**Original Research Article**

**Machine Translation Errors Identification and Pre-editing Strategies in Chinese-to-English Translation of Pharmaceutical Instructions**

**Abstract**

With the increasing share of Chinese drugs in the international market, the translation quality of instructions has become an important factor affecting the internationalization of Chinese drugs. In recent years, machine translation has been widely used in the field of pharmaceutical instructions translation. However, its translation quality is sometimes unreliable and errors occur frequently, which brings potential risks to the accurate use and safety supervision of drugs. In this study, a total of 10 Chinese pharmaceutical instructions and their corresponding English translations were selected by convenient sampling method. Based on the Multidimensional Quality Metrics (MQM) model, this study classifies and analyzes the common errors in the translation of Chinese pharmaceutical instructions via DeepL. The findings show that the errors of Chinese pharmaceutical instructions are mainly concentrated in three aspects: terminology, accuracy and fluency. In response to the above issues, this thesis proposes six pre-editing strategies: explanation, terminology standardization, annotation, word order adjustment, omission, and addition. It is hoped to provide insights into the optimization of machine translation workflows in medical contexts and contribute to improving the translation quality of Chinese pharmaceutical instructions to better meet the international market demand.

**Key Words**:*Machine translation, pre-editing strategies, pharmaceutical instructions, Chinese-English translation*

**1.Introduction**

As has been the case with the process of globalization, global cooperation in the drug business between countries and global trade in medicine have been common, and translation requirements of pharmaceutical instructions into multiple languages have been increasingly needed. Meanwhile, as artificial intelligence has been developing rapidly, the application of machine translation in medicine has been widespread as well. Notably, machine translation was used on a large scale in Chinese-English and English-Chinese pharmaceutical instruction translation.Drug texts are highly technical in nature with characteristics of technical terms, standard format, and specific content, the reliability of translation is of prime concern.

While machine translation can improve translation efficiency, it is still prone to comparatively high error rates when applied to pharmaceutical instructions, which affects drug safety and therapeutic action. Hence, learning how to optimize the quality of source texts and reduce machine translation errors is of immense practical value and importance. At the same time, as China’s pharmaceutical industry continues to go global, improving the accuracy and standardization of translations of pharmaceutical instructions not only meets international standards in the field of medicine but also plays an important role in advancing the internationalization of Chinese drug products. Therefore, this research attempts to improve Chinese-English pharmaceutical instruction machine translation through an in-depth analysis of the types of errors and pre-editing methods. It mainly addresses the following two questions:

RQ1.What are the common errors in the machine translation of pharmaceutical instructions from Chinese to English?

RQ2.How to reduce the errors in the machine translation of pharmaceutical instructions via pre-editing strategies?

**2. Related work**

This section reviews previous studies on pre-editing strategies in machine translation and the translation of pharmaceutical instructions, highlighting key developments, challenges, and research gaps.

With the advancement of artificial intelligence, translation technology has undergone rapid development, bringing transformative changes to the language services industry (Wang & Liu, 2021). Computer-assisted translation (CAT) tools have significantly enhanced translation speed and quality, attracting increasing attention from both academia and industry (Liu, 2021). Consequently, CAT has become a widely adopted approach, with human–machine collaboration emerging as the mainstream model for professional translation (Wang, 2020). When used effectively, machine translation (MT) can alleviate translators’ workload and improve overall efficiency. However, its quality still falls short of that achieved by professional human translators (Cui, 2014), with frequent issues such as incoherence and ambiguity persisting in MT outputs (Zhao, 2023).

The relatively low quality of MT is largely attributed to recurring error types. While MT can approximate the source text’s meaning, common errors such as inappropriate terminology, inaccurate information, unnatural phrasing, and irregular grammar remain prevalent (Yang & Leng, 2024). Wang (2023) found that in the Chinese–English translation of literary texts, MT errors appear across vocabulary, syntax, and discourse levels, indicating that MT cannot yet replace human translators.Beyond linguistic errors, MT systems also face challenges in processing multimodal inputs, such as text combined with visual context. Paul (2024) highlighted that even advanced multimodal MT models struggle with low-resource languages and cultural-specific expressions, reinforcing the need for human intervention, whether in the form of pre-editing or post-editing, to ensure accuracy.

Despite these limitations, MT is increasingly applied in specialized domains, including medicine and law. However, concerns about its reliability remain (Zeng & Wang, 2020). In response to these shortcomings, the present study aims to enhance MT quality and reduce errors through pre-editing strategies, guided by the Multidimensional Quality Metrics (MQM) framework. Specifically, this study evaluates errors in DeepL’s Chinese–English translations to explore ways to optimize pre-editing techniques.

Empirical studies support the value of pre-editing. Wang (2024) observed that MT outputs derived from pre-edited source texts contained fewer errors. Alarcón and Martínez (2016) demonstrated the effectiveness of pre-editing rules in reducing errors in English-to-Spanish news translation. Similarly, LADD (2009) confirmed the positive impact of pre-editing through experimental evaluation. Similarly, Sun and Yang (2023) demonstrated the effectiveness of pre-editing in academic abstract translation, where strategies such as terminology standardization and sentence simplification significantly reduced post-editing effort for Chinese-English outputs. Their findings further validate the role of pre-editing in specialized domains beyond news or medical texts.

A range of pre-editing strategies have been proposed. Zhang (2025) recently synthesized pre-editing research across domains, noting that strategies like segmentation and terminology management are universally effective, while domain-specific adaptations require further exploration. Hu (2012) outlined techniques such as reordering, segmentation, rewriting, paraphrasing, abbreviation, and adding or omitting elements. Liang (2023), analyzing Patient Zero: A Curious History of the World’s Worst Diseases, proposed strategies for pre-editing medical popular science texts, including creating terminology databases, substituting synonyms, restoring pronouns, and restructuring sentences. Cao and Ma (2018), in their study of news release translation, emphasized word order adjustment and supplementing or omitting content as key pre-editing methods.

However, most existing research focuses on English–Chinese translation, with relatively limited attention to Chinese–English pre-editing strategies, particularly in the context of pharmaceutical instruction translation (Gui & Yang, 2022). Existing work on pharmaceutical translation has largely centered on Chinese patent medicine instructions. For instance, Lin, Zhou, and Li (2010) analyzed common error types in the English translation of domestic pharmaceutical instructions. Liu (2021) introduced a "PE + MT + CAT + PE" model for Chinese patent medicine instruction translation and recommended pre-editing measures such as standardizing formats, adding missing sentence elements, and adjusting word order.

The effective application of pre-editing strategies can enhance MT accuracy and efficiency, ensuring patients accurately understand drug-related information and promoting safe medication practices. This has broader significance for the international dissemination of pharmaceutical information and the advancement of cross-cultural medical communication.

Translation studies on pharmaceutical instructions are grounded in various theoretical frameworks aimed at addressing the challenges of precision and readability in professional texts. Zhang (2024) applied Nord’s text analysis model to examine intratextual and extratextual features in pharmaceutical instructions, focusing on passive constructions, medical terms, and ambiguous vocabulary. The study advocated standardization and selective free translation to improve quality. Shi (2023), employing eco-translatology, emphasized adaptive translation choices across linguistic, cultural, and communicative dimensions when handling passive voice in Chinese texts. Similarly, Wang (2023), under Translation Variation Theory, proposed adaptive strategies such as amplification and omission to meet the needs of specific readerships, underscoring the flexibility required in pharmaceutical translation.

Beyond linguistic conversion, pharmaceutical instruction translation must also address cultural differences and audience expectations. Li and Tan (2023), analyzing Olaparib film-coated tablet instructions, highlighted the translator's subjectivity—manifested through self-reliance, passivity, and initiative—as a key factor in effectively conveying medical information. Cheng and Wang (2021) explored the translation of Chinese patent medicine instructions, noting the need to adapt culturally embedded expressions and four-character idioms through strategies such as annotation and domestication, guided by Skopostheorie. Wang (2020) further proposed integrating context theory to balance the standardized nature of source texts with target-language conventions in English TCM instruction translation.

Professional norms and functional goals are essential in pharmaceutical translation. Under the guidance of Skopostheorie, Hu (2023) recommended strategies such as loan translation, transliteration, and formal subject transformation to handle domain-specific terminology and non-subject constructions, ensuring alignment with target readers’ cognitive habits. Wang (2016), using the example of Gongxuening Capsules, also emphasized the primacy of communicative purpose in selecting functionalist translation strategies. Mitka (2007) cautioned that negligence in translating drug instructions can endanger patient safety, thus necessitating careful lexical and syntactic transformation. Collectively, these studies advocate for a standardized, goal-oriented translation process to ensure accuracy, patient safety, and smooth international pharmaceutical communication.

**4. Methods**

Qualitative methods are employed in this study to explore Chinese-English machine translation of pharmaceutical instructions error types and pre-editing methods. By closely reading the text, qualitative methods can facilitate a deeper understanding of internal structure and linguistic features rather than quantitative data presentation. As pharmaceutical instructions contain highly specific terminologies and complex sentence structures, machine translation errors in such kinds of texts should be complex and diversified. By using qualitative analysis, it is possible to identify specific error types and provide pragmatic solutions on how to pre-editing, which can enhance translation quality.

As for data collection, this study used convenience sampling method. It was a non-probabilistic sampling method in which researchers select samples based on availability and convenience, rather than random or stratified sampling (Sudman, 1976). A total of 10 Chinese pharmaceutical instructions and their English translations were selected from the official medical websites. DeepL was selected as the MT tool due to its higher accuracy, compared to other MT systems Google Translator and Microsoft Interpreter (Cotelli et al., 2023).

In the data analysis phase, this study utilized the Multidimensional Quality Metrics (MQM) framework which is a flexible system for evaluating the quality of translated texts. According to this framework, errors were mainly categorized into accuracy, fluency, terminology, style, locale, convention, and other source error (Lori Thicke & Arle Lommel, 2012). Under each of such broad categories, there existed more specific subcategories and descriptions, as listed in Table 1. In order to facilitate such analysis, the present research constructed a Chinese-English Parallel dataset of ten pharmaceutical instructions. By comparing the Chinese original text and two English translations, one reference translation and one DeepL translation, the present research identified a variety of error types that emerged in translation.

Based on these identified errors, the study proposed a set of pre-editing strategies aiming at reducing such errors in machine translation of pharmaceutical instructions. To verify the effectiveness of these strategies,a Python script was employed to automatically calculate and compared BLEU scores of the translations before and after pre-editing. The BLEU score, widely used in assessing MT quality, measures the similarity of word groups (n-grams) between machine-generated and human reference translations (Papineni et al., 2002). Higher BLEU scores mean indicate better alignment with human translations.

|  |  |
| --- | --- |
| Error Category | Description |
| Accuracy | Addition | Translation includes information not present in the source. |
| Omission | Translation is missing content from the source. |
| Mistranslation | Translation does not accurately represent the source. |
| Untranslated text | Source text has been left untranslated. |
| Fluency | Punctuation | Incorrect punctuation(for locale or style). |
| Spelling | Incorrect spelling or capitalization. |
| Grammar | Problems with grammar,other than orthography. |
| Register | Wrong grammatical register(eg,inappropriately informal pronouns). |
| Inconsistency | Internal inconsistency(not related to terminology). |
| Character encoding | Characters are garbled due to incorrect encoding. |
| Terminology | Inappropriate for context | Terminology is non-standard or does not fit context. |
| Inconsistent use | Inconsistent use |
| Style | Awkward | Translation has stylistic problems. |
| Locale | Address format | Wrong format for addresses. |
| convention | Currency formatDate format | Wrong format for currency.Wrong format for dates. |
| Name format | Wrong format for names. |
| Telephone format | Wrong format for telephone numbers. |
| Time format | Wrong format for time expressions. |
| Other |  | Any other issues. |
| Source error |  | An error in the source. |
| Non-translation |  | Impossible to reliably characterize distinct errors. |

**Table 1 MQM Framework**

1. **Error Types Identification in the Translation Process**

In the process of Chinese-English machine translation of pharmaceutical instructions, the identification of error types and the formulation of pre-editing strategies are the key links to improve the quality of translation. Error types not only affect the professionalism and readability of the translation, but also may pose potential risks to the safety of drug use and the effectiveness of information transmission. The proposed pre-editing strategies would help to reduce the error rate of machine translation and improve the overall translation quality.

Based on MQM, this study identified a total of 376 errors in DeepL translation, which were categorized into teminology, accuracy and fluency errors, as illustrated in Figure 1. Specifically, terminology errors accounted for 43%, accuracy errors for 32%, Fluency errors for 25% of the total errors. These three error types collectively formed the focal of the data analysis.



**Figure 1 MQM Error Counts in DeepL**

According to the MQM framework, terminology errors are mainly categorized into two subtypes: inappropriate for context and inconsistent use. In this study, it was found that there were 107 inappropriate terminology errors and 55 inconsistent ones. An example of inappropriate context can be found in the word‘益气活血’, which is translated into ‘Replenishes Qi, activates blood circulation’ by DeepL. This translation lacks semantic depth, as it does not convey the original meaning in the context of TCM. Within the conceptual framework of TCM,‘益气’refers to the function of tonifying qi and enhancing qi, and a more accurate translation should be ‘tonifies Qi’. ‘Tonify’ is a medical term commonly used in TCM, referring to the enhancement or supplementation of body functions, such as qi, blood, yin and yang, through drugs or therapies.

As for accuracy errors, they are mainly categorized into four subtypes: addition, omission, mistranslation, and untranslated text. In this study, it was found that there were 20 addition, 16 omission, 80 mistranslation and 4 untranslated text. Through analysis, it was found that the concept of syndrome differentiation and treatment unique to traditional Chinese medicine is often simply corresponding to the description of western medicine symptoms. An example of mistranslation can be found in the phrase‘消暑利湿’,which is translated into ‘relieve summer heat and promote dampness’ by DeepL. This translation is misleading, as it fails to convey the intended meaning within the context of TCM. In TCM, ‘利湿’ refers to the process of eliminating excess moisture from the body to achieve effects such as clearing internal heat, resolving phlegm, and promoting diuresis.

 Fluency errors are mainly divided into six subtypes: punctuation, spelling, grammar, register, inconsistency, and character encoding. In this study, it was found that there were 15 punctuation, 10 spelling, 12 grammar, 11 register, 46 inconsistency,. An case of grammatical errors is in the sentence ‘服药后请勿驾驶车辆’, which is translated by DeepL as ‘after taking medicine please do not drive vehicles’. This kind of translation contains a word order error, as it does not follow the conventional English syntax for negative constructions, hence weakening the intended warning. According to standard English warning norms, negative words are usually placed in a fronted position to convey urgency and emphasis.

**6. Pre-editing Strategies in Translating Pharmaceutical Instructions from Chinese to English**

Based on the aforementioned common errors in DeepL translations, this study proposed a set of pre-editing strategies to address issues related to terminology, accuracy, and fluency. Before applying the pre-editing strategies, the BLEU score was calculated at 0.1028. After pre-editing, the BLEU score increased to 0.5136. This significant improvement suggests that pre-editing strategies enhanced translation quality by approximately 40%.

**6.1 Pre-editing Strategies in Terminology Errors**

 The pre-editing strategies for terminology errors include explanation, terminology standardization, and annotations. Explanation is to make implicit, vague, or culturally embedded information in the source text explicit in the pre-translation process. Terminology standardization refers to replacing synonymous or inconsistent expressions with domain-approved terminology. Additionally, use of domain-specific glossaries or annotations is to enhance translation fidelity. Here are some examples.

**Example 1**

**Source text:** 辛凉解表

**DeepL translation:** Releases exterior with pungent-cool herbs

**Pre-edited source text:** 性味辛凉的药物解除风热表证

**DeepL translation after pre-editing:** pungent and cooling medicine to resolve wind-heat symptoms

**Reference translation:**resolving exterior with coolness and acridity

In example 1, ‘辛凉解表’ is the core treatment of Wind-Heat Exterior Syndrome in TCM. It refers to the use of pungent-cool drugs to relieve wind-heat pathogens on the surface by sweating and dispelling wind-heat. DeepL translates‘辛凉解表’as ‘releases exterior with pungent-cool herbs’ without conveying the logic of TCM treatment. In English, ‘release’ refers to physical release, which has nothing to do with the treatment of traditional Chinese medicine and is inconsistent with the context of TCM. Pungent drugs have the effects of dispersing, promoting *qi*, and penetrating the surface. Cool drugs can clear heat and reduce fire.

In the context of pathogenic heat, the pathological basis of‘解表’(relieving the exterior) lies in the invasion of wind-heat pathogens into the body’s surface defense system, leading to an imbalance between *Ying* (nutritive *qi*) and *Wei* (defensive *qi*). *Wei qi*, or defensive energy, is responsible for protecting the body’s exterior. When disrupted, it can result in symptoms such as fever, chills, and spontaneous sweating.The treatment goal is to restore the normal function of *Wei qi* by sweating and expelling the invading pathogens. Through explanation, the original text is interpreted as a machine-translatable and contextually appropriate sentence, ensuring that the use of terms aligns with the intended meaning.

**Example 2**

**Source text:** 川贝母

**DeepL translation:** Sichuan fritillary bulb

**Pre-edited source text:** Bulbus Fritillaria cirrhosa

**DeepL translation after pre-editing:** Bulbus Fritillariae Cirrhosae

**Reference translation:** Bulbus Fritillaria cirrhosa

In Example 2, ‘川贝母’ is translated by DeepL as ‘Sichuan fritillary bulb’. In this translation, ‘川’ is mistranslated as a geographical indicator referring to Sichuan Province, while ‘贝母’ is rendered as a plant with bell-shaped flowers. However, in TCM, ‘川贝母’ refers to a specific medicinal herb used to clear heat and moisten the lungs. The correct and standardized term in the medical domain is Bulbus Fritillariae Cirrhosae. To ensure terminological consistency and avoid ambiguity, terminology standardization is essential, as it helps machine translation systems maintain lexical stability and produce accurate outputs in specialized contexts.

**Example 3**

**Source text:** 夏枯草

**DeepL translation:** Prunella vulgaris

**Pre-edited source text:** 夏枯草（中药材：Spica Prunellae）

**DeepL translation after pre-editing:** Xia Ku Cao (Chinese herb: Spica Prunellae)

**Reference translation:** Spica Prunellae

In example 3, ‘夏枯草’ (pinyin: *Xia Ku Cao)* is translated by DeepL as ‘prunella vulgaris’. This botanical term is technically correct from a taxonomical perspective, but it fails to present the specific meaning and usage of *Xia Ku Cao* in the context of TCM. In TCM, *Xia Ku Cao* refers to the dried spike of the herb Prunella vulgaris, rather than the whole plant. It endows with medicinal properties and is traditionally used to clear heat and toxin, alleviate liver-related discomfort and soothe coughing. The correct English term for it is Spica Prunellae. To address this type of terminological error, domain-specific annotation “夏枯草（中药材：Spica Prunellae)” is used during the pre-editing process. Such annotation provides disambiguating context that guides the machine translation system toward contextually appropriate and accurate renderings.

**6.2 Pre-editing Strategies in Accuracy Errors**

The pre-editing strategies for accuracy errors mainly rely on explanation, which is a useful way to make culturally or contextually implied meanings explicit. This strategy is particularly important in enabling machine translation system to produce more accurate translation. The following examples illustrate this.

**Example 4**

**Source text:** 活血祛瘀、通脉活络

**DeepL translation:** Promotes blood circulation, removes blood stasis, unblocks collaterals

**Pre-edited source text:**促进血液循环和消除血瘀，疏通经脉和畅通络脉

**DeepL translation after pre-editing:** Activates blood circulation and resolves blood stasis, dredges meridians and unblocks collaterals

**Reference translation:** Activate blood and remove stasis, unblock meridians and activate collaterals

In Example 4,‘活血祛瘀’is the treatment of blood state, which is embodied in improving circulating blood and eliminating stasis. ‘通脉活络’ belongs to the treatment of vascular system, which is reflected in dredging the trunk, activating the branches, and reflecting the treatment concept of‘气血同治’in traditional Chinese medicine. DeepL mistranslated‘通脉活络’into‘unblocks collaterals’, which confused the hierarchical relationship between‘经脉’and‘络脉’. Through explanation, the study clarifies the implicit information in the original text as “促进血液循环和消除血瘀，疏通经脉和畅通络脉”. Such strategy not only conveys the treatment principles of TCM, but also clearly expresses the relationship between meridians (*jing*) and collaterals (*luo*).

**Example 5**

**Source text:** 风湿在表之痹证

**DeepL translation:** wind-damp obstruction syndrome

**Pre-edited source text:** 外感风湿导致的关节肌肉疼痛综合征

**DeepL translation after pre-editing:** Joint muscle pain syndrome caused by exogenous rheumatism

**Reference translation:** exogenous wind-dampness syndrome

In Example 5,‘风湿在表’clearly points out that the cause of the disease is rheumatism and the location of the disease is on the body surface;‘痹证’is a unique pathological concept of traditional Chinese medicine, which refers to the pain symptoms caused by qi and blood obstruction of meridians. DeepL translates‘痹证’into ‘obstruction’, which fails to accurately express the pain characteristics of ‘痹证’, and then omits the disease location information of‘在表’. Through explanation, the‘外感’corresponding to the‘在表’of the source text is retained, the disease location is clarified, and the‘关节肌肉疼痛综合征’accurately summarizes the symptom characteristics. The overall expression is not only in line with the theory of traditional Chinese medicine but also easy to understand.

**Example 6**

**Source text:** 脉微涩而紧

**DeepL translation:** slightly choppy-tight pulse

**Pre-edited source text:** 脉细弱不畅而紧

**DeepL translation after pre-editing:** weak, not smooth and tight pulse

**Reference translation:** tight and a bit rough pulse.

In Example 6, ‘脉微涩而紧’contains three kinds of pulse information: ‘微’refers to weak pulse, deficiency of qi and blood;‘涩’refers to the pulse blood flow is not smooth, there is blood stasis;‘紧’means that the pulse is particularly tight, and the patient has a cold syndrome or pain syndrome. DeepL mistranslated ‘涩’ into ‘choppy’, and ‘choppy’ was a navigation term, losing the specificity of traditional Chinese medicine pulse. Through explanation,‘微’and‘涩’are interpreted as ‘细弱’and‘不畅’, and difficult terms about TCM pulse are edited into easy-to-understand words, which makes it easier for DeepL to better translate accurate and complete translations.

**6.3 Pre-editing Strategies in Fluency Errors**

The pre-editing strategies of fluency errors contain word order adjustment, omission, and addition. Word order adjustment is to adjust the sentence order according to the target language syntax and grammar. Omission refers to the deletion of synonymous repetition modifiers to meet the simplicity requirements of drug instructions. Addition includes supplementing implicit subject or object and adjusting voice to ensure the integrity and accuracy of sentence structure and effectively improve the fluency of translation. Here are some examples.

**Example7**

**Source text:** 每日三次，每次两片

**DeepL translation:** Daily three times, each two tablets

**Pre-edited source text:** 一次2片，一日3次

**DeepL translation after pre-editing:** 2 capsules each time, 3 times daily

**Reference translation:** 2 capsules each time, 3 times daily

In Example 7, the time adverbial ‘ daily ’ is wrongly limited to ‘ three times ’ rather than the whole administration behavior, which violates the standard expression of ‘verb + dose + frequency’ in English pharmaceutical instructions. Through adjusting the word order, the route of administration is supplemented to construct a complete sentence structure. Then, the word order is reconstructed according to the principle of ‘dose centered, frequency postposition’, and the standard sentence pattern of ‘X units, X times daily’ is formed.

**Example 8**

**Source text:** 服药后可能出现轻微的少许头晕

**DeepL translation:** After taking the medicine, slight minor dizziness may occur

**Pre-edited source text:** 服药后或现轻微头晕

**DeepL translation after pre-editing:** Slight dizziness may occur after taking the medicine

**Reference translation:** Mild dizziness may occur after taking this medication.

In Example 8 , the overlapping use of ‘slight’ and ‘minor’ as synonymous degree adverbs not only violates the EMA’s regulatory requirements for a single degree of modification of the description of adverse drug reactions, but also does not meet the basic principles of conciseness and accuracy in the WHO guidelines for the preparation of pharmaceutical instructions. Through omission, the redundant word ‘minor’ is deleted, and the most accurate ‘slight’ is retained as the only degree modifier. The word order is adjusted to highlight the core symptoms and make the expression more concise and accurate.

**Example 9**

**Source text :** 儿童必须在成人监护下使用

**DeepL translation :** Children must be used under adult supervision

**Pre-edited source text:**儿童必须在成人的监护下使用本品

**DeepL translation after pre-editing:**Adult supervision is required throughout pediatric medication

**Reference translation:** Pediatric use requires adult supervision.

In Example 9, the object ‘本品’ implied in the original Chinese sentence is not explicit, which leads to the machine mistaking ‘儿童’ as the action object, resulting in the ambiguity of children being used; the use of passive voice violates the international norms that pharmaceutical warnings should use active voice. Through addition, a ‘subject-predicate-object’ structure is established to reveal the implicit components in the Chinese pre-editing stage. Specifically, the adding of ‘本品’ helps to clarify the action object, hence reconstructing key grammatical elements that may be omitted in the original sentence.

After applying various pre-editing strategies to DeepL translations, the data indicates a significant improvement in translation quality. Terminology errors decreased from 43% to 8%, accuracy-related errors were reduced from 32% to 6%, and fluency errors dropped from 25% to 4%. These results demonstrate that pre-editing substantially enhances the overall performance of machine translation in specialized domains.

**7. Conclusion**

This study investigated the common MT errors and the pre-editing strategies on improving Chinese–English translations of pharmaceutical instructions. By analyzing three frequent error types, terminology errors, accuracy errors, and fluency issues, the research demonstrates that six pre-editing strategies could enhance translation quality, namely explanation, terminology standardization, annotation, word order adjustment, omission, and addition). The result show that translation quality improved by approximately 40% after applying these strategies. These findings confirm that pre-editing plays a crucial role in mitigating the typical shortcomings of machine translation, improving clarity and readability in medical context.

However, the study's scope was limited to a single MT tool DeepL. Future research could explore additional error categories, evaluate different MT systems, and incorporate broader datasets to further refine and validate pre-editing strategies for wider application in the medical and pharmaceutical translation domains.

Funding Project: 2024 Guangxi Philosophy and Social Science Research [Project number: 24WYL008]; Key Project of the 2024 Open Research Program of the Guangxi Research Institute for the Consciousness of the Chinese National Community [Project number: 2024GXMGY0101]

**COMPETING INTERESTS DISCLAIMER:**

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

Disclaimer (Artificial intelligence)

Author(s) hereby declare that ChatGPT-4 was used to proofread the language and correct the APA format in this manuscript. The input prompts were as such: “You are a professor in translation, please proofread the following text, making them fluency and readable in academic domain” “Please correct the following references into APA formats”.

**References**

1. Cao, Y& Ma, Y. ( 2018 ). pre-editing of Chinese-English translation of Google translation-Taking the English translation of news releases of Beijing Traditional Chinese Medicine Information Network as an example. *Overseas English*, ( 15 ), 153-155.
2. Cheng, A. (2021). Translation strategies of traditional Chinese medicine instructions based on the English translation of Qingfei Paidu Decoction. *Shizhen Traditional Chinese Medicine and Pharmacology*, 32(6), 1534–1536.
3. Cotelli Kureth, S., Delorme Benites, A., Haller, M., Noghrechi, H., & Steele, E. (2023). *“I Looked It Up in DeepL”: Machine Translation and Digital Tools in the Language Classroom*. In Human Translation and Natural Language Processing Towards a New Consensus? Fondazione Università Ca’ Foscari.
4. Cui, Q. (2014). On post-editing in machine translation. *Chinese Translators Journal*, (6), 68–73.
5. Gui, T., & Yang, Z. (2023). Common errors in medical literature machine translation and pre-editing approaches: A case study of Google Translate. *English Plaza*, (13), 37–40.
6. Hu, M. (2012). *Pre-editing of Chinese-English machine translation* ( Master’s thesis ). Zhejiang Normal University
7. Hu, X. (2023). *A translation practice report on Chinese–English translation of drug instructions* (Master’s thesis). Chengdu University of Technology.
8. Julia Mercader Alarcón & Felipe Sánchez Matínez.(2016).Analysis of translation errors and evaluation of pre-editing rules for the translation of English news texts into Spanish with Lucy LT.*Tradumàtica: traducció i tecnologies de la informació i la comunicació*, (14),172-186.
9. Li, X., & Tan, Y. (2023). On translator subjectivity in the translation of imported drug instructions: A case study of Olaparib tablets. *Overseas English*, (10), 27–30.
10. Liang, C. ( 2023 ). *A report on the E-C translation of Patient Zero : A History of the World 's Deadly Infectious Diseases You Don 't Know ( excerpt ) based on the pre-translation editing*(Master’s thesis). Shanghai Ocean University.
11. Lin, Y., Zhou, Y., & Li, C. (2010). Analysis of typical errors and translation strategies in Chinese–English drug instruction translation. *Forum of Science and Technology Associations*, (3), 164–166.
12. Liu, T. (2021). *A study on the translation of traditional Chinese medicine instructions under the PE+MT+CAT+PE assisted translation model* (Master’s thesis). Jiangxi University of Traditional Chinese Medicine.
13. Mitka,M.(2007).For Non-English Speakers, Drug Label Instructions Can Be Lost in Translation.*JAMA: The Journal of the American Medical Association*,(23),2575-257.
14. Paul, B., Rudrapal, D., Chakma, K., & Jamatia, A. (2024). Multimodal Machine Translation Approaches for Indian Languages: A Comprehensive Survey. *Journal of Universal Computer Science (JUCS)*, 30(5),694-717.
15. PHILIP LADD.(2009). Pre-editing of machine translation input:An experimental evaluation. *Multilingua-Journal of Cross-Cultural and Interlanguage Communication*,(4), 217-224.
16. Shi, Y. (2023). *A translation practice report on English–Chinese translation of passive voice in imported drug instructions* (Master’s thesis). Hebei University of Science and Technology.
17. Sun, Y. C., & Yang, F. Y. (2023). Exploring the process and strategies of Chinese–English abstract writing using machine translation tools. *Journal of scholarly publishing*, 54(2), 260-289.
18. Wang X. ( 2016 ). This paper discusses the translation strategies of drug instructions from the English translation of the instructions of 'Gongxuening Capsule'. *English Square*, ( 05 ), 42-43.
19. Wang, H. (2020). Research on translation technology in the era of artificial intelligence. *Foreign Languages and Cultures*, (1), 85.
20. Wang, H., & Wang, Y. (2023). A review of domestic research on computer-aided translation (1980–2021). *Foreign Languages and Literature*, (2), 130–138.
21. Wang, J. (2023). Analysis of error types in Chinese–English machine translation. *Overseas English*, (15), 12–15.
22. Wang, X. (2016). Translation strategies for drug instructions: A case study of the English translation of “Gongxuening Capsule”. *English Plaza*, (5), 42–43.
23. Wang, X. (2023). *A translation practice report on English–Chinese translation of drug instructions under the theory of transformative translation* (Master’s thesis). Xidian University.
24. Wang, X. (2024). A study of pre-editing strategies based on error types: A case study of Border Town under neural network translation. *New Chu Culture*, (21), 56–59.
25. Wang, Y., Liu, A., & Wu, Q. (2020). Research on the writing and English translation of external-use traditional Chinese medicine instructions. *World Journal of Integrated Traditional and Western Medicine*, 15(9), 1756–1760.
26. Yang, B., & Leng, B. (2024). An empirical analysis of machine translation error types in professional texts based on the MQM quality assessment model. *Journal of University of Shanghai for Science and Technology (Social Science Edition)*10, 1–7.
27. Zeng, J., & Wang, Y. (2020). Error type analysis in English–Chinese machine translation: Based on medical corpus. *Overseas English*, (2), 39–41+47.
28. Zhang, Xue-yao (2025). A Review of Research on Pre-editing of Machine Translation. J*ournal of Literature and Art Studies*, 15(1), 36-43.
29. Zhang, Z. (2024). *A translation practice report on Chinese–English drug instruction translation guided by Nord’s text analysis model* (Master’s thesis). Xi’an International Studies University.
30. Zhao, Y. (2018). Exploring interactive translation strategies in Chinese–English machine translation. *Overseas English*, (17), 149–151.