportion of $n$-word phrases appearing in the reference translation to the total number of $n$-word phrases in the machine translated translation, $\omega_n=1/N$, $N$ is the maximum $n$-word grammatical order. Previous studies have demonstrated that, despite some shortcomings of BLEU, experiments have shown a strong correlation between its ratings and manual ratings.

**Research Materials**

The materials are selected from *Translation Theory and Techniques* (He, 2009), and such texts have authoritative reference translations as evaluation criteria, while suitable pre-editing methods can be selected according to their machine translation results.

**Research Process**

Using Google Translation as the machine translation engine, firstly, the original text is input into Google Translation Engine to obtain the first set of translations, and then the pre-editing based on the lexical level is carried out for the problems in the first set of translations, including four aspects of word addition, omitting, replacing and preprocessing terminologies; then the pre-edited text is input into Google Translation Engine to obtain the second set of translations; finally, by comparing the BLEU values of the two sets of translations, it is judged whether the effect of the pre-editing based on the lexical level is significant.

**Results**

**Application of pre-editing methods at the lexical level**

Chinese and English belong to different language families, and there are great differences in writing rules, semantic contrasts, and syntactic structures between the two. According to the previous experience, the result of machine translation can only play an auxiliary role to the translator, and the effective intervention of the translator is an important part of the translation process.

The pre-editing method has been shown in practice that before inputting the original text, the translator can process it according to the characteristics of the original text, thus enabling machine translation to better identify the content of the original text and improve the quality of the translation.

**Additions**

When translating, the translator can add some words and phrases to make the meaning of the translation more accurate according to the differences in the logical structure and expressions between the English and Chinese languages. Therefore, when pre-editing the original text at the lexical level, adding words to the original text is also one of the ways to improve the quality of the translation.
Example 1
SL₁ (Original text): Man, was, is and always will be trying to improve his living conditions.
Reference: 无论是在过去、现在还是将来，人们总是在尽力改善生活条件。
TL₁ (Translation): 过去，现在，将来将永远努力改善他的生活条件。
SL₂ (After pre-editing): Man, no matter was, is and always will be trying to improve his living conditions.
TL₂ (Pre-edited translation): 人，无论过去，现在还是将来都将努力改善他的生活条件。

The BLEU value of the first set of translations was 22.01 and the second set was 38.21, which increased by 16.2 after pre-editing. In this case, “the past, present and future” contains a conditional relationship, which is not only not translated in the first set of translations, but also the subject is ignored. Therefore, this conditional relationship is added in the pre-editing, and the second set of translations is more perfect and precise.

In addition, words that do not express logical relationships in the original English text, but whose implied logical relationships can be judged according to the context, such as transitions, assumptions, concessions, cause-and-effect relationships, etc., can also be appropriately added to the sentences or between sentences when pre-editing, as shown in Example 2.

Example 2
SL₁: Ice is not as dense as water and it floats.
Reference: 冰的密度比水小,因此能浮在水面上。
TL₁: 冰的密度不如水,它会漂浮。
SL₂: Ice is not as dense as water and so it floats.
TL₂: 冰的密度不及水,因此它会漂浮。

In this case, the machine translation result was improved by adding the implied causative word “so” during pre-editing, and the BLEU value increased from 16.91 to 22.77.

Example 3
SL₁: These concepts, values and objectives provide a solid foundation and clear guidance for our mutually beneficial and pragmatic cooperation.
Reference: 这些理念，价值观和目标为我们的互利务实合作奠定了坚实的基础，并给予明确的指导。
TL₁: 这些概念，价值观和目标为我们互利，务实的合作奠定了坚实的基础和明确的指导。
SL₂: These concepts, values and objectives provide a solid foundation and give clear guidance for our mutually beneficial and pragmatic cooperation.
TL₂: 这些概念，价值观和目标为我们互利，务实的合作奠定了坚实的基础，并提供了明确的指导。
In the translation obtained without pre-editing, there were cases of inappropriate verb-object collocation, such as “laying down clear instructions”, so the corresponding verb “give” was added during pre-editing, and the BLEU value increased by 5.05, which indicated that the quality of the translation was improved.

Therefore, when editing before translation, according to the specific meaning of the context, it is necessary to add corresponding verbs before or after certain nouns, and the verbs mentioned here not only refer to the verbs that appear in the sentence but are omitted later, but also include verbs that are not in the original text, in order to make the meaning of the translation clear and complete, and read more smoothly and naturally, in line with the Chinese expression habits.

Example 4
SL₁: We don’t retreat, we never have and never will.
Reference: 我们不后退, 从没有后退, 将来也绝不会后退。
TL₁: 我们不撤退, 我们永远也不会。
SL₂: We don’t retreat, we never have retreated and never will.
TL₂: 我们不撤退, 我们从未撤退, 也永远不会撤退。

In the first group of translations, the machine translation does not fully convey the meaning of the original text, affecting its comprehension. At the same time, Chinese often uses repetitive expressions to enhance the tone, as in this case. When pre-editing, the duplicate content in the original sentences was added, and the second set of translations obtained expressed the meaning in the original text, and the BLEU value increased by 10.93 accordingly. In this case, the vocabulary addition makes the translated text identical to the original in terms of content expression and avoids translation errors. The same method of pre-editing can also make the translated sentences consistent with the original text in terms of cultural background and word association, without affecting the understanding of the original text.

Example 5
SL₁: To explore the moon’s surface, rockets were launched again and again.
Reference: 为了探测月球的表面，人们一次又一次地发射火箭。
TL₁: 为了探索月球表面，火箭一次又一次地发射出去。
SL₂: To explore the moon’s surface, rockets were launched again and again by people.
TL₂: 为了探索月球表面，人们一次又一次地发射了火箭。

In this case, the action of “rocket launch” occurs passively, but without pre-editing, the translation lacks a corresponding logical subject, and after adding the subject, the BLEU value of the translation increases from 46.54 to 68.36, and the sentence components become relatively complete and logical.
In addition, due to the different expression habits, English often omit the subject phenomenon, when pre-editing, it can be supplemented according to different contextual circumstances.

**Omissions**

Omissions, as opposed to addition, refers to the deletion of words in the original that do not conform to the expression habits of the target language or are repetitive and redundant, or are not translated in the translation, but the translation gives the reader the same feeling as the original. The principle is that the integrity of the meaning of the original text should not be affected, and it should be in line with the Chinese expression norms. Under this principle, all words that violate Chinese expression habits should be deleted to make the translation concise.

Example 1

SL₁: The more he tried to hide his mistakes, the more he revealed them.
Reference: 他越是想掩饰自己的错误，就暴露得越明显。
TL₁: 他越想隐藏自己的错误，就越能揭示它们。
SL₂: The more he tried to hide his mistakes, the more he revealed.
TL₂: 他越想掩饰自己的错误，就越能揭示出来。

The first group of translations translated “them”, while the “more it is revealed” in this case is his own mistake, so the translation does not need to use “they” to reflect it, and the semantic meaning is obvious in the translation. By deleting the pronoun “them” in the pre-editing, the machine translation can obtain a translation that does not violate the Chinese language conventions. Among them, the BLEU value of the first group of translations was 34.78, and the second group was 51.12, with significant pre-editing effect. In addition, the verb “reveal” is used in the machine translation, and the complete verb-object pairing is “reveal the error”, while “reveal” means to point out things that are not easy to see to people, and is not generally used with error, so it needs to be adjusted by the post-editing here.
Example 2
SL₁: I had many wonderful ideas, but I only put a few into practice.
Reference: 我有很多美妙的想法，但只有少数付诸实践了。
TL₁: 我有很多很棒的主意，但我只付诸实践。
SL₂: I had many wonderful ideas, but only put a few into practice.
TL₂: 我有很多很棒的想法，只有一些付诸实践。

In this case, the word “a few” was not translated without pre-editing, resulting in an incomplete meaning of the whole sentence. According to Chinese expression habits, if a subject appears in the first sentence, there is no need for it to appear again in the second sentence if it is still the same subject, but the expression of the first group of translations violates this principle. When pre-editing, “I” was deleted, and the second set of translations was basically consistent with the original text and in line with the Chinese expression habits, and the BLEU value increased from 26.71 to 37.15.

Example 3
SL₁: We have made some achievements, and we must guard against complacency.
Reference: 我们取得了一些成绩，但还要防止自满情绪。
TL₁: 我们已经取得了一些成就，我们必须防止自满。
SL₂: We have made some achievements, and must guard against complacency.
TL₂: 我们取得了一些成就，必须谨防自满。

The unedited translation, like example 2, is not in line with Chinese expression habits, so the subject “we” is removed from the second half of the sentence, and the BLEU value of the second set of translations is enhanced, from 38.1 to 40.52.

Example 4
SL₁: A domestic appliance, for instance, a domestic refrigerator or a household smoke exhaust ventilator, can electrify itself.
Reference: 家用电器可能带电，比如电冰箱或抽油烟机等。
TL₁: 家用电器（例如，家用冰箱或家用排烟呼吸机）可以使自己通电。
SL₂: A domestic appliance, for instance, a domestic refrigerator or a household smoke exhaust ventilator, can electrify.
TL₂: 家用电器，例如家用冰箱或家用排烟通风机，可以带电。

In English, some pronouns that are used as objects can often be omitted from translations, regardless of whether they have been mentioned before. In this case, “itself”, as the object of “electrify”, was processed by machine translation as an actual meaningful component in the first set of translations, resulting in a translation error. Therefore, it is removed in pre-editing, and the original text is faithfully expressed in the second set of translations. Among them, the BLEU
value of the first group of translations is 13.85, and the BLEU value of the second group of translations is 17.27, and the translation quality has been improved after pre-editing.

Example 5
SL₁: Cold rolling enables the operators to produce rolls of accuracy and uniformity, and with a better surface finish.
Reference: 冷轧可以生产出精度高、均匀性好、表面光洁度高的轧制品。
TL₁: 冷轧使操作员能够生产出准确的和均匀的轧辊，并具有更好的表面光洁度。
SL₂: Cold rolling enables to produce rolls of accuracy and uniformity, and with a better surface finish.
TL₂: 冷轧能够生产出准确的、均匀且具有更好表面光洁度的轧辊。

In this example, the logical relationship between the object of “enable” and “operators” in the original text is implied in the meaning of the sentence, so it is not necessary to translate it when translating, in order to make the translation more fluent without changing the content of the original text. Therefore, this can be removed during pre-editing, and the BLEU value of the second group of translations grows by 3.8 accordingly, indicating that pre-editing is effective. However, “accuracy” was translated as “correct” in both sets of translations, but in this context the meaning is “high precision”, which needs to be adjusted by the translator in post-editing.

![Figure 2. BLEU values of the omissions](image)

**Replacing**
Because Chinese mostly uses nouns or omissions, English mostly uses pronouns or words with derivative usage, so the translation skill lies in the flexibility to return what is expressed in the original text. In some contexts, it is difficult for the machine to distinguish which part of the content is actually referred to by the pronoun, and the translation is prone to errors and cannot translate the derived meanings of the words. Therefore, when conducting pre-editing, the corresponding referents can be replaced to eliminate ambiguity.
Example 1

SL₁: I chose three modules in the first year because I also had a part-time job.
Reference: 我第一年选择了三门课程，因为我还有一份兼职。
TL₁: 第一年我选择了三个模块，因为我还有一份兼职工作。
SL₂: I chose three courses in the first year because I also had a part-time job.
TL₂: 第一年我选择了三门课程，因为我还有一份兼职工作。

In this example, without pre-editing, machine translation translates “modules” as “block”, but according to the understanding of the original text, it should mean “course” here, so it is replaced with “course” in pre-editing. The BLEU value for the first set of translations is 40.57, and the BLEU value for the second set of translations is 64.33, which is an increase of 23.76.

Example 2

SL₁: Your analysis of current political situation really hits the nail on the head.
Reference: 你对当前政治局势的分析真是一针见血。
TL₁: 您对当前政治局势的分析确实触动了头脑。
SL₂: Your analysis of current political situation is really sharp.
TL₂: 您对当前政治局势的分析确实很敏锐。

In this example, the derived meaning of “hit the nail on the head” is to get to the heart of the matter or to speak pertinently and appropriately, but the machine translation failed to understand its derived meaning correctly, and the first group of translations was just translated word by word, which is far from what the original text wants to express. When pre-editing, this phrase’s derived meaning is directly replaced in its entirety, and the second set of translations was found to be consistent with this context, and the BLEU value increases by 3.22 accordingly.

Example 3

SL₁: The radio wasn’t working because of a loose connection in the wires.
Reference: 因线路接触不良，收音机不响了。
TL₁: 由于电线连接松动，收音机无法正常工作。
SL₂: The radio wasn’t working because of a bad connection in the wires.
TL₂: 由于电线接触不良，收音机无法正常工作。

In this example, the machine translation of “loose connection” without pre-editing does not conform to Chinese expression habits, and in Chinese, loose connection is generally used to refer generally to the malfunction in household appliances due to circuit connection. According to the expression of the first group of translations, the machine translation focuses on the meaning of “loose”, so it is replaced by “bad” in the pre-editing to obtain the second group of translations that conform to the Chinese expression habits, and the BLEU value increased from 14.58 to 38.76.

Example 4
SL₁: This standard does not purport to address all of the safety concerns, if any, associated with its use.
Reference: 本标准未提及同标准使用有关的所有安全问题。
TL₁: 本标准并非旨在解决与使用相关的所有安全问题。
SL₂: This standard does not purport to address all of the safety concerns, if any, associated with its standard use.
TL₂: 本标准并不旨在解决与标准使用有关的所有安全问题。

In this example, without pre-editing, the machine translation omitted the content referred to by “its”, resulting in a lack of meaning and even changing the scope of application of the entire exclusion clause. When pre-editing, the referential meaning of “its” is supplemented, thus limiting the range of sentence meaning. The content of the second group of translations is more accurate, and the BLEU value grows by 16.19.

Example 5
SL₁: Everybody has a responsibility to the society of which he is a part and through this to mankind.
Reference: 每个人都对他所属的社会负有责任，并且通过社会负责任，而对人类负责任。
TL₁: 每个人都应对他所参与的社会负责，并由此对人类负责。
SL₂: Everybody has a responsibility to the society of which he is a part and through taking a responsibility to mankind.
TL₂: 每个人都有责任对他所参与的社会负责，并通过社会对人类负责。

In this example, “this” refers to a responsibility to the society, if it is just translated as “this”, the semantics will be unclear, so replace “it” with “taking a responsibility” and add the verb to make the semantics more coherent and smooth. Among them, the BLEU value of the first group of translations is 26.2, and the BLEU value of the second group of translations is 32.41, which is a significant effect of pre-editing.
Preprocessing terminologies

The procedural rules of machine translation are based on the basic rules of language, and the corresponding translation is calculated by the data in the corpus, which may not be complete for the collection of certain new terms, new technologies, new policies and other terms, so that the translator can directly translate the terms when editing before translation to reduce the error rate of machine translation.

Example 1
SL₁: Non-equilibrium is always fluctuating around the centerline of the equilibrium, thus promoting the economy to move forward and to be infinitely close to the goal of balanced growth.

Reference: 非均衡始终围绕均衡的中心线上下波动，推动经济不断向前发展，无限接近均衡增长的目标。

TL₁: 非平衡总是在平衡的中心线附近波动，从而促进经济向前发展并无限接近平衡增长的目标。

SL₂: Non-balanced is always fluctuating around the centerline of the balance, thus promoting the economy to move forward and to be infinitely close to the goal of balanced growth.

TL₂: 非均衡总是围绕均衡的中心线波动，从而促进经济向前发展并无限接近平衡增长的目标。

“均衡” refers to the equilibrium in economic activities, which is a relatively stable state, while “balance” is a mathematical concept in a sense, and there are obvious differences between the two. The BLEU value of the second set of translations went up by 10.43.

Example 2
SL₁: While the new plant manager has extensive experience improving productivity, many of the workers aren’t ready to hand him a blank check.

Reference: 虽然新厂长在提高生产力方面很有经验，但是很多工人还是不愿意凡事由他全权处理。

TL₁: 新任工厂经理在提高生产力方面拥有丰富的经验，但许多工人还没有准备好给他一张空白支票。

SL₂: While the new plant manager has extensive experience improving productivity, many of the workers aren't ready to hand him 全权处理。

TL₂: 新来的工厂经理在提高生产率方面拥有丰富的经验，但许多工人还没有准备好交给他全权处理。

The word “a blank check” in this example is a common English expression that literally means a check. By extension, “a blank check” can also be interpreted to mean full authority and
discretion. According to the understanding, which is the case in this context, the BLEU values of the two sets of translations raised from 29.38 to 32.39 after pre-editing.

Example 3
SL₁: Valve regurgitation was assessed according to American Society of Echocardiography recommendations.
Reference: 按照美国超声心动描记术学会的建议评价心脏瓣膜返流。
TL₁: 瓣膜返流是根据美国超声心动图学会的建议进行评估的。
SL₂: Valve regurgitation was assessed according to American Society of 超声心动描记术 recommendations.
TL₂: 根据美国超声心动描记术协会的建议评估瓣膜返流。
“Echocardiography” is a medical term, and the BLEU value of the second set of translations increased by 16.56 by early translation substitution during pre-editing.

Example 4
SL₁: Science only learnt why in the 1940s with the discovery of rejection. When human organs are transferred from person to person, the immune system attacks and destroys what it sees as a foreign substance.
Reference：直到20世纪40年代发现了排斥反应后，科学家才明白了导致器官移植失败的原因。当人体器官从一个人移植到另一个人时，受体的免疫系统即将其视作外来物质进行攻击和破坏。
TL₁: 科学直到1940年代才发现排斥的原因。人体器官在人与人之间转移时，免疫系统会攻击并摧毁被视为异物的物质。
SL₂: Science only learnt why in the 1940s with the discovery of 排异反应 When human organs are transferred from person to person, the immune system attacks and destroys what it sees as a foreign substance.
TL₂: 科学只是在1940年代发现了排异反应，才了解了为什么当人体器官在人与人之间转移时，免疫系统会攻击并破坏它视为异物的物质。

The term “rejection” has multiple meanings and is widely used in different professional fields, with different meanings and expressions. In the context of this example, it is a medical term meaning “rejection reaction”, so it was translated by substitution when pre-editing. The BLEU value of the first group of translations is 13.97, and the BLEU value of the second group of translations is 23.36, and the quality of the translations has been improved.

Example 5
SL₁: The calculation result shows that the maximum stress appeared at the edge weld seam.
Reference：根据计算结果显示，最大应力预计在端接焊缝处。
The translation of “edge weld seam” by machine without pre-editing is significantly different from its actual meaning, so it is replaced in pre-editing, and the BLEU value increased by 15.05 accordingly.

![Figure 4. BLEU values of preprocessing terminologies](image)

**Figure 4. BLEU values of preprocessing terminologies**

**Discussion**

A specific analysis is conducted by selecting five examples from each of the four aspects of pre-editing methods at the lexical level in English to Chinese texts. According to the calculation of the BLEU approach, the BLEU values of each aspect improved to different degrees after pre-editing, including an average increase of 11.97 for additions, 7.28 for omissions, the most significant effect of replacing with an average increase of 14.71, and 10.89 for preprocessing terminologies.

(a) In terms of additions: when the implied logical relationship words, verbs and subjects of verb-object collocation are added to the original text, the translation quality of machine translation can be improved and a more standard, fluent and faithful translation can be obtained;

(b) In terms of omissions: delete redundant words in the original text that are dispensable or affect the expression of sentence meaning, so as to make the translation more fluent and conform to the Chinese translation habits;

(c) In terms of replacing: replacing the terms that affect the machine’s understanding of the original text with the original text that has unclear or ambiguous meanings will facilitate translation and thus improve translation quality;

(d) In terms of preprocessing terminologies: by turning out the terms directly during pre-editing, the error rate of machine translation can be reduced. However, the study found that in cases involving perspective shifting, there are also cases of using verb or noun collocations that do
not fit the context, and in this case, post-editing is still needed to make adjustments.

**Conclusion**

In machine translation, the quality of the output can be effectively improved by adopting pre-editing methods of additions, omissions, replacing and terminology preprocessing on the source text. Due to the limitations of the selected texts, the limited number of texts and the fact that they are only applicable to pre-editing strategies at the lexical level without perspective differences, it is not possible to conduct a more detailed and comprehensive discussion of pre-editing strategies, so it is necessary to keep up with the development of machine translation and other related technologies and to continuously update the research on this topic.

**About the authors**

Fan Feifei is a graduate student of Translation at the School of Humanities and Foreign Languages, Xi’an University of Posts and Telecommunications, China. Her research interest is in translation theory and practice. https://orcid.org/0000-0001-9522-2850

Chen Rong is an associate professor of English at the School of Humanities and Foreign Languages, Xi’an University of Posts and Telecommunications, China. She is engaged in research on English for specific purposes.

Wang Xiao is a lecturer of English at the School of Humanities and Foreign Languages, Xi’an University of Posts and Telecommunications, China. Her research interests are English language education and translation.

**References**


