Innovation Humanities and Social Sciences Research, Volume 21, Issue 7, 2025. ISSN: 2949-1282
Proceedings of The 6th International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2025)

Research Trends and Prospects of Corpus Translation: A Visual Analysis Based on CiteSpace (2003-2023)

Zhengfeng Chen[1] Jiaxuan Qiao[2]

[1,2] *School of Foreign Languages, Xidian University, Xi'an 710126, China*
[2] *Corresponding author. Email: 22091213370@stu.xidian.edu.cn*

ABSTRACT
This article aims to review academic papers on Corpus Translation in China from 2003 to 2023 systematically and explore the research progress in this field. An in-depth analysis of the trends are conducted in the number of publications in this field with the visualization software CiteSpace, thus the research activity and shifting focus over different period is revealed in this field. Additionally, this paper analyzes the collaboration among authors and institutions and academic exchanges. Through co-occurrence network analysis of keywords, the research hot-spots and provided insights into future research directions are identified. This study contributes to a better understanding of the significance of corpus translation in the translation domain and offers valuable references and guidance for future research endeavors.

Keywords: *Corpus translation, CiteSpace, Co-occurrence network, Visual analysis.*

1. INTRODUCTION

With the rapid advancement of information technology, corpus-based translation has emerged as a critical topic within the field of translation studies. By leveraging large-scale linguistic resources and tools, corpus-based translation enables translators to achieve greater efficiency and accuracy. Since 2003, extensive research has been conducted in this domain, driving its continuous evolution and innovation.

This paper aims to comprehensively examine the research progress in corpus-based translation, specifically analyzing scholarly articles published from 2003 onward. Utilizing CiteSpace for visual analysis, it identifies key trends, challenges, and potential future directions within the field. Through systematic synthesis and evaluation of existing studies, this review seeks to enhance readers' understanding of the practical advantages, limitations, and challenges of corpus-based translation, thereby providing valuable insights for further research and application. Ultimately, this study intends to present the latest developments in corpus-based translation, foster scholarly exchange and collaboration, and contribute to ongoing innovation within the field of translation studies.

2. CONCEPTUAL FRAMEWORK OF CORPUS-BASED TRANSLATION RESEARCH

Corpus-based translation is an interdisciplinary field integrating linguistics, computer science, and translation studies. Scholars internationally have proposed varying definitions of the concept: Baker (1995) defines it as "the process of utilizing large-scale multilingual text corpora to provide support for translation", emphasizing the corpus role in enhancing translation quality through information retrieval.

Bowker (2002) offers a broader definition: "any form of translation assistance delivered via computers and corpus resources." This encompasses technologies like translation memory tools and automated translation systems. Somers (2003) focuses on the translation process, describing it as "the active process where translators consult corpus resources for information retrieval, reference, and verification during their work." This highlights the interactive relationship

between the translator and the corpus as a decision-support tool. Hutchins (1995) characterizes it as "the process of assisting translators through corpus technology", underscoring the instrumental role of corpora in translation workflows.

Chinese scholars have also contributed distinct viewpoints. Chen Wei (2013) posits that corpus-based translation—through resources like translation memories, terminology databases, and parallel corpora—enables translators to better comprehend source texts while enhancing translation accuracy and naturalness. Regarding advantages, Li Ming (2016) emphasizes that corpus-based tools provide real-time translation suggestions during the workflow, supporting more precise decision-making. Dai Guangrong & Wang Kefei (2021) argue that large-scale corpus analysis allows these tools to capture idiomatic expressions, yielding more authentic translations. On terminology management, Tao Youlan (2010) underscores its critical role in specialized translation, noting that corpus tools ensure terminological consistency and accuracy through automated recognition features. Hu Kaibao (2022) highlights how these tools boost efficiency in translating vast online content amid rapidly evolving demands.

In summary, while scholarly definitions vary, consensus centers on three core elements: corpora, computational technologies, and the translation process. Corpus-based translation thus emerges as a computational process leveraging corpus resources to support information retrieval, reference, verification, and decision-making during translation. This integrated definition comprehensively captures the field scope.

3. RESEARCH DATA AND FRAMEWORK

3.1 Data Sources and Screening

This study draws from CSSCI-indexed journals in the CNKI database, covering publications from 2003 to 2023. Search terms included "corpus-based translation" and "corpus translation". Non-research items (e.g., conference notices, calls for papers) were excluded prior to screening.

Screening criteria:

Papers must directly address corpus-based translation, explicitly referencing related concepts, methods, or practices in titles, abstracts, or full texts. Literature published between 2003–2023 was selected, ensuring coverage of recent developments. Priority was given to papers in reputable journals/conference proceedings or authored by established scholars. Citation metrics, author expertise, and publisher reputation were evaluated.1,068 Chinese-language papers met these criteria.

3.2 Research Methodology and Procedure

This study employs bibliometric analysis to statistically examine the annual publication volume in corpus-based translation research from 2003 to 2023, revealing the field's evolutionary trajectory and development patterns. Time-series quantitative analysis not only reflects cyclical shifts in research activity but also enables forecasting of future research directions through regression modeling.

Knowledge mapping analysis was conducted using CiteSpace (Version 6.2.R4), which transforms bibliographic data into visual network maps through co-occurrence analysis, cluster analysis, and burst detection algorithms. The analytical procedure is illustrated in "Figure 1":

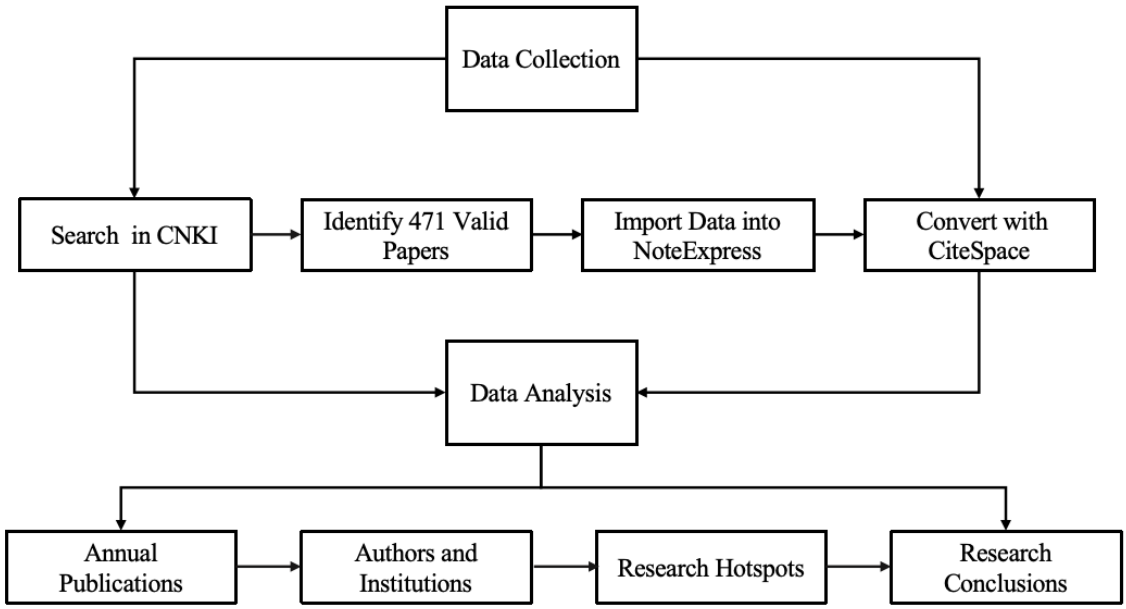

Figure 1 Visual analysis workflow.

4. RESULTS ANALYSIS

4.1 Publication Trends in Corpus-Based Translation Research

The publication trend for corpus-based translation literature (2003–2023) is visualized in "Figure 2":

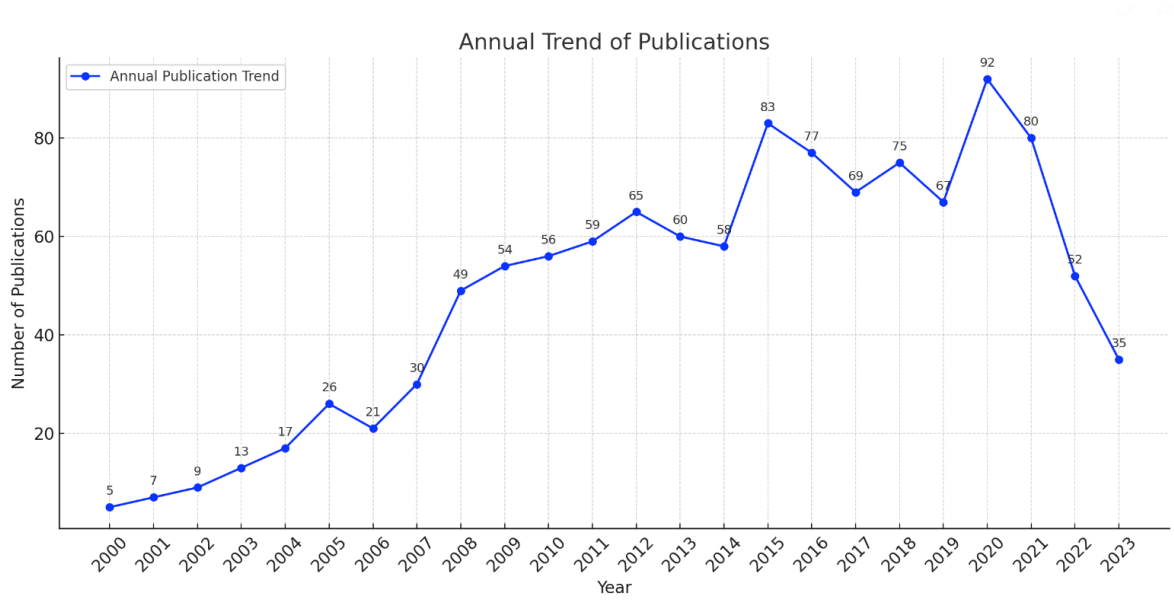

Figure 2 Annual publication volume trend.

As shown in "Figure 2", publications in corpus-based translation demonstrated an overall upward trajectory from 2003 to 2020. This period coincided with rapid technological advancements—including machine translation, natural language processing, and computer-assisted translation tools—that stimulated deeper research and practice, driving sustained growth in scholarly output. However, as these technologies approached maturity, growth in publication volume began to slow.

Notably, the upward trend was punctuated by temporary declines in 2006, 2012, and 2018. These inflection points likely correspond to the emergence of new research hotspots that diverted scholarly attention from corpus-based translation, resulting in reduced publication output.

A significant downturn occurred in 2020 amid the global COVID-19 pandemic, which profoundly impacted academia worldwide. Resource reallocation at research institutions, cancellations/postponements of academic conferences, and shifted research priorities toward pandemic-related topics collectively contributed to reduced publications in this field.

In summary, the decline in corpus-based translation publications stems from interwoven factors: technological maturation, competing research interests, and global disruptions. This trend reflects the dynamic nature of academic research while underscoring how external events and technological cycles influence scholarly focus. Future fluctuations in publication volume may occur as novel research interests emerge and technologies evolve, necessitating ongoing monitoring of field developments.

4.2 Burst Terms in Corpus-Based Translation Literature

Burst terms, keywords exhibiting sharply increased frequency within specific periods, were identified using CiteSpace's Kleinberg burst detection algorithm. As shown in "Figure 3", these terms demonstrate clear temporal clustering (highlighted in red), with Burst Strength values positively correlating with frequency growth rates. This analysis collectively reveals the evolution of research hotspots.

Keywords	Year	Strength	Begin	End	2003 - 2023
Information Retrieval	2004	3.77	2004	2007	
Machine Translation	2004	3.41	2004	2008	
Translation Pedagogy	2004	3.36	2010	2013	
Translator's Style	2011	3.36	2012	2014	
English Translation	2012	2.36	2012	2013	
Translation	2008	2.28	2015	2017	
Artificial Intelligence	2017	3.26	2017	2021	
Translation Style	2018	3.4	2018	2023	
Explicitation	2008	3.17	2018	2019	
Chinese Translation	2018	2.51	2018	2021	
Third Code	2019	2.44	2019	2021	
Digital Humanities	2018	3.02	2021	2023	

Figure 3 Top 12 keywords with the strongest citation bursts.

4.3 Thematic Analysis of Burst Terms

In terms of CiteSpace-generated burst term tables, the researchers categorize keywords into five thematic clusters, which reflects core developmental trajectories in corpus-based translation research. The starting time and ending time on these terms further reveal critical evolutionary shifts. Below is an in-depth analysis of these themes and temporal patterns:

4.3.1 Echnology-Driven Automation

Burst terms machine translation (2004-2008) and artificial intelligence (2017-2021) underscore technology's pivotal role. Research on machine translation emerged early, achieving significant advances in automation. Post-2017, AI gained prominence, enabling further automation and intelligent capabilities in translation technology. This trajectory demonstrates continuous

technological evolution to address increasingly complex translation demands.

4.3.2 Education and Training

Translation pedagogy (2010-2013) and translation style (2018-present) represent educational themes. The former reflects efforts to cultivate new generations of translators, while the latter recent emergence signals growing interest in advanced training and personalized education. Both converge on enhancing translator competencies.

4.3.3 Core Theory and Research

The sustained focus on translation (2015-2017) confirms its status as a foundational research pillar, emphasizing theoretical and practical fundamentals. Its declining prominence post-2017 may indicate disciplinary pivots or emerging subfields.

4.3.4 Cultural and Cross-Cultural Communication

Explicitation (2018-2019) and digital humanities (2021-present) highlight cultural dimensions. Explicitation addresses cultural element visibility in translations, while digital humanities applies computational tools to cross-cultural challenges. These themes emphasize globalization-driven cultural awareness and digital methodologies.

4.3.5 Multilingual Studies

Chinese translation (2018–2021) and third code (2019–2021) address multilingual complexity. The former tackles Chinese-specific translation challenges, the latter explores interlingual mediation. Both respond to demands for translation in linguistically diverse contexts.

4.3.6 Interthematic Connections

Significant synergies exist across clusters. Technology-driven automation intersects with education as technological advances reshape training requirements. Core theory permeates all themes, while cultural communication and multilingual studies overlap through shared focus on language-culture interplay in cross-cultural contexts.

In summary, these thematic and temporal patterns yield critical insights into corpus-based translation research. Cross-disciplinary approaches will be essential to address emerging complexities, driving sustained innovation in the field.

5. COLLABORATION NETWORKS AND KEYWORD ANALYSIS

5.1 Author Collaboration Patterns in Corpus-Based Translation Research

5.1.1 Author Collaboration Patterns

"Figure 4" reveals that Hu Kaibao, Wang Kefei, and Liu Zequan occupy prominent positions, indicating their substantial publication output in the field. Scholars like Pang Shuangzi, Qin Hongwu, and Liu Dingjia also demonstrate high productivity. The collaboration network highlights strong research ties between Hu Kaibao/Wang Kefei and most co-authors. The color gradient from core to periphery of node rings reflects sustained research continuity from 2003 to 2023, alongside growing scholarly engagement. However, broader collaboration remains limited.

Figure 4 Author collaboration network in corpus-based translation research.

5.1.2 Research Hotspots Analysis

In CiteSpace co-occurrence networks, node size represents keyword centrality—a metric of conceptual importance within the research landscape. Connecting lines indicate relational strength between keywords (Chen et al., 2014). Through the Keyword Co-occurrence table cited in "Table 1", the researchers analyze High-frequency terms, which include: translator style, translation studies, machine translation, semantic prosody, digital humanities, literary translation.

Table 1. Keyword frequency table

Order	Count	Strength	Year	Keyword
1	235	0.54	2003	Corpus
2	69	0.07	2008	Translation
3	28	0.08	2006	Translation studies
4	27	0.08	2004	Machine Translation
5	25	0.02	2011	Translator's Style
6	15	0.01	2009	Interpreting Studies
7	13	0.01	2008	Translation Universals
8	13	0.00	2008	Explicitation
9	12	0.00	2012	English Translation
10	9	0.00	2010	Semantic Prosody
11	9	0.00	2017	Artificial Intelligence
12	9	0.00	2018	Translation Style
13	8	0.01	2018	Digital Humanities
14	8	0.00	2010	Translation Strategy
15	7	0.01	2011	Translatology

5.2 Thematic Cluster: Technology-Driven Keywords

5.2.1 Machine Translation and Artificial Intelligence Epitomize Technology Pivotal Role

Machine translation leverages computational algorithms and large-scale corpora for automated translation. Artificial intelligence has gained prominence since 2017, introducing intelligent capabilities that enhance speed and accuracy in processing massive texts.

5.2.2 Keywords on Translation Strategies and Stylistic Features: Translator Style, Translation Style, Translation Strategy

These keywords spotlight the significance of strategies and styles in translation. Translation is a blend of art and science, where the translator personal choices and style play a crucial role. Researchers delve into the differences in translation styles and strategies among translators and their impact on the final translation. This is key for enhancing translation quality and fostering personalized translation.

5.2.3 Keywords on Explicitation and Semantic Features: Explicitation, Semantic Prosody

These keywords center on research into explicitation and semantics in corpus-based translation. Explicitation deals with conveying cultural elements and semantic traits in translation, which is vital for accurate text processing. Semantic prosody concerns the phonological aspects of language and is essential for transliteration and phonetic handling in translation. Research in these areas boosts translation accuracy and context - adaptability.

5.2.4 Keywords on Interdisciplinary Studies: Interpreting Studies, Digital Humanities

These keywords pertain to interdisciplinary research. Interpreting studies focus on the process and skills of interpreting, linked to oral communication in corpus - based translation. Digital humanities emphasize the use of digital tools in translation and cultural spread, tied to corpus digitization. The intersection of these fields broadens the research scope and offers a more comprehensive understanding.

5.2.5 Keywords on Translation Universals and Norms: Translation Universals

These keywords focus on universals and norms in translation. Translation universals research explores common features and rules across different contexts, helping set translation standards. Corpus - based translation studies can uncover these universals and guide translators on best practices.

In summary, the categorization and analysis of these keywords mirror the diversity and complexity of corpus - based translation studies. By studying these keywords, we can gain in - depth insights into translation techniques and strategies, explore cross - cultural and semantic aspects, and drive the advancement of translation studies. They also reveal the intersection of corpus - based translation with other research areas, spurring interdisciplinary research progress.

6. FUTURE RESEARCH DIRECTIONS

Corpus-based translation studies have achieved remarkable progress since 2003, yet the field continues to present extensive opportunities and challenges. Future research may advance this discipline by focusing on the following areas:

6.1 Development and Application of Multilingual Corpora

Multilingual corpora offer broad applicability. By analyzing such corpora, researchers can deepen their understanding of grammatical, lexical, and semantic variations across languages, thereby enhancing translation accuracy and fluency. Additionally, these corpora facilitate cross-cultural research and communication, enabling nuanced interpretations of linguistic expressions and meaning transfer within diverse cultural contexts. Efforts should be devoted to constructing parallel corpora for more language pairs—particularly low-resource languages—to address varied translation needs.

6.2 Integration of Corpus-Based Translation and Natural Language Processing (NLP)

Advancements in NLP present opportunities for deeper synergy with corpus-based translation. This includes leveraging pre-trained language models (e.g., BERT, GPT) to refine translation quality and integrating NLP applications—such as automated summarization and question-answering systems—to enable comprehensive language processing solutions. Furthermore, breakthroughs in corpus-driven machine translation may be achieved through deeper adoption of deep learning techniques.

6.3 Innovative Approaches to Translation Quality Assessment

Translation quality assessment remains pivotal to corpus-based translation research. Future studies should explore novel methodologies, particularly hybrid frameworks combining automated and human evaluation. For instance, deep learning could refine automated assessment models to holistically gauge translation accuracy, fluency, and semantic coherence. Research should also investigate how user feedback and subjective evaluations can complement systematic quality assessment.

6.4 Interdisciplinary Collaboration

The field of corpus-based translation requires interdisciplinary research approaches and collaboration. Future studies should foster closer partnerships with experts in computer science, linguistics, sociology, and cultural studies to address complex translation challenges. Such interdisciplinary cooperation can provide researchers with diverse perspectives and methodological tools to navigate the multifaceted nature of translation.

6.5 Ethical and Cultural Considerations

Translation extends beyond technical proficiency to encompass ethical and cultural dimensions. Future research should prioritize ethical concerns in translation, including cultural sensitivity and handling potentially discriminatory language. Scholars should explore how corpus-based approaches can address these ethical challenges while developing robust ethical guidelines for practitioners. Additionally, deeper investigation into cultural adaptation during cross-cultural communication is needed to ensure translations accurately convey cultural meaning.

7. CONCLUSION

This study examines core journal publications (2003-2023) in corpus-based translation studies, revealing that while scholars worldwide have explored this domain from multiple perspectives, opportunities remain for deeper integration of methodological approaches. Advancing this field will require enhanced interdisciplinary collaboration. Corpus-based translation will continue driving progress by providing researchers with increasingly sophisticated tools. By addressing current challenges, we can better serve multilingual societies, foster global cultural exchange, promote language learning, and propel corpus-based translation studies to new heights.

ACKNOWLEDGMENTS

Funded Projects: Xi’an Social Science Fund Project (25YZ17): Research on the Digital Transformation of Red Culture Translation in Xi’an and the Enhancement of International Communication Efficacy in the New Media Context in 2025;Shaanxi Philosophy and Social Sciences Research Special Project: Foreign Publicity Translation of Shaanxi's Intangible Cultural Heritage from the Perspective of Communication Studies in the Intelligent Media Era(2024HZ0841) in 2024; Key Teaching Reform Project of Xidian University : Collaborative Education Models, Pathways, and Evaluation Mechanisms Between "Curriculum-based Ideological and Political Education" in College English and Ideological-Political Courses Under the "Macro-Ideological Education" Framework (A2313) in 2023; Industry-Academia Collaborative Education Project in 2024: Building a New Smart Teaching Ecosystem for College English in Western Chinese Universities in the GAI Technology Era ;Key Fundamental Research Project of Xidian University : Cultural Security and International Discourse Power Construction from an Area Studies Perspective: Strategies for China's Cultural Overseas Communication in 2025 and 2025Xidian University Specially Funded Project for Interdisciplinary Exploration(TZJHS202502).

REFERENCES

[1] Chen Wei. Corpus-Based Study of Chinese-English Legal Performative Verbs [Doctoral dissertation]. Central China Normal University. 2013.

[2] Chen Yue, Chen Chaomei, Hu Zhigang. Principles and Applications of Citation Space Analysis: A CiteSpace Practical Guide. Science Press. 2014.

[3] Dai Guangrong, Wang Kefei. Comparing Linguistic Intuition and Corpus Approaches in Translation Studies. Technology Enhanced Foreign Language Education, 2021, (03), 96–101+15.

[4] Hu Kaibao, Zhang Lili. Corpus-Based Rhetorical Studies: Features, Topics, and Significance. Contemporary Rhetoric, 2022, (04), 1–10.

[5] Hu Kaibao. Language Data Science and Applied Linguistics: Characteristics, Domains and Methodologies. Foreign Language World, 2022, (03), 37–44.

[6] Li Ming. Corpus-Based Semantic Prosody Study of Synonyms: A Case Study of the Word “Become, Get, and Go”. Research on Composition Teaching, (06), 117.

[7] Qin Hongwu. Development and Application of Bilingual Corpora. Foreign Language Teaching and Research Press. 2021.

[8] Tao Youlan. Corpus-Based Development of Interpretation Textbooks for Translation Majors. Foreign Language World, 2010, (04), 2–8.

[9] Xiao Zhenhua, Dai Guangrong. A New Framework for Translation Teaching and Research: Corpus Translation Studies Review. Foreign Language Teaching Theory and Practice, 2011, (01), 8–15.

[10] Xiao Weiqing, Qian Jiajun Xiao. Advances and Trends in Translation Technology Pedagogy (2000–2020): Comparative Analysis of Core International and Domestic Journal Publications. Foreign Language World, 2021, (01), 62–70.

[11] Baker, M. "Language and translation in the context of a post-colonial literature curriculum." [J]. Language and Literature, 1995,4(3):199-214.

[12] Bowker, L., & Pearson, J. Working with specialized language: A practical guide to using corpora[J]. Routledge, 2002.

[13] Gaspari, F. Human translator and machine translation: Friends, foes or strange bedfellows? [J]. Perspectives: Studies in Translatology,2017.

[14] Hutchins, W. J. "Machine Translation: A Concise History."[J]. The Oxford Handbook of Computational Linguistics, 1995:419-437

[15] Somers, H. Computers and translation: A translator's guide[J]. John Benjamins Publishing, 2003.