A Study on the Translation of Biomedical Terminology under the MT+PE Mode

Ying Mi[1] Wei Xu[2]

[1,2] *Hubei University of Chinese Medicine, Wuhan, Hubei 430070, China*
[1]*Corresponding author.*

ABSTRACT

The digital and intelligent era has fostered the robust development of computer technology, deepening its application across various industries. At present, computer-aided machine translation is increasingly utilized in everyday translation practice; however, there are still some deficiencies in how machine translation handles highly specialized and technical scientific texts. Therefore, guided by the Communicative Translation Theory, this paper investigates the translation of biomedical terminology under the "MT+PE" (Machine Translation + Post-Editing) model. Using three mainstream machine translation tools commonly employed in domestic translation practices — Youdao Translator, Baidu Translator, and DeepL Translator — as examples, the study examines the strengths and weaknesses of MT in the English-Chinese translation process and proposes post-editing methods. This research aims to provide insights for medical English translation and further improve the quality of translations in medical texts.

Keywords: *"MT+PE" translation model, Medical English, Machine translation.*

1. INTRODUCTION

With the rapid development of science and technology, the application of artificial intelligence has permeated various aspects of production and daily life. In this context, the proliferation of computer translation software has greatly improved the efficiency of translators but has also posed significant challenges and threats. Currently, AI-driven natural language processing tools have brought about transformative changes in the development of artificial intelligence. Due to the enhancement of translation systems and the refinement of computer-assisted technologies, machine translation (MT) has made substantial progress and is increasingly gaining recognition in the translation industry. However, due to the emphasis on logic and standardization in medical texts belonging to the scientific genre, which have high professional requirements, machine translation still has certain limitations when handling such texts.

Machine translation (MT), also known as automatic translation, is a technology that uses computers to convert text from one language (source language) to another language (target language) [4]. Computer software analyzes the input text at the syntactic and semantic levels, then uses translation memories and terminology databases set up by developers to identify the text requiring translation and convert the two languages to produce the translated output. The essence of computer translation lies in the utilization of translation memory technology, delegating tasks such as term matching, consistent translation of names, and reproducing highly similar sentences to powerful computers [7]. However, the quality of the translation output solely from machine translation still contains errors and deficiencies in certain domains, leading to the emergence of the "MT+PE" translation model, which can effectively improve translation quality and accurately convey the connotations of biomedical language and expressions.

2. MEDICAL VOCABULARY CHARACTERISTICS

Vocabulary is one of the fundamental units that constitute the meaning of language, and understanding vocabulary is an important step in accurately grasping the meaning of a text. One of

the reasons why medical English texts are difficult to understand lies in the differences in the composition, meaning, and usage of vocabulary compared to general English. The composition of medical English vocabulary mainly includes three parts: abstract vocabulary, medical terminology (including term abbreviations and complete medical terms), and ambiguous vocabulary (also known as semi-medical vocabulary). Due to the particularity of the vocabulary and writing style, machine translation software also has limitations in processing, thus post-editing personnel need to pay attention to the translation of such vocabulary.

2.1 Unambiguous Nature of Medical Terminology

Medical English covers a wide range of fields, including traditional Chinese medicine, Western medicine, clinical medicine, pathology, and more, resulting in a large number of specialized terms. These terms can be categorized into complete medical vocabulary and medical vocabulary composed of initialisms, which include a large number of proprietary terms. In summary, the former is formally complete and often incorporates Greek and Latin word roots as compound factors. These terms can be analyzed for their word formation, broken down into roots, and matched with equivalent Chinese translations based on their meanings. Additionally, medical English frequently utilizes professional English acronyms composed of initials, which have the characteristic of capitalizing the initial letters of the constituent words to form a special vocabulary. Medical terminology is an important feature of medical English, with semantic unambiguity that cannot be altered.

2.2 Ambiguity in Amphibious Vocabulary

Apart from specialized medical terminology, English medical texts often contain a plethora of amphibious words with multiple meanings, presenting significant differences in translation compared to common English. Directly applying translations from ordinary English to these words are inappropriate; instead, suitable translations should be sought based on contextual changes to achieve accuracy and fidelity. Furthermore, certain amphibious words carry different meanings; when encountered in different disease contexts or linguistic environments, they adopt distinct interpretations, thereby adding a layer of complexity to medical translation.

3. RESEARCH CONTENT

The sentences cited in this paper are extracted from papers and research reports published in well-known international journals, including Science and Nature. These two journals are renowned as the top-tier publications worldwide, representing the pinnacle of human natural science research. All selected sentences in this paper are from biomedical texts. We utilized three popular machine translation software tools in China, namely Youdao Translation, Baidu Translation, and DeepL Translation, for English to Chinese translation. A comparative analysis of the strengths and weaknesses of each software is conducted, along with post-editing suggestions.

4. APPLICATION OF "MT+PE" IN MEDICAL TEXT TRANSLATION

4.1 Terminology with Proper Nouns

In Western medical terminology, there exists a plethora of terms with proper nouns, primarily categorized into personal name terms, deity name terms, geographical name terms, and terms named after biblical or literary figures. Such terms are commonly found in various aspects including common diseases, treatment methods, clinical trial techniques, bacteria, and viruses. As the name suggests, personal name terms are named in honor of individuals who have made outstanding contributions in a specific field or who were the first to discover a certain disease, often named after them; deity name terms, as well as terms named after biblical or literary figures, are named based on the characteristics and images of various gods or different characters; geographical name terms are often named after a place where a certain disease was first discovered or prevalent. In general, medical proper noun vocabulary is an essential component of medical terminology, with a large quantity and lack of unified standards.

- Sample 1: Very recently, scRNA-seq of fibroblasts infected with Chikungunya virus showed an extremely narrow window of opportunity for the cells to express IFNs before viral protein production shuts off the antiviral system.

  Youdao Translation: 最近，对感染奇孔古尼亚热病毒的成纤维细胞的 scrna 测序显示，在病毒蛋白的产生关闭抗病毒系统之前，细胞表达 ifn 的机会窗口非常窄。

  Baidu Translation: 最近，感染基孔肯亚病毒的成纤维细胞的 scRNA-seq 显示，在病毒蛋白产生

关闭抗病毒系统之前，细胞表达 IFN 的机会窗口极窄。

DeepL Translation: 最近，对感染基孔肯雅病毒的成纤维细胞进行的 scRNA 序列分析表明，在病毒蛋白生成关闭抗病毒系统之前，细胞表达 IFNs 的机会窗口极其狭窄。

Post-Edited Translation: 近期对感染基孔肯雅病毒（Chikungunya virus）进行了研究，根据其成纤维细胞的单细胞 RNA 测序结果显示，在病毒蛋白生产导致抗病毒系统关闭前，细胞表达 IFNs 的几率极低。

In this sentence, “Chikungunya virus” is a specialized medical term named after a geographical location. This virus was first discovered during an outbreak in Tanzania in 1952 and was isolated as the Chikungunya virus in 1953. The term “基孔肯雅” is transliterated from the Swahili language spoken in Tanzania, describing the characteristic posture of patients bending due to joint pain. All four translation software platforms provided interpretations of the term in Chinese, with some differences in translational standards, but all translations are understandable within the industry. In terms of standardization of professional terminology and overall fluency, the version translated by DeepL appears to be the most natural. Drawing from the translations of all three versions and performing post-editing by adding the English content within parentheses, the sentence is made more complete. Guided by the theory of communicative translation, this approach allows for greater consideration of the living environment and cultural background of readers in different contexts.

4.2 Abbreviations

Biomedical texts are characterized by accuracy, specialization, and conciseness. The content of these articles often contains a large number of repetitive, complex, and difficult-to-understand vocabulary and sentences. To achieve these characteristics, abbreviations are commonly used to construct words and phrases. The formation of medical abbreviations mainly includes three types: initialism abbreviations, root letter abbreviations in compound words, and consonant letter + initialism abbreviations. Generally speaking, for abbreviations, transliteration and transference methods are mainly used. However, in modern medical English translation, the transference method is commonly used, which means that the abbreviation of the original language is directly transferred into the target language without translation. However, these abbreviations have a positive meaning for readers in the industry but may hinder the reading experience of some readers to a certain extent. Additionally, apart from the varying meanings of some abbreviations in different industries, in the field of medicine, the same abbreviation can also have different meanings in different disease contexts and contexts. For example, “PID” can represent “pelvic inflammatory disease” or “prolapsed intervertebral disc”. Therefore, for ambiguous abbreviations, post-editors need to judge whether machine translation can identify the most appropriate translation based on the input sentence.

- Sample 2: But in many ways influenza represents a different and more formidable challenge than SARS-CoV-2, with the hypermutating feature of its hemagglutinin head and immune evasion propensity of its stem.
- Youdao Translation: 但在许多方面，流感代表着与 SARS-CoV-2 不同且更强大的挑战，其血凝素头部具有超突变特征，其茎部具有免疫逃避倾向。
- Baidu Translation: 但在许多方面，流感代表着与严重急性呼吸系统综合征冠状病毒 2 型不同且更艰巨的挑战，其血凝素头部具有高突变特征，其干细胞具有免疫逃避倾向。
- DeepL Translation: 但在许多方面，流感与 SARS-CoV-2 相比，是一种不同的、更可怕的挑战，因为流感的血凝素头部具有高突变特性，其茎具有免疫逃避倾向。
- Post-Edited Translation: 但在许多方面，流感代表着与 SARS-CoV-2（严重急性呼吸综合征冠状病毒 2 型）不同的、更可怕的挑战，其血凝素头部具有超突变特征，其茎部具有免疫逃逸倾向。

In this sentence, “SARS-CoV-2” is a common medical abbreviation. Its full expression is “severe acute respiratory syndrome coronavirus 2” which is translated into Chinese as “严重急性呼吸综合征冠状病毒 2 型”. As shown in the example above, Youdao Translator and DeepL Translator directly retain the abbreviation in the translation, which can effectively save time and space. Baidu Translator, on the other hand, adds an explanatory translation of the full expression, which helps with understanding. It is well known that “SARS-CoV-2” is familiar to professionals in the medical field. Therefore, to achieve the purpose of communicative translation, when translating such terms, it is necessary to consult the full name and translation of the abbreviation. An explanation can be provided during the first translation, but zero translation should be used for subsequent occurrences. Therefore, the translations provided by the three translators are all worth considering, but

the translator needs to make their own choices during the post-editing process.

4.3 Amphibious Words

English belongs to the Indo-European system and has a complex lexicon due to its incorporation of numerous Greek, Latin, and Saxon elements throughout its development and popularization. Consequently, the phenomenon of polysemy is prevalent. Within this context, translating amphibious words frequently encountered in medical English poses a significant challenge. Machine translation primarily relies on terminological databases and memory banks, making it difficult to accurately translate amphibious words based on context when inputting limited sentences or paragraphs. Machine translation often selects high-frequency translations from general English, which fails to convey the most contextually accurate meaning, thereby preventing readers from experiencing the same comprehension as native speakers. Thus, post-editing plays a crucial role. Under communicative translation guidelines, the principles for editing amphibious words can be specifically divided into two categories: (1) determining word meaning based on contextual environment, and (2) determining word meaning based on common medical collocations. The former refers to analyzing the textual features and the context to understand the position and function of the vocabulary within the discourse, considering coherence and readability, and then choosing an appropriate Chinese expression for the word's meaning by analyzing its structural components in the sentence. The latter requires translators to consult specialized dictionaries to derive different meanings based on various collocations.

- Sample 3: All segments of the population were naïve to SARS-CoV-2, whereas older children and adults have protective immunity to established viruses because of prior exposures and vaccines.
- Youdao Translation: 所有人群对 SARS-CoV-2 均为 naïve，而年龄较大的儿童和成人由于先前接触和接种疫苗而对已知病毒具有保护性免疫力。
- Baidu Translation: 所有人群对严重急性呼吸系统综合征冠状病毒 2 型都很天真，而年龄较大的儿童和成年人由于之前接触过病毒和接种过疫苗，对已建立的病毒具有保护性免疫力。
- DeepL Translation: 所有人群对 SARS-CoV-2 都不敏感，而年龄较大的儿童和成年人由于以前接触过病毒并接种过疫苗，对已确定的病毒具有保护性免疫力。
- Post-Edited Translation: 所有年龄段群体都未感染过 SARS-CoV-2，然而由于年龄较大的儿童和成人既往接触过病毒并接种过疫苗，因此对现有病毒具备了保护性免疫力。

As shown in Example 3, three translation software tools provided different translations for “naïve”. Upon examination, it was found that this word is a typical example of an amphibious term. Its high-frequency translation in general English is “innocent” or “naïve”, but in medical contexts, its translation varies with different usages and collocations. For instance, “naïve KO mice” translates to “幼生基因消除小鼠”, “biologic-naïve” translates to “生物制剂初治”, and “treatment-naive patients with achalasia” translates to “在未经治疗的贲门失弛缓症患者”. Clearly, only DeepL translation for this term is relatively close, but it still may not be easily understood by readers. By analyzing the context, it is evident that there is a comparative relationship emphasizing exposure to viruses and vaccines. Therefore, in this situation, post-editing the term "naïve" based on its contextual meaning to “未接触过的” is more comprehensible to readers, thereby achieving the goals of fluency and fidelity in communicative translation.

4.4 Abstract Terms

Medical vocabulary is characterized by its professionalism, richness, complexity, and comprehensiveness. The language used in medicine emphasizes logical coherence and standardization, hence a preference for nouns and phrases [4]. Given that abstract nouns are concise, succinct, and highly generalizable, they are widely utilized in medical contexts to denote actions, processes, states, and various abstract concepts. Medical texts demand completely objective descriptions of phenomena or medical changes, avoiding subjective interpretations and emphasizing objective existence. When translating such terms, appropriate conversions are necessary to render the abstract into the concrete. This may involve adding words according to Chinese expression habits or converting word classes to ensure fluency and clarity. By concealing the agent's role and adopting noun-centric structures, the narrative is made as objective as possible, tightening the logical relationships between different parts. Guided by Newmark’s communicative translation theory, conversion of word classes is often required to achieve smooth and accurate translations.

- Sample 4: Analysis of peripheral blood immune cells in pediatric and adult COVID-19 patients showed an increased

frequency of naïve T cells, depletion of natural killer (NK) cells, and a lower frequency of cytotoxic T cells in children compared with adults.

- Youdao Translation: 对儿童和成人 COVID-19 患者外周血免疫细胞的分析显示，与成人相比，儿童中 naïve T 细胞频率增加，自然杀伤(NK)细胞耗竭，细胞毒性 T 细胞频率较低。
- Baidu Translation: 对儿童和成人新冠肺炎患者外周血免疫细胞的分析显示，与成人相比，儿童幼稚 T 细胞的频率增加，自然杀伤（NK）细胞的耗竭，细胞毒性 T 细胞的出现频率降低。
- DeepL Translation: 对儿童和成人 COVID-19 患者外周血免疫细胞的分析表明，与成人相比，儿童中幼稚 T 细胞的频率增加，自然杀伤（NK）细胞减少，细胞毒性 T 细胞的频率降低。
- Post-Edited Translation: 对感染 COVID-19 儿童和成人患者外周血免疫细胞进行分析，结果显示，与成人相比，儿童的初始 T 细胞生成频率增加，自然杀伤（NK）细胞衰竭，细胞毒性 T 细胞生成频率降低。

Due to the differences in language and cultural backgrounds between Chinese and English, there also exist discrepancies in the selection of word classes. In Example 4, the original Chinese text employs four noun structures, enhancing its logicality and scientific rigor. All three machine translations adequately convert word classes, effectively avoiding stiffness in the translations. Newmark's communicative translation theory emphasizes that translations should be target-language-oriented, closely aligning with the target culture, hence providing valuable reference and greatly improving post-editing efficiency. In the PE translation, aside from adjusting word choices for related word class conversions, "analysis of" is translated into a verbal phrase, enhancing the readability of the translation and better reflecting the dynamic characteristics of the Chinese language.

5. CONCLUSION

Medical English, as an important branch of scientific and technical English, is characterized by its professionalism, standardization, and complexity. Its purpose is mainly to communicate information, provide instructions, and persuade, thus emphasizing the accurate and effective transmission of information to readers during translation. Medical vocabulary contains numerous technical terms, abbreviations, amphibologies, and abstract terms, which increase the difficulty of translation. With the development of science and technology, artificial intelligence has been able to alleviate human workload in many aspects, including the field of translation.

This study focuses on the translation practice of biomedical texts and adopts the "MT+PE" translation mode. Guided by Newmark's communicative translation theory, it analyzes the translation activities of three domestic translation software: Youdao Translation, Baidu Translation, and DeepL Translation, and proposes the importance of emphasizing the cooperation between machine translation and post-editing. In general, when dealing with medical texts, the mainstream machine translation software in China, including Youdao Translation, Baidu Translation, and DeepL Translation, can provide relatively fluent and appropriate translations that are worth learning from in many aspects. However, due to the complexity of medical vocabulary, there may still be omissions, errors, or misunderstandings in their translations. Therefore, after machine translation, post-editing based on communicative translation principles can not only improve translator efficiency, but also enhance the readability, accuracy, and professionalism of the translation, as well as ensure the coherence of context translation, ultimately producing translations that are closest to the original meaning.

REFERENCES

[1] Carl A. et al. COVID-19 and Children [J] Science, 377 (6611).DOI: 10.1126/science.ade1675.

[2] Loske, J., et al. Pre-activated Antiviral Innate Immunity in the Upper Airways Controls Early SARS-CoV-2 Infection in Children. [J]Nature Biotechnology (2021).

[3] Cui Qiliang. On the Post-Editing of Machine Translation[J]. Chinese Translators Journal, 2014(6): 68-73.

[4] Chen Meixia. Characteristics of Biomedical English Vocabulary and Translation Strategies[J]. Chinese Science and Technology Translators Journal, 2021, 34(04): 4-6+14.DOI: 10.16024/j.cnki.issn1002-0489.2021.04.002.

[5] Gui Tingting, Yang Zhan. Exploration of Common Errors in Machine Translation of Medical Literature and Pre-Editing Approaches—Taking Google Translation as an Example[J]. English Square, 2023, (13): 37-40.DOI: 10.16723/j.cnki.yygc.2023.13.011.

[6] Li Fengxi. Comparative Study on the Quality of Human-Computer Translation in the Era of Artificial Intelligence. Foreign Languages Journal, 2022, (04): 72-79.

[7] Luo Qichen, Jin Enguang, Li Yafen. Analysis of Computer-Aided Translation Software under the Background of Artificial Intelligence—Taking Youdao Dictionary as an Example[J]. Computer Age, 2022, (08): 43-45+55.DOI: 10.16644/j.cnki.cn33-1094/tp.2022.08.010.

[8] Zhai Qingyong, Feng Yifan. Translation Research on Scientific and Technical English Texts under the "MT+CAT+PE" Translation Mode[J]. English Square, 2023, (22): 41-45.DOI: 10.16723/j.cnki.yygc.2023.22.027.

[9] Wang Hesi, Ma Kexin. Comparative Study on the Application of Artificial Intelligence Translation—Taking Biomedical Texts as an Example[J]. Chinese Science and Technology Translators Journal, 2023, 36(03): 23-26.DOI: 10.16024/j.cnki.issn1002-0489.2023.03.001.