

# Research on Big Data Empowerment of Modern Translation Technology in the Context of New Liberal Arts

Yanan Wu

*Russian Center of Sanya University, Sanya, 572022, China*

**Abstract:** *This article aims to explore the application of big data technology in modern translation, with the background of a new liberal arts course. This article analyzes the challenges and opportunities that big data technology brings to translation, and discusses possible methods and approaches to integrating big data into translation practice. Through comprehensive literature review and case analysis, this article concludes that big data technology has great potential to promote the development of modern translation, and recommends that more attention be paid to and efforts be made to explore and utilize the role of big data in translation research and practice.*

**Keywords:** *Big Data, Modern Translation, New Liberal Arts Courses, Data Driven Translation, Machine Translation, Human-computer Collaboration*

## 1. Introduction

The rise of big data technology has completely changed many industries and fields, and translation is no exception. With the exponential growth of digital data and the progress of machine learning algorithms, big data technology has created new opportunities and challenges for translation practice and research. In the context of a new liberal arts curriculum that emphasizes interdisciplinary and data driven education, exploring the application of big data technology in translation and promoting the integration of translation and big data is of great significance. Against this background, this article aims to answer the following research questions: What are the challenges and opportunities that big data technology brings to modern translation? How to integrate big data technology into translation practice and research? What is the impact of big data technology on the future development of translation?

## 2. Overview of big data technology and translation

Big data refers to large-scale and complex data sets generated and collected from various sources, including social media, sensors, and digital devices. Big data technology includes data mining, machine learning, and natural language processing, which can analyze and process large-scale data sets with high speed, diversity, and authenticity.

In the field of translation, big data technology has been applied in various ways, such as machine translation, corpus based translation, and term extraction. Machine translation relies on statistical or neural machine learning models to automatically translate text. The availability of large-scale parallel corpora and advanced machine learning algorithms greatly improves the effectiveness of machine translation. Corpus based translation involves the analysis and comparison of large-scale multilingual corpora, which can help identify translation patterns, improve translation quality, and optimize translation workflow. Term extraction is a technique for extracting relevant terms and concepts from a large-scale corpus, which can help translators identify key concepts and terms in a specific field.

The translation technology enabled by big data faces the following challenges and opportunities:

### 2.1. Data privacy and security

In the process of big data driven translation, it is necessary to process a large amount of user data and sensitive information, such as user evaluations, medical literature, and trade secrets. The privacy and security of these data is an important challenge. When processing these data, it is necessary to

adopt safe and reliable data storage, transmission, and processing methods to protect the privacy and sensitive information of users. In addition, it is also necessary to comply with data protection regulations and standards, such as GDPR in Europe and HIPAA in the United States, to ensure the legality and security of data.

## ***2.2. Data quality and accuracy***

In the process of big data driven translation, it is necessary to handle large-scale and diverse language and text data. The quality and accuracy of these data are important factors in the quality of translation. Data quality and accuracy issues include data cleansing, deduplication, labeling, and parity issues. When processing these data, it is necessary to adopt appropriate data cleaning and preprocessing techniques to improve data quality and accuracy.

## ***2.3. Multi language and multi domain data mining***

In the process of big data driven translation, data from multiple languages and domains needs to be processed. The mining and analysis of these data needs to adapt to the characteristics and laws of different languages and domains. For example, there are differences in the grammar and vocabulary usage of different languages, as well as differences in terms and languages in different fields. Therefore, different data mining and analysis technologies need to be adopted to adapt to data in different languages and fields.

## ***2.4. Translation quality and cultural adaptability***

In the process of big data driven translation, it is necessary to achieve high-quality and culturally adaptive translation results. This requires consideration of the use and dissemination patterns of different languages and cultures, as well as the subjectivity and cultural characteristics of translation. When dealing with large-scale multilingual corpora, it is necessary to adopt appropriate language and cultural adaptation techniques to improve translation quality and cultural adaptability.

## ***2.5. Combination of human translation and machine translation***

In the process of big data driven translation, both human translation and machine translation play an important role. Artificial translation has the flexibility and subjectivity to handle some difficult issues in machine translation, such as the cultural and rhetorical characteristics of language. Machine translation is efficient and accurate, and can handle large-scale language and text data. The combination of human translation and machine translation can achieve better translation results and improve the efficiency and quality of translation.

## ***2.6. Construction and management of corpus***

Big data driven translation requires the support and construction of a large-scale multilingual corpus. The construction and management of corpus requires a lot of time and effort, including the collection, collation, labeling, and storage of corpus. When processing corpus, it is necessary to adopt appropriate data mining and analysis techniques to improve the quality and accuracy of data. In addition, issues such as copyright and usage rights of the corpus need to be considered to ensure the legitimacy and availability of the corpus.

## **3. Methods and approaches for big data driven translation**

In order to integrate big data technology into translation practice and research, the following methods and approaches are proposed:

### ***3.1. Utilizing a combination of machine translation and human translation***

Machine translation is one of the key technologies for big data driven translation. Machine translation systems can utilize large-scale corpus and machine learning algorithms to achieve automated translation and improve translation efficiency and quality.[1] However, the quality of machine translation needs to be improved, especially when dealing with complex linguistic and cultural

issues. Therefore, a combination of human translation and machine translation can be used to fully utilize the advantages and opportunities of both. For example, a machine translation system can be used for preliminary translation, followed by manual translation for later modification and polishing to improve translation quality and cultural adaptability.

### ***3.2. Utilizing a large-scale multilingual corpus***

Large multilingual corpora are important resources for big data driven translation. These corpora can include various forms of text and voice data, such as news reports, social media, e-books, and audiovisual materials. Using these corpora, translation can be automated and intelligent, and translation efficiency and quality can be improved. In addition, these corpora can also be used for language and cultural research, in-depth understanding of the use and dissemination patterns of different languages and cultures to improve cultural adaptability and translation quality[2].

### ***3.3. Utilizing natural language processing technology***

Natural language processing technology is an important support for big data driven translation. Natural language processing technology can handle various issues of natural language texts, such as the grammar and semantics of language, the recognition and extraction of terms and phrases, and the recognition and analysis of emotions and emotions. We can use natural language processing technology, translation tasks such as automated translation, terminology management, and quality control can be realized to improve translation efficiency and accuracy.

### ***3.4. Utilizing artificial intelligence technology***

Artificial intelligence technology is the future trend of big data driven translation. Artificial intelligence technology can achieve higher level translation tasks, such as robot translation and automatic text generation. For example, generating adversarial networks (GANs) can be used to achieve automatic text generation and generate high-quality translation results. Using deep learning and neural network technology we can achieve more accurate language and cultural adaptability, and improve translation quality and accuracy[3].

### ***3.5. Utilizing data visualization technology***

Data visualization technology can help translators better understand and process large-scale and diverse data. Data visualization technology can convert data into visualized graphs or charts for users to better understand and analyze data. During the translation process, data visualization technology can be used to achieve tasks such as terminology management, data mining, and analysis to improve translation efficiency and quality.

### ***3.6. Utilizing man-machine collaboration***

Human-computer collaboration can combine the advantages of human translation and machine translation systems to achieve better translation results and further improve the efficiency and quality of translation. Human-computer collaboration can achieve a combination of automated translation and manual translation to fully utilize the advantages and opportunities of both. For example, a machine translation system can be used for preliminary translation, followed by manual translation for later modification and polishing to improve translation quality and cultural adaptability.

## **4. Discussion on application examples of translation technology driven by big data**

### **Case 1: Translation of online user evaluations**

In this case study, a translation company used big data technology to translate a large-scale online user evaluation dataset for a multinational e-commerce company. The dataset includes over 1 million evaluations in multiple languages, such as English, Chinese, Spanish, and French. To translate these evaluations, the company used a combination of machine translation and manual post editing. The machine translation system is trained based on a large-scale user evaluation parallel corpus, and uses fine-tuning for specific domains and language pairs. Manual post editors are responsible for correcting errors and improving fluency and naturalness.

The use of big data technology enables the company to efficiently and accurately process large and diverse data sets. Machine translation systems provide basic translation quality faster and cheaper than human translation. Manual post editors can focus on more complex and subjective aspects of translation, such as style, intonation, and cultural adaptability. The use of big data technology also enables the company to continuously improve machine translation systems by collecting and analyzing feedback and post editing data.

#### Case 2: Translation of medical documents

In this case study, a medical translation team used big data technology to translate a large-scale medical document corpus for a global pharmaceutical company. The corpus includes various types of medical documents, such as clinical trial reports, drug labels, and patient information. The translation team used machine translation and manual post editing methods to translate the corpus. The machine translation system is based on large-scale medical literature corpus training, and uses fine-tuning of specific fields and language pairs. Manual post editors are responsible for correcting errors and improving translation quality.

The use of big data technology enables the team to efficiently process large and complex medical texts. Machine translation systems can quickly generate basic translation results, reducing the time and cost of manual translation. Artificial post editors can focus on more complex and critical aspects of translation, such as medical terminology and cultural adaptability. The use of big data technology also enables the team to leverage the diversity and richness of medical texts to improve translation quality and accuracy by analyzing and comparing texts of different types and languages.

#### Case 3: Translation of cross-cultural communication

In this case study, an international organization used big data technology to translate online content in multiple languages for its cross-cultural communication activities. These include articles, blogs, and social media posts designed to spread the organization's philosophy and values. To translate these content, the organization used a data driven translation approach, analyzing large-scale online texts in multiple languages, and combining translation results with localization and cultural adaptability. The organization also uses a combination of machine translation and human translation to improve translation quality and efficiency.

The use of big data technology enables the organization to gain an in-depth understanding of the use and dissemination patterns of different languages and cultures. By analyzing and comparing online texts in multiple languages and domains, the organization can improve translation quality and cultural adaptability. The use of big data technology also enables the organization to achieve real-time translation of online content in multiple languages and cultures, continuously improving translation quality and effectiveness by analyzing user feedback and data.

The case study described in this article shows that big data technology has a potential impact on the efficiency, quality, and innovation of modern translation. Big data technology provides new opportunities for translation research and practice, which can help translators better understand and process large-scale and diverse language and text data. By analyzing and utilizing big data, translators can gain an in-depth understanding of the use and dissemination patterns of different languages and cultures to improve translation quality and cultural adaptability.

In addition, future translation research and practice will increasingly focus on human-computer collaboration and the development and application of intelligent translation tools. Human-computer collaboration can combine the advantages of human translation and machine translation systems to achieve better translation results and further improve the efficiency and quality of translation. Intelligent translation tools can improve the efficiency and accuracy of translation through automation and intelligence, and gradually achieve higher level translation tasks, such as robot translation and automatic text generation.

## 5. Conclusion

This article aims to explore the application and impact of big data technology in modern translation. By analyzing the challenges and opportunities in translation, it introduces the application and impact of big data technology in translation, and expounds its practical application and effects through case studies. The analysis and discussion in this article indicate that big data technology has a potential impact and role on the efficiency, quality, and innovation of modern translation.

However, the application and development of big data technology still faces some challenges and problems. For example, issues such as data privacy and security, data quality and accuracy, and data mining in multiple languages and domains need to be further explored and resolved. In addition, the application and impact of big data technology also requires consideration of issues such as human values, cultural differences, and social impact, in order to achieve sustainable and responsible translation practices.

### **Acknowledgement**

**Fund Project:** This article is the Research Result of the 2022 Collaborative Education Project of the Ministry of Education, "Research on Big Data Empowerment of Modern Translation Technology in the Context of New Liberal Arts", Project Number: 221005618134159.

### **References**

- [1] Bi H, Kuang W, Bi C. *Translation opportunities brought by cloud computing and big data [J]. Journal of Tangshan Normal University 2021; 43(03):87-89.*
- [2] Liu H, Xu Q. *Research on the development of translation technology in the context of big data era [J]. Information recording materials. 2019; 20(10):137-138.*
- [3] Wang H. *The development and enlightenment of translation technology in the era of big data [J]. Oriental translation 2016; (04):18-20.*