Practice and research of computer-aided medical translation based on big data

Zikai Guo¹ and Chen Chen¹,∗
¹Center for Medical Language and Culture Study, Xi’an Medical University, Shaanxi Province, China
∗Corresponding author’s e-mail: dsgkevin@xmu.edu.cn
∗Corresponding author’s ORCID: https://orcid.org/0000-0002-8526-2805

Abstract. Medical translation significantly influences medical teaching and medical communication, so high-quality medical translation plays an important role in promoting medical development and innovation. Based on the big data era background, this paper studies the practical effects of computer-assisted translation in medical translation. This article first introduces the relevant computer-assisted translation tools in the era of big data, then introduces the innovations and shortcomings of machine translation in the field of translation, and then explains the characteristics of medical translation. In addition, it also analyzes the application of computer-assisted translation in medical translation. Finally, the medical translation effects of machine translation, traditional translation, and computer-aided translation are compared through experiments. The results show that the accuracy rate of machine translation is 67.5%, the accuracy conventional translation rate is 95.4%, and the accuracy rate of computer-aided translation is 87.3%.

1. Introduction
With the rapid development of Internet technology, machine translation has become an auxiliary translation tool in many industries. However, machine translation is often hard to translate, and it is common to cause errors in the translation. With the rise of big data, information technology tends to be intelligent, and computer-assisted translation (CAT) has developed rapidly, and corresponding translation software has emerged, replacing traditional manual translation in many aspects. Medical translation is highly professional. International medical cooperation is getting closer and closer with the promotion of globalization and the Belt and Road Initiative. With the emergence of new technical equipment, diagnosis and treatment methods and drugs, it is inevitable for Chinese doctors to learn new technologies by translating foreign language documents. Not only that, Chinese medical documents must also be translated into foreign languages to promote my country’s advanced medical technology to the world, the demand for medical translation is gradually increasing.

Regarding computer-assisted translation, many scholars have done a lot of research. R Krüger proposed a general model of the usability of CAT tools and a more specific model of the usability of a translation memory system (TMS), and captured all relevant aspects of the usability of CAT tools from a user-oriented perspective. These models are derived from the ISO standard 9241 "Human-Computer Interaction" Ergonomics", he wants to use this method to study the contextualization of CAT tools [1]. When Park K studied the vocabulary conversion between Korean and English, he proposed a constraint that a group of verb assigners may not perform vocabulary
transfer. At the same time, he expects that the translation software will also be equipped with the proposed claims that English verbs are actually borrowed as nouns or defective verbs to escape the direct attachment of inflectional morphemes. Finally, he pointed out that vocabulary transfer constraints have a very large impact on CAT translation [2]. Samad SS, Mohammed OS, Mahdi HS and others have studied the attitudes of professional translators and students in Yemen towards computer-assisted translation tools. They used questionnaire surveys and online interviews to investigate the opinions and opinions of professional translators and students on CAT tools. The results show that they have a positive attitude towards CAT tools [3].

The application of computer-assisted translation in medical disciplines is a very important and meaningful link in the development of modern society. With the continuous advancement of science and technology, people's awareness of things is getting deeper and deeper. As a modern technical means and information processing method in the medical field, it has been used incisively and vividly to show its strong vitality, great potential and versatility. It has been widely used and achieved good results; computer-assisted translation can not only improve work efficiency and quality can also provide an essential basis for the development of medical science.

2. Practice and Research of Computer-Assisted Medical Translation Based on Big Data

2.1. Computer-Aided Translation Tools in the Context of Big Data

With the highly developed information technology, there are endless translation tools on the market, but the most important translation tools are Tradosi and Google. Trados is memory-based translation software that uses special tools to store the work done by the translator to form a corpus and terminology database. When translating, Trados matches the translation memories of these two databases. If the matching degree reaches 100%, it will replace it; if the matching degree does not reach 100%, it will put forward new translation suggestions based on the actual matching degree. The provided translation suggestions translate sentences and store them in the database after completion. This translation method is a human-machine translation method. The system only exists as a translation aid, but it also improves the efficiency and accuracy of translation. Trados can also change the date, time and unit [4]. Google Translate is carried out on the basis of statistics. The content of Google's corpus comes from language texts appearing on the Internet, so Google's corpus is very large and continues to be supplemented. After entering a language in Google Translate, it can detect the type of language and display the most commonly used translations. Google’s language knowledge is accumulated after calculating all sentences in the corpus through a probability model [5].

2.2. Machine Translation under Big Data

2.2.1. Advantages. In the era of big data, machine translation is based on parallel corpora, which can ensure translation accuracy while achieving translation diversity. The corpus not only provides real corpus and data statistics for translation activities, but also has the function of verifying current theories and constructing new theoretical models. For example, "He smiled and said" appears many times in an article, which can be translated as "He smiled and said". If the whole article is translated in this way, the vocabulary is not repeated, and the literary charm will disappear. Translators can obtain smile parallel vocabulary, such as giggle, grin, sneer, beam, etc., through the Internet or from native English-speaking friends. The translator can choose the appropriate vocabulary based on the context and character characteristics of the characters. Through the memory function of the translation software, translators can share termbases, avoid repeated translations, improve translation efficiency and reduce translation costs. In addition, the memory function of the translation software also makes the terminology consistent and improves the quality of translation. As far as Google is concerned, as a search engine, it has created a favorable environment for expanding the corpus.
2.2.2. Limitations. Although big data provides a lot of convenience for translation work, it also has obvious shortcomings. First, the size and domain of the corpus affect the quality of the translation. Only when the information sample is large enough can we approach complex language rules to translate sentences with high matching degrees. Second, the translation work with Trados needs to rely on translation memory. If these translation memories are not correct or do not conform to the current language norms, the accuracy of the translation will also be greatly affected. Third, translation under big data is easy to ignore the grammatical rules, and the translation result does not conform to the language specification, which reduces the intelligibility of the translation. Fourth, the emergence of big data translation software makes translators blindly dependent on search and lack of independent thinking. This not only limits the improvement of the quality of the translation, but also limits the improvement of the translator's translation level and translation innovation ability. Fifth, language cannot convey the psychological changes and emotional characteristics of characters, but machine translation cannot perceive these emotions well and express them in another language. Sixth, machine translation cannot feel the cultural background of the source language. Tang poetry has a very significant historical feature, and at the same time it is a combination of literature and art. Machine translation is either unable to translate or is very crude translation. The translation does not show its historical and cultural color, literary connotation and artistic beauty at all.

2.3. Features of Medical English Translation

2.3.1. More professional vocabulary. Medicine has many professional vocabulary and terminology in various languages, and they all have specific semantics. In Chinese, medicine includes traditional Chinese medicine vocabulary and western medicine vocabulary. There are many medicinal materials, acupoints, and diagnosis and treatment methods in the Chinese medicine vocabulary. In addition, the vocabulary in western medicine has its vocabulary in addition to the introduction of foreign terms. Whether it is translating Chinese medicine into foreign language medicine, or vice versa, the translator needs to have a comprehensive grasp of the different language expressions of medical terms in order to ensure the accuracy of the translation [6-7].

2.3.2. Abbreviations are widely used. Because the original professional terms are too complicated, in order to simplify the writing and expression, a large number of abbreviations are widely used in medical English. These abbreviations often appear in western medicine prescriptions and medical reports. The content of abbreviations generally includes medical journals, medical institutions, drug names, academic groups, etc. [8].

2.4. Syntactic Features of Medical English

2.4.1. Nominal structures appear frequently. Chinese medical literature usually uses verbs to express the occurrence of sports, while in English, medical English generally uses nominal structures, the most prominent of which are nouns evolved from verbs. This helps reduce the frequency of sentences and terms, and can convey a lot of information [9].

2.4.2. Passive voice is used more. In medical English literature, passive voice is often used to express objective facts, which avoids the influence of individual subjective emotions and ultimately achieves a good and clear meaning. This highlights the subjectivity of behavioral objects including diseases and patients, which is also a point of urgent concern [10-11].

2.4.3. Long sentences are common. Parallel sentences and compound sentences are used in medical English literature to express the logical relationships in the process of changing behavior objects. These sentences are long in length, have a particularly large number of clauses, and have complex sentence structures, including side-by-side structure, non-predicate structure, ellipsis, etc. [12].
2.5. Advantages and Characteristics of Computer-Aided Medical Translation

Compared with traditional translation methods, computer-assisted translation has a very significant advantage in improving work efficiency. In traditional translation, translators need to look through a large number of dictionaries and documents, and translate sentence by sentence to complete a translation job. This translation mode requires a lot of time and energy from the translator, which seriously hinders the improvement of translation efficiency. Computer-aided translation solves this problem to a large extent. In addition, the advantages of computer-aided translation include the following:

2.5.1. Provide a large amount of vocabulary for translation work. The essence of the CAT system can be said to be a huge multilingual dictionary. Translators do not need to read the dictionary anymore, and can know the target language corresponding to the source language by searching directly. CAT tools can also use professional knowledge other than grammatical structure and vocabulary to form new language expressions.

2.5.2. It can be translated in a large space. Machine translation is difficult to achieve full text translation, while traditional translation takes a long time. The CAT tool can use translation memory technology to retrieve sentences in the database, analyze the data through the search engine, and obtain the data content related to the source language text in a short time.

2.5.3. Provide terminology management. Medical texts have a lot of professional terminology. Translation memory technology in CAT provides the function of terminology management. It can standardize the terminology usage of the entire article, which is convenient for readers to read, and avoids the problem of disjointed understanding of the article caused by inconsistent terminology.

2.5.4. Provide project management. Medical translation is not just a mutual conversion between languages. In CAT tools, a translation work is integrated into a systematic project. The project management includes document import, field analysis, pre-translation, quality inspection and proofreading, and document export. In addition, CAT tools need to process terminology in advance, and translators can use manual or automatic methods to generate terms and import them into the relevant terminology database. Project management is conducive to step-by-step medical translation work, which can improve translation efficiency while ensuring the standardization and specialization of translations.

3. Experimental Research on the Effect of Computer Aided Medical Translation Based on Big Data

3.1. Experimental Content

In order to explore the applicability of computer-aided translation in medical translation under big data, this article designs experiments for three translation modes: machine translation, traditional translation, and computer-aided translation. After investigation and negotiation, a total of 27 medical translators were selected as experimental subjects for this experiment. Twenty-seven translators used different methods to translate the same medical literature, and analyzed the results in terms of the translation process, duration, accuracy, etc.

3.2. Experimental Process

According to the experimental content, one suitable Chinese medical literature and one English medical literature are selected as the translation text. The 27 translators are divided into three groups: A, B, and C. Group A uses machine translation methods for translation, and Group B and C use traditional translation methods and computer-assisted translation methods for translation. Record the time spent by each translator during the translation process, and finally evaluates the results of their
translation, and performs data statistics and data analysis on the results, expounding the characteristics of the three translation methods in medical translation from an objective perspective. In the process of data processing, the following formulas were used:

\[
\text{Sample variance formula: } s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \tag{1}
\]

\[
\text{Sample standard deviation formula: } s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{2}
\]

4. Analysis of the Effect of Computer-Aided Medical Translation Based on Big Data

4.1. Analysis of the Degree of Complexity and Simplicity of the Translation Process

After the experiment, three groups of translators were interviewed, including their translation steps, how to deal with new words, how to deal with professional terms, and so on. Finally, according to their answers, the translation process is divided into three levels: "easy, general, and cumbersome". The results are as follows:

![Figure 1](image_url)

**Figure 1.** Analysis of the degree of complexity and simplicity of the translation process

As shown in Figure 1, the experimental results show that the computer-aided translation method is the simplest, followed by machine translation, and the most cumbersome is the traditional manual translation method.

4.2. Analysis of Translation Efficiency

| Category | Simple | General | Cumbersome |
|----------|--------|---------|------------|
| A        | 2      | 3       | 5          |
| B        | 1      | 2       | 7          |
| C        | 1      | 1       | 8          |

**Table 1.** Translation time

|   | Max (h) | Min (h) | Average (h) |
|---|---------|---------|-------------|
| A | 2.5     | 2.0     | 2.2         |
| B | 3.6     | 3.0     | 3.3         |
| C | 1.5     | 0.9     | 1.1         |
As shown in Table 1 and Figure 2, for the same medical text, the maximum duration of group A is 2.5h, the minimum duration is 2.0h, and the average duration is 2.2h; group B takes the longest time, with a maximum duration of 3.6 h, The minimum duration is 3.0h, and the average duration is 3.3h; the most efficient group is Group C. The maximum duration of this group is only 1.5h, the minimum duration is 0.9h, and the average duration is 1.1h.

4.3. Analysis of Translation Accuracy

| Category | Average(h) | Min(h) | Max(h) |
|----------|------------|--------|--------|
| A        | 2.2        | 2.5    | 3.0    |
| B        | 3.3        | 3.6    | 4.0    |
| C        | 1.1        | 0.9    | 1.5    |

Table 2 shows the evaluation results of the accuracy of the three groups of translation results. From the perspective of the results, group B has the highest accuracy rate, with an accuracy rate of 95.4%; followed by group C, with an accuracy rate of 87.3%; the accuracy rate is the lowest. In group A, the accuracy rate of this group is only 67.5%.

5. Conclusions

Computer-aided translation has an indispensable position in medical translation, especially in the era of high-efficiency big data. It is convenient and fast, improves the efficiency of medical translation, and avoids unnecessary repetitive query work. But it still has some shortcomings. For example, the choice of vocabulary is not precise and sentence patterns are repeated. If you just want to understand and read the document, computer-assisted translation is a very useful tool. If you want to translate the document into a target language version for teaching or distribution, then the translation quality must be improved.

Acknowledgment

Research Program on Humanities and Social Science of Education Department of Shaanxi Provincial
Government (21JK0343);
“English Language and Literature”, the 4th Key Construction Program of Xi’an Medical University ([2019]96);
Counterpart Supporting Program of Xi’an Medical University (2018PT51, 2018PT52);
Education and Teaching Reform Program of Xi’an Medical University(2018JG-48, 2018JG-51);
Innovation and Entrepreneurship Program of Education Department of Shaanxi Province (S20201840043, S20201840029);
Enterprise-University Cooperative Education Program of Ministry of Education (202002268025).

Reference
[1] R Krüger.(2016) Contextualising Computer-Assisted Translation Tools and Modelling Their Usability[J]. Transkom, 9(1):114-148.
[2] Park K. (2016) A Constraint on Lexical Transfer: Implications for Computer-Assisted Translation (CAT)[J]. Journal of the Korea Society of Computer and Information, 21(11):9-16.
[3] Samad S S, Mohammed O S, Mahdi H S. (2020) The attitudes of professional translators and translation students towards computer-assisted translation tools in Yemen[J]. Dil ve Dilbilimi Çalışmaları Dergisi, 16(2):1084-1095.
[4] Bell T. (2019) TAUS New York speakers say translation's biggest problem isn't tech—it's people [J]. Multilingual Computing & Technology, 30(4):9-9.
[5] Bundgaard K, Christensen T P, Schjoldager A.(2016) Translator-computer interaction in action—an observational process study of computer-aided translation[J]. The Journal of Specialised Translation, (25):106-130.
[6] MÁ Jiménez-Crespo, M Tercedor Sánchez.(2017) Lexical variation, register and explicitation in medical translation: A comparable corpus study of medical terminology in US websites translated into Spanish[J]. Translation and Interpreting Studies, 12(3):405-426.
[7] Tracey D.(2018) An Early History of Medical Translation[J]. Journal der Deutschen Dermatologischen Gesellschaft, 16(10):1300-1301.
[8] Dalton-Oates B M.(2017) Medical translation: The neglected human right [J]. International Journal of Human Rights in Healthcare, 10(4).
[9] Ali L, Milani M.(2019) Medical writing and medical translation - two crossing paths [J]. Medical Writing, 28(1):42-45.
[10] Cai Y, Guo P.(2020) The Application of Actor Network Theory in Medical Translation[J]. Open Journal of Modern Linguistics, 10(5):599-605.
[11] Muoz-Miquel A, V Montalt, I García-Izquierdo.(2020) Fostering Employability through Versatility within Specialisation in Medical Translation Education[J]. Hermes (Denmark), 60:141-154.
[12] Tracey D. (2018) An Early History of Medical Translation[J]. Journal der Deutschen Dermatologischen Gesellschaft, 16(10):1300-1301.