Research on English Translation of Computer-aided Classics Based on Natural Language Processing

Shengqin Bi
Wanjiang University of Technology, Ma’anshan, Anhui, 243000, China
Corresponding author’s e-mail: Lily76272647@163.com

Abstract. The advantages of computer-aided translation of scientific and technological classics into English are becoming increasingly obvious. The production and development of English translation of computer-aided science and technology classics is closely related to computer information technology. International exchanges and cooperation have gradually increased. Computer-aided translation of scientific and technological classics can ensure translation quality, greatly improve efficiency, and reduce complex mechanical work. In the case of machine translation with certain defects, computer-assisted translation of classic English texts is a good alternative, providing effective choices for translators. Among the core technologies of mainstream computer-assisted translation of English texts for scientific and technological classics are translation memory, terminology management, translation project management and other technologies. Based on the principles and characteristics of computer-aided translation, this paper analyzes the technology of computer-aided English translation of classic books based on natural language processing technology, and provides a reference for the use and promotion of computer-aided translation of English classics.

1. Introduction
After the introduction of artificial intelligence, the machine will do some intelligent work as a key research object. Natural language processing (NLP) can analyze and process text at different levels of scale to meet the needs of different levels. The rapid development of information technology and the continuous maturity of new technologies such as big data, cloud computing, and the Internet have led to a fundamental change in the objects, themes, and methods of translation. Translation is no longer limited to individual behavior or a single text, and has been extended to accept and manage the localization of translation projects and products, combined with translation technology, to form large-scale language service activities. With the innovation of information technology, more computer-assisted translation systems have emerged, which has improved the quantity and quality of translations. In order to achieve better development, the workers of English translation of scientific and technological classics and the participants of related research need to continuously strengthen the improvement of information technology capabilities.

2. Overview of Computer-aided Translation
In the development of electronic sci-tech book materials and translation processes, computer-assisted translation has been continuously developed due to the beginning of machine translation. Computer-aided translation of scientific and technological classics is developed through machine translation. Computer-aided Translation, Referred to as CAT. After the translation materials became electronic, computer-aided translation technology was developed. Its content includes original text analysis and
processing technology, translation memory technology, and desktop publishing technology. In a narrow sense, CAT technology refers to specialized software technology related to the translation process. Computer-assisted translation of scientific and technological classics to a certain extent has made up for the quality of machine translation. Computer-aided translation of computer-aided science and technology classics uses computer software technology to realize automatic translation, ensure the quality of manual translation, improve the efficiency of manual translation, and achieve good management of the translation process. English translation of computer-aided science and technology classics mainly includes translation memory, terminology management, alignment tools, and translation project management. Its core technology is translation memory, which works by comparing the fragments to be translated with existing fragments in the database to provide translators with translations with a high matching rate. Multilingual-based translation terminology comparison table is included in the terminology management system, thereby reducing the cost of localization and translation. Computer-assisted translation of scientific and technological classics has many advantages, and is highly valued in related fields.

3. Computer-aided translation technology
Translation memory technology mainly completes the storage of original text and translation. The translation memory technology stores the original text and the translation into a memory bank in contrast. When translating a new original text, it can be reused. The higher the repetition rate of text content, the more obvious the advantages of translation memory. The translation memory compares the fragments to be translated with existing translation fragments in the database, and provides the English translation staff of science and technology classics with translation fragments that meet the matching rate standard. The translation staff adopts, edits, or discards them.

Translation memory is mainly divided into two parts: determining candidate sets and calculating similarity. Due to the large corpus size, an inverted index of the translation memory is used to determine the candidate set, and then the similarity is calculated to select the translation unit. A statistical-based fuzzy translation memory method, incorporating statistical machine translation technology into the translation results, retaining as many high-quality translation fragments as possible in the memory, and performing statistical-based machine translation operations on fragments that cannot be successfully matched.

Statistics-based fuzzy translation memory edits character-based similarity calculation methods to retrieve results from translation memories and enters them into statistical machine translation systems as reference answers. By modifying the phrase-based statistical machine translation, the correct translation part is retained as much as possible, combined with other model scores, the incorrect translation result is corrected, and the translation quality is improved.

In the design of computer-aided English translation memory technology for technical classics, the division of the statistical fuzzy translation memory system can be performed according to the dashed line in Figure 1, which is mainly divided into two modules: translation memory fuzzy retrieval and improved statistical machine translation. The architecture of a fuzzy translation memory system based on statistics is shown in Figure 1 below.

![Figure 1. Statistics-based fuzzy translation memory system architecture](image-url)
4. Terminology Management Technology for English Translation of Computer-aided Classics

Terminology management is a systematic collection of descriptions that process records, store presentations, and query specialized vocabulary in specific areas of expertise. Based on the characteristics of modern translation projects, modern term management was decided with the help of advanced term management technology. The terminology management technology tool completes the management of terminology data, stores, processes, uses, and services the terminology data, and makes use of terminology data in the field of computer-assisted English translation of science and technology classics. The modern terminology management system can help English translators of scientific and technical classics to complete terminological-related tasks, realize the rapid creation of termbases, fast searches, real-time updates and sharing, and ensure the consistency of terminology usage, thereby improving translation efficiency. Terminology management systems are roughly divided into three categories: terminology data management systems, terminology management modules for computer-aided translation tools, and independent terminology management systems. The terminology management of English translation of computer-aided science and technology classics is divided into terminology management before translation, terminology management in translation and terminology management after translation.

The existing computer-assisted technical translation of English translation term extraction technology includes the following: 1) dictionary-based extraction technology, matching entries in professional dictionaries to complete extraction; 2) statistics-based extraction technology, using the statistical properties of terms, Terminology identification and extraction; 3) Linguistic-based extraction technology, using term context information and internal composition information of terms to identify and extract terms; 4) Machine learning-based extraction technology, such as using hidden Markov models to achieve Terms are identified and extracted. In order to maximize the effect of term extraction, multiple techniques are often used to complete the term extraction. After the term extraction is completed, a termbase needs to be created, terminology extraction and termbase creation are completed before translation, and terminology sharing is implemented to avoid the occurrence of inconsistent terminology during translation.

Terminology Management in Translation. In the implementation process of English translation of science and technology classics, the term recognition module in computer-aided science and technology classics English translation uses data structures to complete term matching and labeling, and provide translators with term translations to improve translation efficiency. After identifying new terms in the translation process, we will discuss the newly identified terms in a timely manner, determine acceptable translations, fill in the term details, add them to the termbase, complete the sharing of new terms, and realize the synchronous update of terms in translation collaboration.

Terminology management after translation. After the translation project is completed, the termbase should be reviewed and backed up again.

| Before translation | Translating | After translation |
|--------------------|-------------|------------------|
| Terminology extraction technology | Add terminology technology | Terminology Maintenance Technology |
| Format conversion technology | Terminology recognition technology | Terminology backup technology |

Figure 2. Flow chart of terminology management technology application

5. Computer Aided Science English Translation Scoring Design

The computer-assisted technical classics English translation intelligent scoring system modules mainly include: natural language processing and interaction module, translation scoring module, model file module and result feedback module.

English translation scoring system based on natural language processing to achieve effective communication between people and computers. A comprehensive analysis of the modules of each level of the scoring system shows the overall architecture of the English translation scoring system for science and technology classics based on natural language processing as shown in Figure 3 below. The user
uploads the translation through the user terminal, and after computer natural language processing and information interaction, enters it into the English translation scoring model of scientific and technological classics, uses the model file to complete the scoring, and returns the scoring and evaluation results to the user terminal; the manual review and evaluation terminal can use information interaction. The module consults the English translation files of user science and technology classics, selects the file conversion format and the training model translation set data to improve the overall effect of the scoring module.

![Figure 3. Overall architecture of translation scoring system](image)

6. Conclusion
From the perspective of current cultural dissemination, scientific and technological classics are an important part of Chinese cultural classics. The interpretation and translation of cultural elements are of great significance. In the process of translating scientific and technical books into English, different methods can be used. This paper uses developed computer technology, combined with the currently mature natural language processing technology, to study the computer-aided translation of scientific and technical classics based on natural language processing, to improve the efficiency of English translation of scientific and technological classics while ensuring the quality of translation.

References
[1] Xu Mingwu, Wang Yanmo. Research on English Translation of Chinese Science and Technology Books (1997-2016): Achievements, Problems and Suggestions[J]. Chinese Foreign Languages, 2017 (2): 96-103.
[2] QU Weiguang, ZHOU Junsheng, WU Xiaodong, et al. Survey on abstract meaning representation [J]. Journal of data acquisition & processing, 2017, 32(1):26-36.
[3] Wang Huashu, Leng Bingbing, Cui Qiliang. Further Research on Applied Translation Research System in the Information Age. Shanghai Translator, 2013(1): 9.
[4] BOWKER L. Computer-aided translation technology: a practical introduction[M]. Ottawa: University of Ottawa Press, 2002.
[5] Ma Yanping, Liu Yingli. Discussion on the construction scheme of computer-aided translation memory and service system for petroleum equipment scientific and technological literature[J]. Shandong Chemical Industry, 2016, 45 (20): 135-136.
[6] CHEN Zhiming, LI Maoxi, WANG Mingwen. Sentence-level machine translation quality estimation based on neural network features [J]. Journal of computer research and development, 2017, 54(8):1804-1812.