Theory and Practice of Science and Technology
2026, VOL. 7, NO. 2, 110-115
DOI: 10.47297/taposatWSP2633-456922.20260702

Paradigm Shift in Domestic Linguistic Research Driven by Intelligent Technology: A Bibliometric Analysis Based on CiteSpace

Qin Guo[1], Yaoyou Zhang[2]

[1]College of International Cultural Exchange, Xinjiang Normal University, Urumqi, Xinjiang, 830000, China

[2]Aba Teachers University, School of Physics and Electronic- Electrical Engineering, Aba Tibetan and Qiang Autonomous Prefecture, Sichuan, 623002, China

ABSTRACT

In the digital era, intelligent technologies are propelling paradigm shifts in linguistic research. This study selected relevant literature from the CNKI database spanning 1979 – 2025, manually screened 488 valid papers, and employed CiteSpace for bibliometric and visualization analysis to examine the current state, hotspots, and trends in this field. Findings reveal three developmental phases: preliminary exploration (1979–2002), gradual development (2003–2022), and rapid growth (2023–2025). A core author group has emerged, including Liu Haitao and Li Yuming, though intergroup collaboration remains weak. Research is predominantly driven by diverse types of institutions such as Beijing University of Chinese Medicine, and cross-institutional participation is prominent. Research hotspots revolve around artificial intelligence and natural language processing, with current focus on cutting-edge directions such as linguistic intelligence and interdisciplinary studies. Concurrently, the field faces challenges including insufficient cross-group coordination, inefficient resource integration, inadequate theoretical exploration, and low-level repetition. Future efforts should strengthen multi-party collaboration, deepen interdisciplinary integration, enrich theoretical frameworks through frontier research, and advance the high-quality development of linguistic research paradigms.

KEYWORDS

Intelligent technology; Linguistics; Citespace

1 Introduction

Linguistics is the discipline that studies language, primarily examining its functions, usage, history, and properties to systematically address various language-related issues. The deepening two-way interaction between large language models and linguistics can drive technological breakthroughs and theoretical innovations in both fields, opening new pathways for linguistic research and applications in the intelligent era. Currently, linguistics research in China faces challenges such as difficulties in inheriting the multi-ethnic languages in China and the inefficiency of traditional research methods.The rise of digital technologies like artificial intelligence, big data, and large language models presents a crucial opportunity to overcome these challenges and drive research paradigm innovation, making their integration inevitable.

Since the reform and opening-up in China, linguistic research has been continuously deepened, forming a scientifically sound theoretical research system . Its research scope has been steadily expanded and has gradually achieved in-depth integration with cutting-edge technologies such as artificial intelligence.This convergence has overcome the fragmented limitations of traditional research, propelling the discipline from rule-based approaches toward data-driven methodologies and from singular qualitative analysis toward integrated qualitative-quantitative frameworks, laying the groundwork for deeper integration. The advent of the digital age, driven by technological waves like artificial intelligence, big data computing, and large language models, has directly catalyzed a paradigm shift in linguistic research. Intelligent technologies not only provide a robust computational foundation for linguistic research but also optimize research processes, enabling more precise and efficient data processing and analysis. This has deepened the vertical research depth of linguistics, with applications in fields like ancient script decipherment and endangered language preservation further highlighting the value of technological empowerment.

While the integration of these two fields has become a hot topic in academia, existing research lacks a systematic review and quantitative analysis of the overall landscape, evolutionary trajectory, and key frontiers in this domain. To address this gap, this study employs CiteSpace software to conduct a quantitative analysis of relevant literature, supplemented by visualization techniques. A knowledge map is utilized to reveal the field's evolutionary trajectory, core hotspots, and future trends, thereby filling research deficiencies and providing empirical support for scholars.

2 Research Design and Methods

2.1 Data Sources

This study retrieved relevant literature on intelligent technology-empowered linguistic research since the reform and opening-up in China, and all data were sourced from the CNKI database. To ensure the comprehensiveness and accuracy of the search results, the advanced search function in the academic journal database of CNKI was adopted, with the search subjects set as "Intelligent Technology" and "Linguistics" and the search mode set as "Accurate".The search time frame was limited from January 1, 1979 to December 31, 2025, and the search scope covered 30 types of journals

(including extended editions). A preliminary search yielded 543 literature records. To ensure the research relevance and data validity, non-academic documents such as calls for papers, book reviews and conference announcements were excluded after manual screening, and 488 valid documents were finally obtained.Each paper includes basic information such as author, institution and keywords. After the selected literature is exported in Refworks format recognizable by CiteSpace software, it is used as the analysis sample of this study.

2.2 Research Methods

CiteSpace is a visual analysis tool for academic literature, focusing on citation network and co-occurrence analysis, which helps researchers gain an insight into the knowledge structure, hotspots and trends of a research field. Its core functions include literature co-occurrence analysis and knowledge graph construction. Through visual maps, the software analyzes the potential dynamic mechanisms and detects the disciplinary development frontiers, thus exploring the key paths and inflection points of the evolution of the research field. The author adopted the CiteSpace tool to construct visual maps of the number of published papers, authors, institutions and keywords of relevant documents, aiming to address the following two research questions: First, the distribution and cooperation characteristics of published paper output, authors and research institutions; Second, the research hotspots and future trends of domestic linguistic research driven by intelligent technology.

3 Results and Analysis

3.1 Statistical Analysis of Publication Volume

Figure 1 presents the annual number of publications in the field of intelligent technology and linguistics in the CNKI database, which intuitively reflects the temporal evolutionary characteristics of domestic research in this field. Annual publication volume is an important indicator to measure the development trend, popularity and maturity of a certain research field. Combined with the visual analysis results of CiteSpace, linguistic research driven by domestic intelligent technology can be divided into three development periods, and the characteristics of each stage are highly consistent with the iteration of intelligent technology and the evolution of linguistic research paradigm.

The first stage is the preliminary exploration period (1979 – 2002). The research was in an embryonic state, with a relatively low annual number of publications, amounting to 32 papers in total, accounting for 6.5% of the total number of valid documents. There were no relevant publications in some years. At that time, intelligent technology worldwide was in the early symbolic stage, and domestic intelligent technologies had not been applied on a large scale, with no sufficient support for the cross-integration of intelligent technologies and linguistics. Related research was scattered, focusing on basic language information processing, and did not form a systematic framework and clear theme.

The second stage is the period of slow development (2003-2022), with fluctuating growth in the annual number of articles published, with a total of 273 articles, up to 56%, climbing from 1 article in 2003 to 37 articles in 2020. With the application of statistical learning, word vector and other technologies to language processing, academic circles began to pay attention to the application value of technology, and research turned to the initial integration of technology and disciplines, and interdisciplinary consciousness appeared, laying the foundation for the subsequent explosive growth.

The third stage is the rapid growth period (2023–2025), during which the annual number of publications has surged, accounting for 37.5% of the total valid documents in three years. Generative AI and large language models have driven the iteration of intelligent technologies, whose disruptive impact on linguistic research paradigms has attracted wide academic attention. Interdisciplinary integration has intensified, with research focusing on paradigm shifts and applications of large language models, marking the field's entry into a mature stage of deep integration.

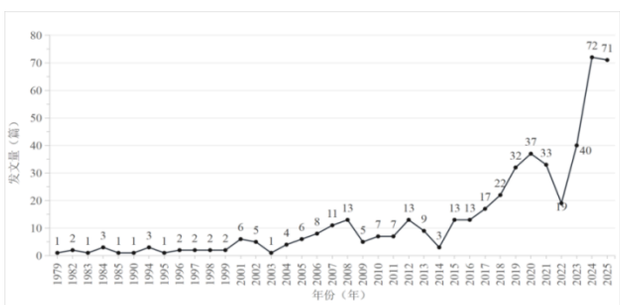

Figure 1 Annual Publication Trends in Intelligent Technology and Linguistic Research

3.2 Core Author Analysis

Core authors (also known as prolific authors) refer to scholars who have published a number of papers in a specific field and have certain academic influence . This study took "authors" as the research node and constructed an author co-occurrence network knowledge graph with a node number (N) of 401, a link number (E) of 198 and a network density of 0.0025.

Specifically, Liu Haitao and scholars such as Li Zuowen have formed a stable collaboration cluster, focusing on interdisciplinary fields like computational linguistics, with their work having significant influence in areas such as language modeling; the collaboration group composed of Li Yuming and Dai Weidong conducts research around language planning, intelligent foreign language education, and other directions, reflecting trends of technological integration; the collaboration network of Yuan Yulin and Feng Zhiwei, among other scholars, concentrates on natural language processing theory and intelligent education applications, highlighting the deep integration of traditional linguistics and modern technology.

From the perspective of network structure, the internal cooperation of each core group is stable, but cross-cluster collaboration is weak due to disciplinary barriers and field differentiation. Peripheral authors such as Ye Zi have scattered cooperation and mostly publish papers independently, failing to integrate into the core collaboration network, which restricts the scaling of research outcomes and the dissemination of academic influence. In summary, the current collaboration network reflects both the field-leading role of core scholars and issues such as insufficient group synergy and inadequate integration of peripheral forces, providing directions for optimizing future interdisciplinary collaboration.



Figure 2 Author Co-occurrence Network Knowledge Map

3.3 Analysis of Research Institutions

Counting the number of publications of research institutions can help identify the core institutions in a specific discipline and also indirectly reflect the frontier research positions and directions of the discipline. To identify the core institutions in the current research field, we conducted a statistical analysis on the number of publications of each research institution (see Figure 3). In the field of intelligent technology and linguistics research, there are 20 institutions with more than 10 publications. Among them, Beijing University of Chinese Medicine ranks first with 284 publications, accounting for about 58% of the total number of publications, followed by Hunan University of Chinese Medicine and Wuhan University. The publication scale of these institutions represents the main force of research in this field, and their results have effectively promoted the development of interdisciplinary research between the two areas.

The types of institutions show characteristics of cross-field participation, covering traditional Chinese medicine universities, normal universities, and comprehensive universities. Traditional Chinese medicine universities occupy three of the top five in terms of publication volume, reflecting the depth of application of intelligent technology in areas such as Chinese medicine language information processing and the digitization of ancient texts; normal universities focus on language intelligence education and computational linguistics theory, forming a stable research team; comprehensive universities, relying on their multidisciplinary advantages, stand out in theoretical integration and methodological innovation.

The regional distribution of core institutions is concentrated in provinces and cities such as Beijing, Hubei, and Hunan, which basically aligns with the distribution pattern of higher education resources in China. Beijing leverages relevant universities to form a research hub, while Hubei and Hunan provinces develop distinctive advantages through sustained investment by key institutions. This dual distribution by region and type provides diverse support for research in this field and lays a foundation for subsequent cross-regional and cross-institutional collaborative innovation.

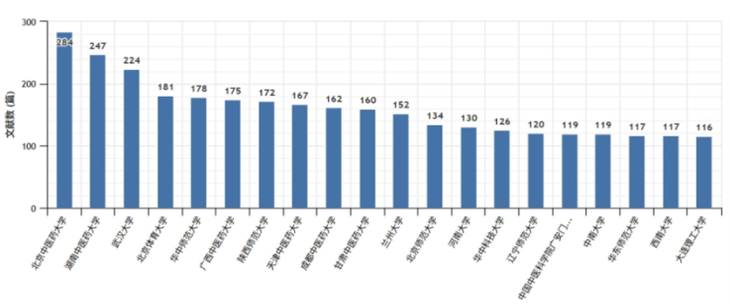

Figure 3 1979 Map of Research Institutions Distribution in 2025

4 Keyword Analysis

4.1 Keyword Co-occurrence Analysis

Keyword frequency can directly reflect research hotspots, and high-frequency keywords usually represent the core research focuses of a field. This study used CiteSpace software and took keywords as nodes to generate a keyword co-occurrence map, with the node number N=509, link number E=877 and network density=0.0068. The relatively low network density indicates that although several core topics have been formed in this field, there is considerable room for expansion in various sub-directions, showing a structural characteristic of 'prominent cores with multiple radiating points.

As can be seen from Figure 4, this field exhibits characteristics of interdisciplinary focus centered around artificial intelligence, with 'artificial intelligence' being a hot topic and 'natural language processing' and 'large language models' as core focal points. In terms of the strength of connections between nodes, 'artificial intelligence' is most closely associated with nodes such as 'natural language processing' and 'machine translation,' indicating that relevant application scenarios have formed stable research communities. Additionally, emerging nodes such as 'generative artificial intelligence' have risen, reflecting a shift of research focus toward the next generation of intelligent language technologies.

In terms of node colors, the purple series corresponds to the early foundational fields of linguistics, while the red series corresponds to recent research hotspots such as "generative artificial intelligence". Research has gradually shifted from the traditional theoretical framework of early "corpus linguistics" to application scenarios such as "language intelligence", which highlights that intelligent technologies have promoted the transformation of linguistic research from theoretical research to practical application and provided a clear context for the subsequent innovation ofresearch paradigms.

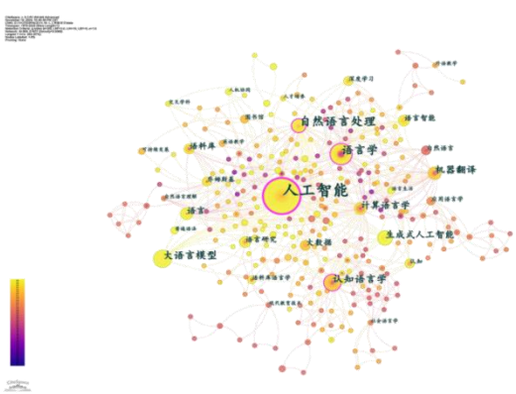

Figure 4 Keyword Co-occurrence Network Knowledge Graph

4.2 Keyword Clustering Analysis

This study conducted a cluster analysis of keywords and generated a knowledge graph with a modularity value Q= 0.8192 and an average silhouette value S=0.9387 (Figure 5). It is well established that when Q>0.3 and S>0.5, the clustering structure of the graph is significant and reliable. Both the Q value and S value are far above the critical threshold, indicating that the clustering results have strong internal homogeneity and high external differentiation, effectively reflecting the knowledge structure characteristics in the interdisciplinary field of intelligent technology and linguistics. The 488 paper samples formed a total of 12 clusters, corresponding to the hot topics in this field since the reform and opening-up, mainly including artificial intelligence, linguistics, and natural language processing.

From the perspective of cluster size and core degree, #0 Artificial Intelligence and #1 Linguistics are the core hubs of the knowledge graph, with high spatial overlap and dense node connections between them. This indicates that intelligent technologies have penetrated the core fields of linguistic research and thus become key drivers of research paradigm innovation. Clusters such as #2 Natural Language Processing and #3 Machine Translation form secondary clusters around these core hubs, demonstrating the practical application of intelligent technologies in language use. Traditional branches like #4 Cognitive Linguistics and #5 Applied Linguistics, through cross-connections with technology-related clusters such as #6 Artificial Neural Networks and #9 Language Intelligence, achieve the expansion of disciplinary boundaries and methodological innovation. In addition, emerging clusters such as #10 Modern Educational Technology reflect the extension of research into practical scenarios, enriching the practical dimension of the study.

Overall, this map clearly presents the distribution of hotspots in the field, revealing the inherent logic of 'technological empowerment—theoretical innovation—application expansion': artificial intelligence provides new tools for linguistics, promotes breakthroughs in application areas, and offers opportunities for the integration of cross-school theories; at the same time, the interaction between traditional linguistic theories and intelligent technologies deepens, giving rise to new directions such as language intelligence; meanwhile, clusters like modern educational technology reflect the transformation of research outcomes into practice, highlighting the trend of research paradigms shifting from theory-driven to a combination of theory and application.

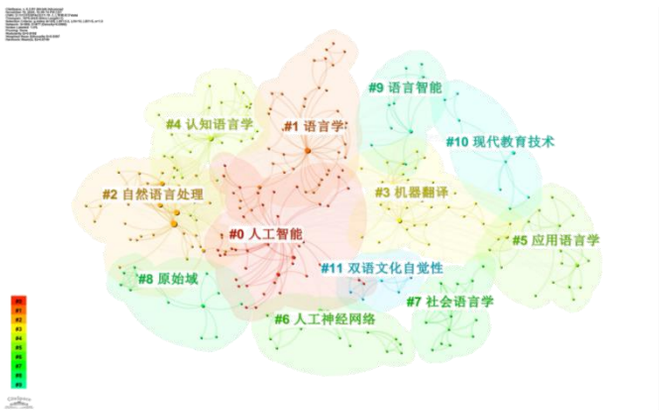

Figure 5 Keyword Clustering Network Knowledge Graph

4.3 Analysis of Keyword Emergence

Keyword burst detection can identify sudden research growth in a disciplinary field and help understand its frontier research trends and evolutionary processes.Through burst detection, a total of 25 burst keywords emerged in the field of domestic linguistics research driven by intelligent technology, and their burst duration and intensity intuitively reflect the evolution of research hotspots, as shown in Figure 6.

From a temporal perspective, between 2004 and 2009, keywords such as 'natural language' and 'natural language understanding' emerged first, with intensities of 2.63 and 1.21 respectively. This marked the early stage of intelligent technology intervention in linguistics, focusing on basic language processing, supported by early intelligent technologies, promoting the initial integration of language research and technology, and laying the foundation for subsequent paradigm shifts. Between 2012 and 2016, keywords such as 'data mining' and 'applied linguistics' continued to emerge, with 'applied linguistics' reaching an intensity of 1.55, reflecting the integration of data-driven approaches driven by big data and machine learning technologies. This not only expanded the dimensions of applied research but also promoted quantitative and empirical research in related fields.

Since 2021 – 2025, the burst intensity of "language intelligence", "interdisciplinary studies" and "embodied cognitive linguistics" has increased significantly, with the burst intensity of "language intelligence" reaching 2.26, and the burst cycles of the latter two extending to 2025. In this period, the new generation of intelligent technologies has driven the transformation of research paradigms, interdisciplinary integration has become a core trend, and the burst of embodied cognitive linguistics marks the upgrade of research from technological application to theoretical innovation, which has become a current research hotspot.

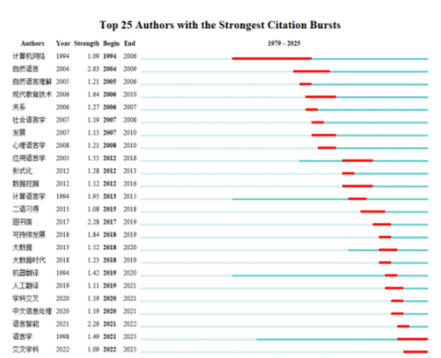

Figure 6 Keyword Prominent Network Knowledge Graph

4.4 Keyword Timeline Analysis

To reveal the evolutionary trajectory of domestic linguistic research paradigms driven by intelligent technologies, this study used CiteSpace to construct a keyword timeline map. In Figure 7, keywords of the same cluster are placed on the same horizontal line, the node size is positively correlated with the keyword frequency, and the connecting lines represent the co-occurrence relationships between keywords.Based on the map, the integrated research of the two fields can be divided into three stages: the technology introduction stage (1994 – 2005), which focused on foundational fields such as computational linguistics and machine translation and achieved preliminary integration between intelligent technologies and linguistics; the theoretical deepening stage (2005 – 2015), with the rise of keywords such as natural language processing and the research focus shifting to theoretical construction and practical expansion; the paradigm transformation stage (2015 – 2025), with the emergence of keywords such as generative artificial intelligence and the research focusing on human-computer collaboration and other frontier topics.

From the perspective of development trends, future research will present three core characteristics: first, the core connections will continue to strengthen, and the core network of 'artificial intelligence—linguistics—natural language processing' will further deepen, becoming the main driving force for the transformation of research paradigms; second, application scenarios will continually expand, with research extending to social application areas such as language governance and intelligent education, highlighting the practical value of linguistics; third, theoretical innovations will continue to emerge, and the bidirectional interaction between large language models and linguistics will promote technological breakthroughs and theoretical innovations, opening new paths for research and applications in the intelligent era.

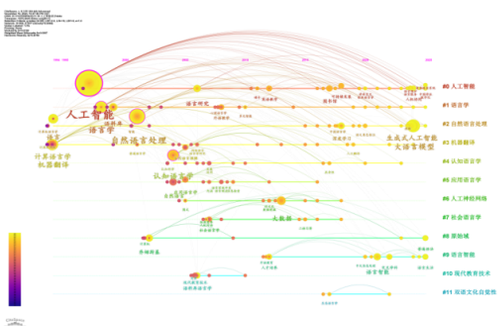

Figure 7 Keyword Clustering Timeline Map

5 Conclusion

Based on the CNKI database, this study conducted bibliometric and visual analysis using CiteSpace and found that domestic linguistic research driven by intelligent technologies keeps pace with the digital era and presents diverse and interdisciplinary development characteristics. This field has gone through three developmental stages: preliminary exploration, gradual development and rapid growth. The number of publications has witnessed explosive growth since 2023, and a core author group with Liu Haitao and Li Yuming as the core has been formed. Research is primarily driven by institutions of diverse types, establishing a research focus system centered on artificial intelligence and natural language processing, with current emphasis on cutting-edge directions like linguistic intelligence.

This study also identified several problems in this field, including weak collaboration among core authors, low efficiency of institutional coordination, superficial application of intelligent technologies and low-level repetitive research. To solve these problems, future research efforts should strengthen academic collaboration among core authors, establish cross-institutional collaborative platforms, deepen the integration of linguistics with related disciplines, and focus on in-depth theoretical research in frontier fields to promote the high-quality innovative development of research paradigms in this field.

About the Author

Qin Guo, Gender: Female, Education: postgraduate, Research Direction: International Chinese Language Education.

References

[1] WU J M, ZHU J X. Two-way interaction between large language models and linguistics: technological empowerment and theoretical feedback [J]. Foreign Language World, 2025(06):52-58.

[2] YAN C Y. Research on Chinese language teaching from the perspective of applied linguistics: a review of Contemporary Applied Linguistics[J]. Language Planning, 2021(06):86.

[3] ZHENG W, LIU J J. A multi-dimensional study of language use in the era of artificial intelligence and big data[J]. Contemporary Rhetoric, 2024 (02):45-59.

[4] HUANG L H, CHE Y R. Development and reflections on artificial intelligence empowering gerontolinguistics[J]. Journal of Yunnan Normal University (Philosophy and Social Sciences Edition), 2024,56(05):33-43.

[5] WAN Y K, LI X, HE S. Visualization analysis of modern logistics research in China based on CiteSpace[J]. Logistics Research, 2025(05):66-72.

[6] CHEN Y, CHEN C M, LIU Z Y, et al. Methodological functions of CiteSpace knowledge mapping[J]. Studies in Science of Science, 2015,33(02): 242-253.

[7] YANG Y. Visualization analysis of research on the elderly sports in China based on CiteSpace[J]. Bulletin of Sport Science & Technology, 2025,33 (09):84-88+115.

[8] WU J H, ZHU X M. Bibliometric analysis of domestic research papers on cognitive linguistics[J]. Journal of Harbin University, 2013,34(09):77-81.

[9] YANG Y N. Linguistics and artificial intelligence: the next decade[J]. Foreign Language Teaching Theory and Practice, 2024(03):1-9.

[10] CHEN Y, WANG Z Q, ZHANG L W. Research on Chinese innovation management academic groups from the perspective of social network[J]. Studies in Science of Science, 2015,33(07):985-992.

[11] WU J M, ZHU J X. Two-way interaction between large language models and linguistics: technological empowerment and theoretical feedback[J]. Foreign Language World, 2025(06):52-58.