

# Post-Editing Research in China (2007–2024) - A Visualization Analysis Using VOSviewer

Jing Li, Rong Chen\*

School of Foreign Studies, Xi'an University of Posts and Telecommunications, Xi'an, Shaanxi, 710121, China

## ABSTRACT

In recent years, with the widespread adoption of neural machine translation (NMT) technology, post-editing of machine translation (MTPE) has emerged as a prominent topic in translation technology research in China. This study conducts a bibliometric and visualization analysis using VOSviewer, based on literature related to MTPE published in China National Knowledge Infrastructure (CNKI) from 2007 to 2024. By examining patterns of author collaboration and keyword co-occurrence, the study outlines the current state of Chinese research on MTPE. The findings reveal that although research on MTPE in China began relatively late, it has gradually gained academic attention. However, the degree of collaboration among researchers remains low, and interdisciplinary research is not yet well developed. Moreover, the research focus tends to be narrow, with limited application domains. Based on these observations, the paper also offers a prospective outlook on future directions in the field.

## KEYWORDS

Post-editing; VOSviewer; Research status

## 1 Introduction

In the context of globalization and localization, the demand for high-quality, efficient translation has grown, making post-editing (PE) a vital method for enhancing machine translation (MT) output. With AI-integrated MT systems like Google Neural Machine Translation and Baidu Translate boosting efficiency, post-editing has become key to ensuring translation quality and addressing diverse needs. While PE research began internationally in the mid-1980s, China's academic focus on the topic emerged later. The earliest domestic study appeared in 1995, but significant scholarly attention started in the early 2000s, marked by Wei Changhong and Zhang Chunbai's 2007 article. Since then, interest has grown, prompting questions about the current state, key themes, and trends in PE research.

This study uses bibliometric analysis and the visualization tool VOSviewer to map China's post-editing research landscape, aiming to identify its current status, highlight research gaps, and offer insights for future studies.

## 2 Theoretical framework

### 2.1 Machine translation

Machine Translation (MT) is the automated translation of text using computer technology, with origins dating back to the 1950s. It has evolved from rule-based and statistical methods to today's dominant neural machine translation (NMT) powered by deep learning<sup>[1]</sup>. The introduction of the Transformer architecture has enabled NMT systems to produce more fluent, natural translations by learning semantic and syntactic patterns. Notably, Google's 2016 neural MT system marked a breakthrough in translation quality across language pairs.

However, challenges remain in applying MT to formal and professional contexts. Issues such as semantic ambiguity and cultural nuances often result in inaccuracies, and effective post-editing still demands specialized human expertise<sup>[2]</sup>.

### 2.2 Post-editing

Post-editing (PE) involves manually refining machine translation outputs to improve accuracy, fluency, and readability. Beyond fixing surface errors, it addresses semantic integrity, coherence, style, and cultural relevance.

With advances in neural machine translation, the demand for PE has grown steadily. The Blue Book of China's Language Service Industry (2023) identifies PE as a crucial part of the translation service value chain, now widely offered by translation companies. Reviewing PE research helps track industry trends, anticipate technological shifts, and support market and talent development. In translation education, such reviews can guide curriculum design and training to prepare qualified PE professionals.

---

\* **Corresponding Author:** Rong Chen, chenrong@xupt.edu.cn

### 3 Methodology

#### 3.1 Data sources

The data for this study were collected from academic journal articles indexed in the China National Knowledge Infrastructure (CNKI). A precise search was conducted using “译后编辑” (post-editing) as both the subject and keyword, covering the period from 2007 to the present. To ensure the accuracy and relevance of the dataset, the retrieved results were manually reviewed and screened. Duplicate entries, conference presentations (PPTs), standards, interviews, and documents in non-English languages that were unrelated to the research topic were excluded. After this filtering process, a total of 166 valid articles were retained for analysis.

#### 3.2 Research methods

Knowledge mapping, developed on the basis of citation analysis and information visualization technology, employs tools such as SPSS, CiteSpace, and VOSviewer. This study adopts bibliometric methods in combination with VOSviewer for visualization analysis. Developed by Van Eck and Waltman from the Centre for Science and Technology Studies at Leiden University, VOSviewer is a software tool for bibliometric and knowledge visualization analysis. It offers powerful capabilities for generating cluster, label, and density views to explore and interpret research patterns<sup>[3]</sup>. It is particularly effective in keyword analysis, cluster analysis, thematic extraction, and author profiling<sup>[4]</sup>, and can process large datasets while clearly presenting clustered network structures and node relationships.

Accordingly, this study uses VOSviewer 1.6.20 to perform visual analyses of author collaboration and keyword co-occurrence, aiming to objectively summarize the characteristics of research across different stages and to reveal research perspectives and emerging trends in the field.

### 4 Results and discussion

#### 4.1 Analysis of annual publication volume

The number of publications per year in a given knowledge domain is a key indicator of scientific progress. It can reflect the developmental stage, current status, and growth trends of a research field. Based on the valid documents obtained from CNKI, the annual publication data are plotted as a line chart (see Figure 1), with the horizontal axis representing the years (2007–2024) and the vertical axis indicating the number of publications per year.

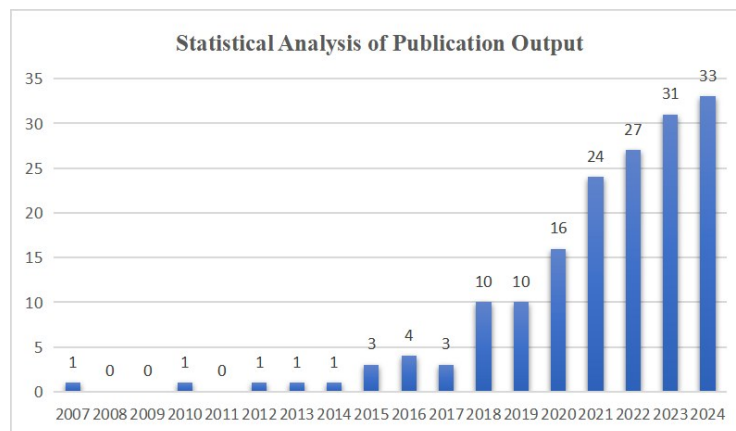


Figure 1 Statistical analysis of publication output

Figure 1 shows a clear upward trend in domestic post-editing (PE) publications from 2007 to 2020, with a sharp increase after 2020. The development of PE research in China can be divided into four stages: the initial stage (2007–2014), slow-growth (2015–2017), rapid-growth (2018–2022), and steady-growth (2023–2024).

From 2007 to 2014, PE studies remained limited, with few annual publications, reflecting its low recognition at the time. Between 2015 and 2017, research output grew modestly as the rise of machine translation (MT) highlighted the need for quality improvements. A major surge began in 2018, with publications peaking at 27 in 2022, driven by the expanding use of MT, market demand for high-quality translation, and supportive national policies—such as Xi Jinping’s 2018 call to enhance China’s global communication efforts.

Since 2023, PE has matured as a research area, gaining traction among scholars and institutions. Continued advances in large language models suggest that PE will remain a vital and evolving field in the future.

## 4.2 Analysis of major authors

Analyzing co-authorship reveals key researchers and collaboration patterns in post-editing (PE) research. Using VOSviewer on CNKI data (with a minimum of two publications per author), only 20 out of 267 authors met the threshold. The co-authorship map (Figure 2) shows limited collaboration, with small clusters around authors such as Feng Quangong, Cui Qiliang, Zhang Huiyu, Wang Xiangling, Jia Yanfang, Yin Jian, and Qi Guanhua.

Overall, author distribution is fragmented, indicating weak collaboration and limited interdisciplinary engagement. To further develop PE research in China, stronger academic cooperation and integration of diverse disciplinary perspectives are needed.

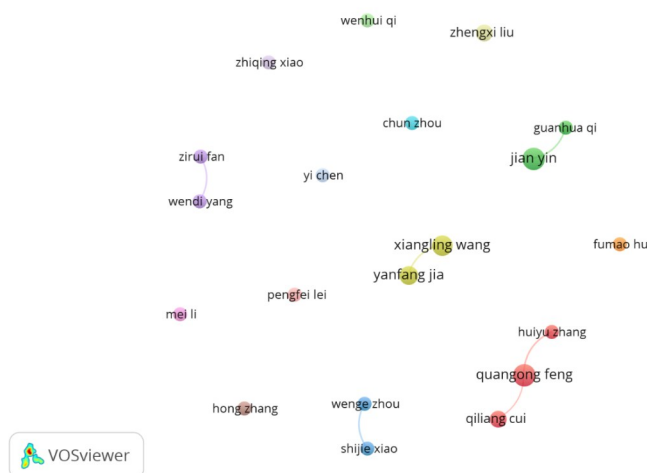


Figure 2 Co-authorship network of major authors

## 4.3 Keyword clustering analysis

"Keywords serve as a condensed summary of a paper's content. If a particular keyword repeatedly appears in a set of documents over a given period, it indicates that the associated research topic is a hotspot within the field" [5]. In this study, the valid documents retrieved from CNKI were imported into VOSviewer for keyword clustering analysis. A total of 305 keywords were extracted. After setting the minimum occurrence threshold to 2, only 49 keywords met the criterion.

To enhance the accuracy of the results, these 49 keywords were manually refined. Terms with low thematic relevance—such as "DeepL," "启示" (inspiration), and "目的论" (skopos theory)—were removed. Additionally, redundant or semantically overlapping terms were merged, retaining only the most representative ones (e.g., "译后编辑" [post-editing] was kept, while "机器翻译译后编辑" [machine translation post-editing] was excluded). After preprocessing, a keyword co-occurrence network was generated, as shown in Figure 3.

In the co-occurrence network, the size of each node reflects the frequency of the keyword—larger circles indicate higher frequencies. The lines between nodes represent co-occurrence relationships; thicker lines signify stronger associations between keywords. Different colors denote different thematic clusters. As illustrated in Figure 3, post-editing research in China is primarily organized into five major thematic clusters.

**Thematic Cluster 1: Post-Editing Competence and the Application of AI Technologies (Red)**

This cluster primarily includes keywords such as post-editing competence, ChatGPT, translation competence, post-editing course, and translation quality. The focus is on post-editing competence and the integration of artificial intelligence technologies. Post-editing competence refers to the knowledge base and cognitive literacy required to revise raw machine translation output. It involves three interrelated dimensions: cognitive, knowledge-based, and skill-based. These dimensions are coordinated through cognitive mechanisms, enabling the transformation of knowledge into practical skills [6]. The post-editing process involves minimal human modification of machine-generated translations to improve accuracy and readability, primarily addressing linguistic and formatting errors, and polishing the text to conform to the conventions and context of the target language.

The development of AI—particularly deep learning-based machine translation—has significantly improved translation efficiency but has also presented new challenges for post-editing. In response, Qi Guanhua and Yin Jian (2024) emphasized the importance of general translation literacy in guiding effective post-editing with ChatGPT and provided case studies demonstrating how human-machine collaboration can enhance translation quality [7]. Other scholars have proposed new models of translator competence. Hu Minyu and Yin Jian (2024) argued that, under a human-machine collaborative translation paradigm, post-editors should possess dynamic awareness of model ideology, ethics, and

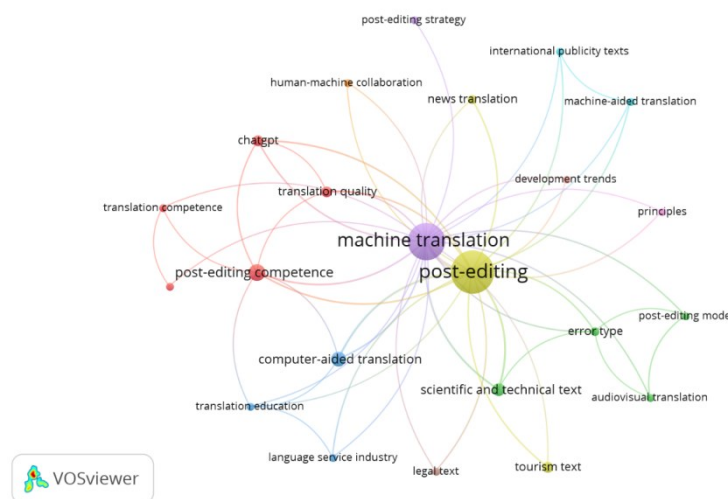


Figure 3 Keyword co-occurrence knowledge map

cultural bias, along with the ability to express needs and correct errors—thus promoting a redefined competence model for post-editing <sup>[8]</sup>.

As AI technologies continue to evolve, so too must post-editing competence. Even MTI (Master of Translation and Interpreting) training models need to adapt. Post-editing courses should not only meet the demands of the language service industry but also enhance the professional competitiveness of translation graduates <sup>[9]</sup>.

#### Thematic Cluster 2: Industry Development and Translation Education (Blue)

This cluster includes keywords such as language service industry, computer-assisted translation, and translation education. The focus is on the evolution of the translation industry and its educational implications. Scholars in this area have examined the current landscape of the translation market. For example, Wang Huashu and Chen Nie'ao (2021) conducted a survey of 218 Chinese language service providers, analyzing the market scale, usage frequency, application scenarios, investment levels, and expectations regarding machine translation and post-editing. Their findings indicate key expectations from MT systems: greater intelligence, specialization, integration, interactivity, and user-friendliness. They proposed recommendations from three perspectives—industry, academia, and practitioners—to promote the integration of MT and post-editing, covering areas such as technological innovation, awareness, and talent development <sup>[10]</sup>.

From an educational standpoint, however, the curriculum remains misaligned, with a notable imbalance between theory and practice. Most MTI programs in Chinese universities currently lack dedicated training in post-editing competence. Institutions should respond to the changing landscape by incorporating post-editing skills into training frameworks and fostering collaborations among universities, enterprises, and research institutions to explore synergistic innovation through government–industry–academia partnerships.

#### Thematic Cluster 3: Post-Editing in Scientific and Audiovisual Texts (Green)

This cluster centers on keywords such as scientific and technical text, audiovisual translation, post-editing model, and error type. The focus is on applying post-editing practices to scientific and audiovisual content. Both domains contain common translation errors that require post-editing intervention. For scientific texts, typical issues include mistranslation of terminology and proper nouns, lexical inaccuracies, and syntactic errors <sup>[11]</sup>. In audiovisual translation, common mistakes include mistranslated idioms, omissions of cultural elements, and missing conjunctions or subjects <sup>[12]</sup>.

In response, scholars have proposed specific post-editing models for each text type. For example, when dealing with culturally embedded terms that are difficult to translate directly, a combined translation + annotation strategy is often employed to improve comprehension. Such measures have proven effective in enhancing reader understanding and engagement.

#### Thematic Cluster 4: Machine Translation and Post-Editing Strategies (Purple)

This cluster is built around two core keywords: machine translation and post-editing strategy. The discussion here takes a macro-level approach to examining the interplay between MT and PE. Machine translation excels in efficiency and scalability, allowing rapid processing of large volumes of text. However, it often lacks precision, fluency, and idiomatic expression, and struggles with domain-specific terminology, cultural nuances, and stylistic coherence.

Post-editing thus emerges as a corrective mechanism to address these shortcomings. At the macro level, post-editing strategies are typically explored across lexical, syntactic, and discursive dimensions. However, current research remains limited in scope, offering only partial solutions that may not fully capture the complexity of real-world translation challenges. Consequently, the practical value and generalizability of these strategies are often constrained.

#### Thematic Cluster 5: Post-Editing in Domain-Specific Texts (Yellow)

This cluster focuses on keywords such as post-editing, news translation, and tourism text, emphasizing the application of post-editing in specific content domains. News translation, characterized by high timeliness, accuracy, objectivity, and clarity, places stringent demands on post-editing quality. One key aspect is pre-editing—conducted before translation—since news is often republished without the opportunity for downstream revision. Pre-editing strategies tailored to sentence structure and information flow are therefore crucial. For example, inserting or removing modifiers based on sentence length and structure enhances the readability and conciseness of the final translation<sup>[13]</sup>.

Tourism texts, by contrast, are often vivid and information-dense, including elements such as scenic descriptions, itineraries, and travel guidance. Here, readability, coherence, and cultural adaptation are critical. Such texts frequently contain cultural-specific references (e. g., place names, customs, and folklore) that require not only accurate comprehension but also culturally appropriate rendering in the target language. While domestic research on tourism-related post-editing has grown, it still lacks sufficient focus on professional standards and real-world applicability, often failing to address the genre-specific needs of tourism discourse.

## 5 Conclusion

This study used VOSviewer to analyze 166 post-editing (PE) journal articles from CNKI (2007–2024), focusing on publication trends, author collaboration, and keyword co-occurrence.

PE research in China began in 2007 and has progressed through four stages: initial (2007–2014), slow-growth (2015–2017), rapid-growth (2018–2022), and steady-growth (2023–2024). Though initially slow to develop, interest has risen due to increased translation demand and the widespread use of machine translation.

Future research should expand into literary PE and prioritize developing intelligent PE tools using deep learning to enhance MT correction and reduce human workload. Additionally, more attention is needed on human–computer interaction and ethical issues in PE. As Wang Xiangling and Jia Yanfang (2018) note, human–machine interaction is central to PE<sup>[14]</sup>, and the broader human relationships involved warrant careful ethical consideration in the digital age.

## Disclosure Statement

The authors declare no conflict of interest.

## References

- [1] Zhang Y, 2023, A Review of the Past Decade in Machine Translation: A Visualization Analysis Based on CiteSpace. *Modern Linguistics*, 11(3): 1171–1176.
- [2] Li X, Gao CY, 2024, Knowledge Mapping and Development Trends in Foreign Machine Translation Research. *Shanghai Journal of Translators*, 175 (2): 41–47.
- [3] Gao K, 2015, Application Study of Bibliometric Analysis Software VOSviewer. *Sci-Tech Information Development & Economy*, 25(12): 95–98.
- [4] Zong QJ, Yuan QJ, Shen HZ, et al., 2011, Research Hotspots in Chinese Information Science in 2010 from a Knowledge Mapping Perspective—Part I of a Study on Contemporary Disciplinary Trends Based on Knowledge Mapping. *Journal of Information*, 30(12): 48–53.
- [5] Diao H, 2017, A Review of Domestic Research on Translation Technology. *Journal of Beijing International Studies University*, 39(6): 69–81, 125.
- [6] Feng QG, Liu M, 2018, A Three-Dimensional Model of Post-Editing Competence. *Foreign Language World*, 186 (3): 55–61.
- [7] Qi GH, Yin J, 2024, Research on ChatGPT Post-Editing Instructions Based on Human-Machine Coupling. *English Square*, 2024(23): 44–47.
- [8] Hu MY, Yin J, 2024, Exploring Post-Editing Competence for Machine Translation Based on ChatGPT. *English Square*, 2024(20): 43–46.
- [9] Chen Y, Pan LF, 2023, Research on the Curriculum Design of Post-Editing for Machine Translation: Based on the Analysis of Needs, Competence, and Content Elements. *Journal of Hunan University of Science and Engineering*, 44(4): 121–124.
- [10] Wang HS, Chen NA, 2021, A Survey on the Application of Machine Translation and Post-Editing in Chinese Language Service Enterprises. *Journal of Beijing International Studies University*, 43(5): 23–37.
- [11] Ming SY, Ma HJ, 2022, A Preliminary Study of Post-Editing in English-Chinese Scientific and Technical Text Translation. *Taste -Classics*, 2022 (12): 32–34.
- [12] Li W, 2019, A Study on Post-Editing Models in Audiovisual Text Translation. *Journal of Shijiazhuang Tiedao University (Social Science Edition)*, 13(4): 29–37.
- [13] Feng QG, Li JW, 2016, A Study on Post-Editing Models in News Translation. *Computer-Assisted Foreign Language Education*, 2016(6): 74–79.
- [14] Wang XL, Jia YF, 2018, Empirical Studies on Machine Translation Post-Editing Abroad in the 21st Century. *Journal of Hunan University (Social Sciences)*, 32(2): 82–87.