International Journal of Education and Humanities (IJEH), 4(2) 2024:137-146
http://i-jeh.com/index.php/ijeh/index
E-ISSN: 2798-5768

A Review of Research in Translation Technology Based on Citespace

Ding Lijie[1]

Abstract

To explore the research hotspots and trends in translation technology, this study utilizes the literature visualization analysis software CiteSpace to conduct a bibliometric analysis of relevant literature on translation technology from 2013 to 2023 in the Web of Science and China National Knowledge Infrastructure. This research indicates a continuous increase in annual publications from the Web of Science, while the publications from China National Knowledge Infrastructure show no clear upward or downward pattern over the past 10 years. This study also reveals that scholars in the field of translation technology mainly publish papers independently. Most importantly, this study discovers that machine translation, translation technology, artificial intelligence, computer-assisted translation, post-translation editing, neural machine translation, and talent training emerge as research hotspots in the field. Post-editing, google translate and natural language processing are the research frontiers in the field of machine translation in the last three years. This study can provide scholars in the field with the latest hotspots and frontiers so that they can conduct more innovative research.

Keywords: *Translation technology, Machine translation, CiteSpace*

A. Introduction

Translation technology is a rapidly developing field that makes use of computer software and other tools to advance the translation process. Machine translation (MT) falls within the realm of computational linguistics and is characterized as a system capable of autonomously translating texts from one language to another. (Munkova, Hajek, Munk, & Skalka, 2020). Bowker (2002) pointed out that computer-aided translation technology encompasses all types of computerized tools that translators employ to assist them in their tasks. This study believes that translation technology is a versatile tool employed by translators, playing an empowering role in facilitating the translation process. Due to the limitation of space, this study focuses on the most representative translation technologies such as machine translation technology and computer-aided translation technology.

The Web of Science (WoS) is considered the most reliable independent global citation database, while the China National Knowledge Infrastructure (CNKI) stands as China's largest and most significant database (Zhang et al.,2020). Existing reviews have predominantly concentrated on Chinese literature in machine translation within CNKI (Fan & Yun, 2022; Zhang, 2017; Diao, 2017). However, almost all the papers in CNKI are written by Chinese scholars, and these studies can hardly reflect the research hotspots and frontiers of international translation technology. In addition, some scholars have systematically reviewed the literature on translation studies in WoS (Deng & Yu, 2022; Dong & Chen, 2015; Huang & Liu, 2019; Rovira-Esteva, Orero, & Franco Aixelá, 2015; Van Doorslaer & Gambier, 2015).

[1]Southwest University of Political Science and Law, Chongqing, China, 2532336501@qq.com

However, a gap exists in systematically reviewing and comparing literature on translation technology from both CNKI and WoS. To address this gap, this study used CiteSpace, a scientometric analysis tool, to visualize and examine the knowledge components of translation technology research in China and other countries. The aim is to organize the research network in this discipline, track research trends, and identify leading-edge topics, offering strategic insights for future studies.

This study on translation technology trends and hotspots is fundamentally important for several reasons. Firstly, it serves as a crucial foundation for future research by identifying key areas of focus and emerging trends, thereby guiding scholars and practitioners in developing relevant and impactful research questions. This not only enhances the quality of academic inquiry but also directs resource allocation more effectively. Academic institutions and funding bodies, informed by this study, can make more strategic decisions about where to invest resources to foster innovation. Furthermore, the highlighted trends in areas like machine translation, artificial intelligence, and natural language processing have broader implications, driving advancements that transcend translation and impact global communication and content globalization. This study also facilitates stronger ties between academia and industry, encouraging collaborations that can lead to practical technological innovations. Additionally, by spotlighting the need for talent training, it underscores the importance of adapting educational programs to prepare a workforce skilled in modern technologies and methodologies. The global impact of improved translation technologies cannot be overstated, as they enhance cross-cultural communication and understanding, critical in an increasingly interconnected world. Moreover, with a focus on enhancing translation quality through post-editing and improved tools like Google Translate, the study addresses vital aspects of translation accuracy, which is paramount in avoiding miscommunications in international relations and other crucial areas. Overall, this research is instrumental in shaping the future of translation technology, fostering professional growth, and enhancing global communications, making it a pivotal piece in the advancement of both academic inquiry and practical application in the field.

B. Methods

Data Sources

To ensure the original data's comprehensiveness, accuracy, and high credibility, this study opted for the Social Science Citation Index (SSCI) within the Web of Science Core Collection (WoSCC) to collect English literature and the Chinese Social Science Citation Index (CSSCI) within CNKI for Chinese publications.

In CNKI, the search keywords including “machine translation”, “computer-aided translation” and “translation technology” were used to retrieve the needed documents for the study and the retrieval time was from 2013 to 2023. The language was specified as Chinese. Non-research pieces, like book reviews and article solicitations, along with articles unrelated to the topic, were manually filtered out. A total of 240 Chinese articles were successfully gathered, each providing critical information like the author's name, affiliated institution, keywords, title, abstract, and the year it was published.

In WoS, the search formula was TS = (“translation technology” OR “computer-assisted translation” OR “machine translation” OR “computer-aided translation” OR “neural machine translation”). The search timeframe was defined as “2013 to 2023.” The types of documents were specified as either "article" or "review article," and the language of publication was

designated as “English”. To focus the selection of studies, this research refined the dataset to include categories like linguistics, language linguistics, education and educational research, and computer science artificial intelligence. In the end, 661 references were gathered. Each article was downloaded and stored in TXT format, encompassing information like titles, keywords, authors, abstracts, descriptors, identifiers, among others.

Data Analysis

CiteSpace is a robust tool commonly utilized for visualization and bibliometric analysis. It facilitates bibliographic and citation data sourced from prominent databases like Web of Science, Scopus, and Dimensions (Chen, 2006). CiteSpace software creates knowledge maps that mainly consist of nodes and connections. Here, nodes symbolize analytical elements like authors, institutions, countries, keywords, cited references, and so forth. The connections between nodes indicate relationships of collaboration, co-occurrence, or co-citation. Node and line colors correspond to various years, while the node sizes indicate the frequency with which they occur or are cited.

The operational stages of CiteSpace software include time slicing, thresholding, modeling, pruning, merging, and mapping. This study used the latest version of CiteSpace 6.2.R6 (64-bit) Advanced. In this study, the CiteSpace software parameters were configured as follows: the time slicing covered the years “2013–2023”, with each slice representing one year. This study set the term source to “All selections”, configured the node type option to “Choose 1 at a time”, used the “Top 50 objects” as the selection criteria, applied the “Pathfinder” method for pruning, and displayed the visualization map in “Cluster view-static, show merged network.”

Typically, the Q value (Q) and silhouette value (S) are employed to assess the scientific validity and practicality of visualization knowledge maps. When the Q value exceeds 0.3, the network's cluster structure is considered significantly meaningful. Likewise, an S value above 0.5 indicates that the clustering results are highly reliable (Hu et al., 2020).

C. Findings and Discussion

Analysis of Publication Outputs

Figure 1 illustrates that the annual number of publications from WoS has shown an upward trend over the past decade. Specifically, there was a general upward trend in the annual output related to machine translation research from 2013 to 2023. The research trends can be separated into two phases. During the initial phase spanning from 2013 to 2018, there was a consistent rise in publications related to MT studies. Subsequently, from 2018 onwards, there was a notable and significant increase in the quantity of publications within this particular field. The quantity of articles released in 2023 (111 publications) has surpassed twice the number from 2018 (49 publications).

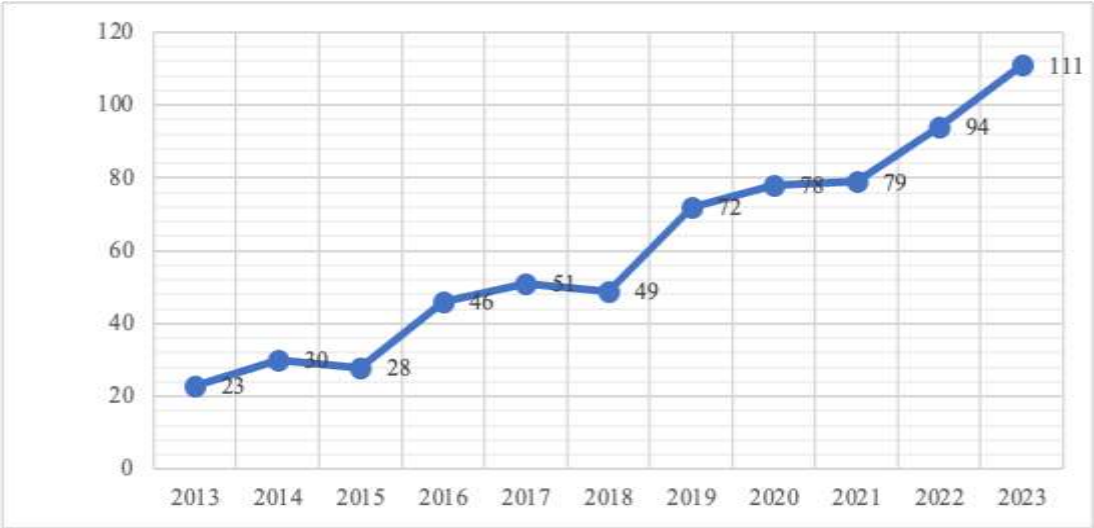


Figure 1. The annual number of publications on machine translation in WoS published from 2013 to 2023

As illustrated in Figure 1 and Figure 2, the quantity of articles acquired from CNKI is lower than that obtained from WoS. The overall trend in published articles on machine translation in CNKI has exhibited fluctuations without a clear upward or downward pattern throughout the last decade. From 2017 to 2019, the number of publications on MT experienced rapid growth, increasing from 12 articles to 34 articles. However, starting in 2019, there has been a decline in the number of publications, averaging about 25 articles on translation technology per year in CNKI's CSSCI from 2020 to 2023. The publication count in 2023 saw a slight decrease compared to that of 2022.

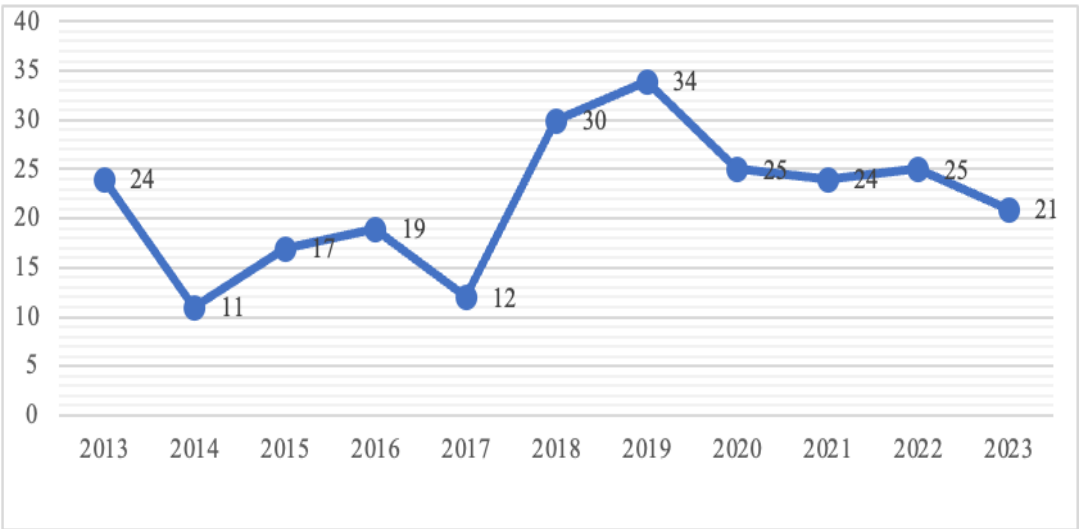


Figure 2. The annual number of publications on machine translation in CNKI published from 2013 to 2023

Analysis of Countries and Institutions

Figure 3 depicts the collaborative network among countries or regions, featuring 71 nodes connected by 255 link lines. Table 1 outlines the leading ten countries based on publication quantity and their centrality. The ten countries or regions with the highest number of research articles were the USA (128), England (106), Spain (72), Germany (69), Netherlands (31), Turkey

(24), Denmark (22), Switzerland (21), Scotland (21) and France (21). In terms of centrality, among the top ten countries, the USA had the highest centrality with 0.39 in the network. England, Spain, Germany, and the Netherlands had centralities of 0.3, 0.19, 0.17, and 0.13, respectively, suggesting that these countries served as crucial bridges in facilitating international cooperation. Despite Switzerland, France, Finland, Sweden, and Saudi Arabia having published a large number of papers, their centrality was below 0.1, suggesting limited collaboration with other countries.

Figure 3. Analysis of Countries on machine technology in WoS

Table1. Top 10 countries/regions on translation technology in WoS

Ranking	Frequency	Country/region	ranking	Country/region	Centrality
1	128	USA	1	USA	0.39
2	106	England	2	England	0.3
3	72	Spain	3	Spain	0.19
4	69	Germany	4	Germany	0.17
5	31	Netherlands	5	Netherlands	0.13
6	24	Turkey	6	Switzerland	0.09
7	22	Denmark	7	France	0.07
8	21	Switzerland	8	Finland	0.05
9	21	Scotland	9	Sweden	0.05
10	21	France	10	Saudi Arabia	0.05

Table 2 illustrates the top 10 institutions that have published research on translation technology in WoS. Dublin City University leads the list with 17 publications, indicating its significant contribution to this field. Following closely are Aarhus University and the Autonomous University of Barcelona, each with 11 publications, suggesting their active involvement in translation technology research. The University of London and Google Incorporated share the third position, both having published 10 articles on this topic. The Centre National de la Recherche Scientifique is ranked fourth with 9 publications. Finally, in fifth place, we have four institutions: University College London, Guangdong University of Foreign Studies, University of Bristol, and Universitat Politecnica de Valencia, each with 7 publications. These institutions represent a mix of universities and private sector organizations from various

countries, reflecting the collaborative and international nature of research in translation technology. Figure4 reveals a modest amount of collaboration among these publishing institutions, yet the cooperation between them is minimal.

Table2. Top 10 institutions on translation technology in WoS

Ranking	Frequency	Institution
1	17	Dublin City University
2	11	Aarhus University
2	11	Autonomous University of Barcelona
3	10	University of London
3	10	Google Incorporated
4	9	Centre National de la Recherche Scientifique
5	7	University College London
5	7	Guangdong University of Foreign Studies
5	7	University of Bristol
5	7	Universitat Politecnica de Valencia

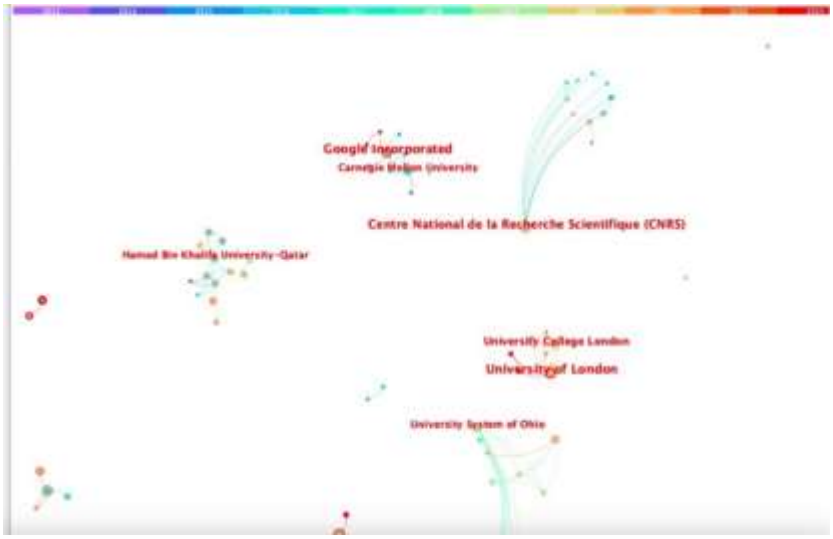


Figure 4. Cooperation between institutions on translation technology in WoS

Figure 5 contains 193 nodes and 88 connections. The connections between institutions were limited in China, with 193 institutions included in the network, and 88 of them having published only one article. As indicated in Table 3, the institutions that published the most articles were Guangdong University of Foreign Studies (17), Shanghai International Studies University (9), Beijing Foreign Studies University (7), and Zhejiang University (7). Meanwhile, Figure 5 illustrates that cooperation and communication between these institutions were infrequent.

Table3. Top 10 Institutions on translation technology in CNKI

Ranking	Frequency	Institution
1	17	Guangdong University of Foreign Studies
2	9	Shanghai International Studies University
3	7	Beijing Foreign Studies University
4	7	Zhejiang University

5	6	Peking University
6	5	Beijing Normal University
7	5	Central South University
8	4	University of International Business and Economics
9	4	Dalian University of Technology
10	4	Nankai University

Figure 5. Analysis of Institutions on translation technology in CNKI

Analysis of Hotspots

Keywords play a crucial role in identifying research themes and areas of intense focus over a specific period, and burst words, which are identified from all keywords, act as significant markers for forecasting leading-edge subjects and upcoming trends (Synnestvedt et al., 2005). As illustrated in Figure 6 and Figure 7, machine translation, translation technology, artificial intelligence, computer-assisted translation, post-translation editing, neural machine translation, and talent training emerge as research hotspots in both CNKI and WoS over the past ten years. Additionally, google translate is identified as one of the research hotspots in WoS.

It's noted that the terms exhibited by CiteSpace are associated with the utilization of MT in translation studies. Terms like “artificial intelligence”, “language”, “machine learning”, “post-editing”, “translation quality”, “neural machine translation”, “evaluation”, and “translator training” improve the quality of MT outputs. This discovery aligns with earlier systematic reviews (Rivera-Trigueros, 2022), which illustrated that NMT was the most frequently employed MT tool in earlier research.

Additionally, this research discovered that Google Translate was frequently cited as a key term by authors in the selected studies, suggesting that it has become the prevailing type of MT, especially after it started using NMT technology in 2016. In 2016, Google implemented NMT to enhance the accuracy of machine translation by developing AI-based algorithms that predict following phrases and segments, reducing ambiguities associated with synonyms. Moreover, the examination of studies on this subject showed that NMT frequently appeared as a key term. This observation is consistent with the findings of Rivera-Trigueros (2022), which emphasized that Google Translate, employing NMT, became the most prominent machine translation tool used

by researchers in translation studies and language education. This suggests a broad acceptance and acknowledgment of NMT as an influential technology in these fields.

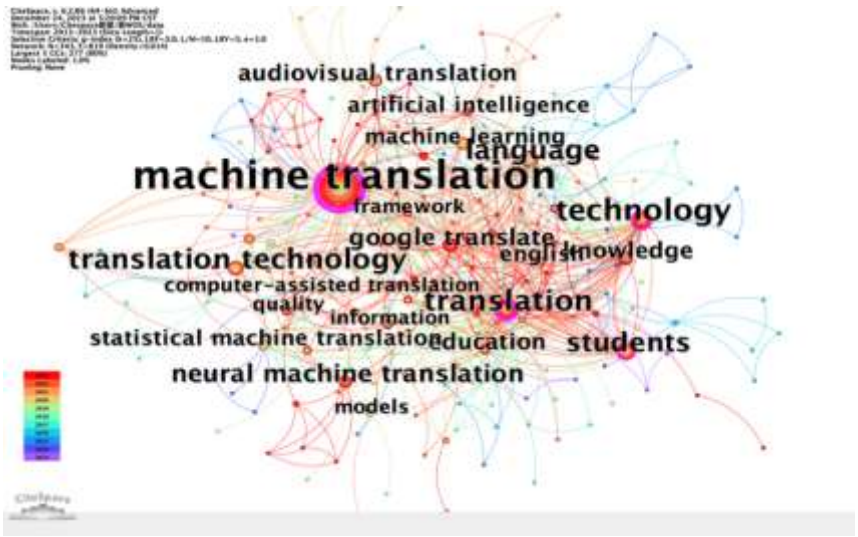


Figure 6. Analysis of keywords on translation technology in WoS

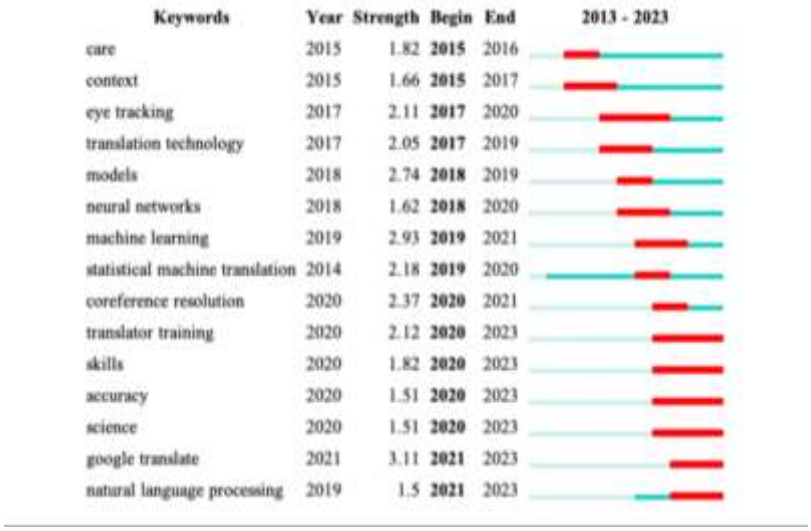
Figure 7. Analysis of keywords on translation technology in CNKI

Analysis of Frontier Research

CiteSpace’s “mutation detection calculation” can identify sudden spikes in research interest, providing an accurate reflection of the latest trends within a specific research field (Li, 2014). Figure 8 and Figure 9 display the top keywords exhibiting the most significant citation bursts in MT research over the last decade

Figure 8 shows that the keyword with the highest mutation intensity in the last decade was “machine learning” in WoS. “Google translate” and “natural language processing” have emerged as forefronts of research in machine translation in the last three years. For example, Cornelison et al. (2021) found that clinicians need to be aware of possible errors in translation when using Google Translate to convert usage instructions and typical advice for prescription medications into Arabic, Chinese (Simplified), and Spanish.

Figure 9 shows that the keyword with the highest mutation intensity in the past 10 years was “computer-aided translation” in CNKI. Over the last three years, “Post-editing” has become the research frontier in the realm of machine translation in China. Cai and Wang (2023) classified and analyzed the errors in light of the machine translation results, and proposed the corresponding post-editing modes. Some Chinese scholars have studied post-translation editing. For example, Wang, Chen and Zhou (2023) drew a research map studies on cognitive post-editing, which highlights three major themes: cognitive processes in post-editing, measurement and assessment of post-editing, and pedagogical approaches to post-editing. Jia and Sun (2022) identified three sets of factors that influenced the difficulty of post-editing, namely, texts (including source text and its machine translation), post-editors, and environments.



Keywords	Year	Strength	Begin	End	2013 - 2023
care	2015	1.82	2015	2016	
context	2015	1.66	2015	2017	
eye tracking	2017	2.11	2017	2020	
translation technology	2017	2.05	2017	2019	
models	2018	2.74	2018	2019	
neural networks	2018	1.62	2018	2020	
machine learning	2019	2.93	2019	2021	
statistical machine translation	2014	2.18	2019	2020	
coreference resolution	2020	2.37	2020	2021	
translator training	2020	2.12	2020	2023	
skills	2020	1.82	2020	2023	
accuracy	2020	1.51	2020	2023	
science	2020	1.51	2020	2023	
google translate	2021	3.11	2021	2023	
natural language processing	2019	1.5	2021	2023	

Figure 8. Top 15 keywords with strongest citation bursts in WoS.

Keywords	Year	Strength	Begin	End	2013 - 2023
计算机辅助翻译	2013	3.73	2013	2016	
课程设置	2015	2.15	2015	2016	
语言服务行业	2015	2.15	2015	2016	
翻译教学	2015	2.02	2015	2019	
“一带一路”	2015	1.44	2015	2018	
人才培养	2017	1.27	2017	2017	
深度学习	2016	1.52	2018	2018	
大数据	2013	1.3	2018	2018	
人工智能	2018	3.67	2019	2021	
翻译	2014	1.28	2020	2020	
翻译技术	2013	2.78	2021	2021	
译后编辑	2015	1.69	2021	2023	

Figure 9. Top 12 keywords with strongest citation bursts in CNKI

As shown in Figure 8, keywords with the strongest citation bursts including “care”, “context”, “eye-tracking”, “translation technology” “models” were once the research frontiers in this field from 2015 to 2019. “Neural networks”, “machine learning”, “statistical machine translation” and “conference resolution” are the research hotspots from 2019 to 2021. “skills”,

"accuracy", "science", "google translate" and "natural language processing" are the current research frontiers in translation technology in the last three years.

D. Conclusion

We gathered 240 high-quality Chinese academic papers from CNKI's CSSCI and 661 English research articles from WoSCC, covering the period from 2013 to 2023, as the subjects of our study. This study conducted scientometric analysis to present the trend of publication outputs, analyze countries and institutions, explore research hotspots, and highlight the differences between domestic and international research in translation technology over the past decade. This study draws the following conclusions: (1) It reveals a consistent growth in yearly publications from the Web of Science, whereas the publications from CNKI exhibit no distinct trend, either upward or downward, throughout the last decade; (2) Scholars who have published papers in the field of translation technology mainly publish papers independently; (3) Research focal points within translation technology include machine translation, translation technology, artificial intelligence, computer-assisted translation, post-translation editing, neural machine translation, and talent training; (4) Post-editing, Google Translate, and natural language processing represent the forefronts of research in the field of machine translation over the past three years.

The study has some limitations. First, the study is constrained by the selection of databases; as only two representative databases were chosen, excluding others. Subsequent research should tackle this constraint by broadening to more inclusive databases like Scopus. Second, there is a possibility of overlooking relevant terms that could have been used to identify additional research papers. Third, due to space constraints, this study did not analyze co-cited literature, authors, and journals. These aspects could be explored in future studies.

References

Bowker, L. (2002). *Computer-aided translation technology: A practical introduction.* Ottawa: University of Ottawa Press.

Cai, Y. & Wang, H. (2023). An Investigation into the Post-editing Mode of Neural Machine Translation of Science and Technology Texts. *Translation Research and Teaching*, (02), 120-127.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*(3), 359-377.

Cornelison, B. R., Al-Mohaish, S., Sun, Y., & Edwards, C. J. (2021). Accuracy of google translate in translating the directions and counseling points for top-selling drugs from English to Arabic, Chinese, and Spanish. *American Journal of Health-System Pharmacy*, *78*(22), 2053-2058.

Deng, X., & Yu, Z. (2022). A systematic review of machine-translation-assisted language learning for sustainable education. *Sustainability*, *14*(13), 7598.

Diao, H. (2017). An Overview of China's Translation Technology Studies. *Journal of Beijing International Studies University*, (06), 69-81+125.

Dong, D., & Chen, M. L. (2015). Publication trends and co-citation mapping of translation studies between 2000 and 2015. *Scientometrics*, *105*, 1111-1128.

Fan, L., & Yun, M. (2022). The past, present and future of machine translation in China: A visualization study based on CNKI literature (1959− 2021). *International Journal of English Linguistics*, *12*(6), 98-111.

Hu, Y., Yu, Z., Cheng, X., Luo, Y., & Wen, C. (2020). A bibliometric analysis and visualization of medical data mining research. *Medicine*, *99*(22).

Huang, Q., & Liu, F. (2019). International translation studies from 2014 to 2018: A bibliometric analysis and its implications. *Translation Review*, *105*(1), 34-57.

Jia, Y. F. & Sun, S. J. (2022). Measuring the Difficulty of Machine Translation Post-Editing: A Research Framework. *Foreign Languages in China*, (03), 16-24.

Li, H. M. (2014). The Cutting-edge Research Interests and Topics of International Translation Studies: A Visualized Analysis of Relevant SSCI Data (2008-2012). *Chinese Translators Journal*, 35(2), 21-26.

Munkova, D., Hajek, P., Munk, M., & Skalka, J. (2020). Evaluation of machine translation quality through the metrics of error rate and accuracy. *Procedia Computer Science*, *171*, 1327-1336.

Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. *Language Resources and Evaluation*, *56*(2), 593-619.

Rovira-Esteva, S., Orero, P., & Franco Aixelá, J. (2015). Bibliometric and bibliographical research in translation studies. *Perspectives*, *23*(2), 159-160.

Synnestvedt, M. B., Chen, C., & Holmes, J. H. (2005). CiteSpace II: Visualization and knowledge discovery in bibliographic databases. *AMIA Annual Symposium Proceedings, 2025, 724-728.*

Van Doorslaer, L., & Gambier, Y. (2015). Measuring relationships in translation studies. On affiliations and keyword frequencies in the translation studies bibliography. *Perspectives*, *23*(2), 305-319.

Wang, X. L. Chen, G. J. & Zhou, X. Y. (2023). Mapping the Structure of Cognitive Machine Translation Post-Editing Studies (2011—2021). *Journal of Foreign Languages*, 46(5), 90-100.

Zhang, D., Xu, J., Zhang, Y., Wang, J., He, S., & Zhou, X. (2020). Study on sustainable urbanization literature based on Web of Science, scopus, and China national knowledge infrastructure: A scientometric analysis in CiteSpace. *Journal of Cleaner Production*, *264*, 121537.

Zhang, X. (2017). Field composition and development trend of research hotspots of translation technology in China-based on co-word visualization analysis of relevant academic journals from CNKI published from 1999 to 2017. *Theory and Practice in Language Studies*, *7*(12), 1210-1221.