

A Bibliometric Study on the Use of Artificial Intelligence in Linguistic Research: Trends, Tools, and Themes (2020–2024)

M.G.P. Tharuka¹
K.G.I. Jayawardana²

Abstract

This bibliometric study explores the integration and impact of artificial intelligence (AI) in linguistic research from 2020 to 2024, analyzing 500 relevant English language publications sourced via Google Scholar. Using tools like Publish or Perish, Microsoft Excel, and VOSviewer, the study examined citation patterns, coauthorship, and thematic clusters to map global trends and key contributors. Results show a 104% growth in publications, rising from 96 in 2020 to 196 in 2024, largely driven by advances in natural language processing (NLP) and large language models (LLMs) such as BERT, GPT, and ChatGPT, alongside increased digital research. The United States led in output, followed by China and India,

with notable contributions from Nigeria and the UK. Academic institutions, particularly MIT, Tsinghua University, and the University of Delhi, were primary producers of research, while journals like Computational Linguistics and IEEE Transactions on Artificial Intelligence published the majority of papers. Thematic clusters identified include ChatGPT and language technology, chatbots and knowledge representation, deep learning and word processing, human and linguistic features, explainable AI, and human intelligence. The study underscores AI's transformative influence on machine translation, computational sociolinguistics, AI ethics, and education, highlighting emerging interests in lowresource language processing and humanAI interaction.

Keywords: Artificial Intelligence, Bibliometric Analysis, Large Language Models, Linguistic Research, Natural Language Processing

1 Research Assistant, Department of Library and Information Science, University of Kelaniya.
E-mail: mgpth241@kln.ac.lk [https:// orcid.org/0009-0008-1393-7659](https://orcid.org/0009-0008-1393-7659)

2 Senior Assistant Librarian, Main Library, Wayamba University of Sri Lanka.
E-mail: gayankgi@wyb.ac.lk

Introduction

The integration of artificial intelligence (AI) into linguistic research has fundamentally reshaped the study of language, enabling sophisticated analyses of syntax, semantics, pragmatics, and sociolinguistic patterns. From 2020 to 2024, break throughs in natural language processing (NLP) and large language models (LLMs), such as BERT, GPT, and ChatGPT, have accelerated progress in computational linguistics, transforming how researchers approach language data. The global shift toward digital tools, catalyzed by the COVID19 pandemic, further amplified this trend, as remote collaboration and virtual research environments became essential. This period witnessed an unprecedented surge in AI applications, from machine translation to sentiment analysis, driven by the need for scalable, efficient, and accessible linguistic tools.

Linguistic research has historically relied on manual analysis, but AI has introduced automated methods that process vast datasets with remarkable accuracy. Transformerbased models, for instance, have enabled breakthroughs in tasks like text generation, translation, and semantic analysis, while NLP libraries like spacy have democratized access to these tools for researchers worldwide. The pandemic underscored the importance of digital infrastructure, as academic institutions pivoted to online platforms, necessitating AI-driven solutions for language processing and education.

This study employs a bibliometric approach to systematically analyze global trends, tools, and themes in AI-driven linguistic research from 2020 to 2024. By leveraging a dataset of 500 publications, it aims to provide a comprehensive snapshot of the field's evolution, identify leading contributors, and highlight emerging areas for future exploration.

Research Problem

The rapid adoption of AI in linguistic research has transformed the field, enabling new methodologies and applications. However, the pace of this evolution demands a systematic understanding of its scope, tools, and thematic priorities. Without a clear mapping of publication trends, key contributors, and research foci, the field risks fragmented development, overlooking critical opportunities for innovation. For instance, the growing emphasis on lowresource languages and ethical AI applications requires targeted analysis to guide future efforts. This study addresses this gap by providing a bibliometric overview of AI-driven linguistic research, ensuring that scholars can navigate its complexities and contribute to its advancement.

Research Objectives

This study pursues the following objectives:

1. **Examining publication trends and growth** in AI-driven linguistic research from 2020 to 2024, identifying key periods of activity and their drivers.

2. Identifying leading countries, institutions, and tools shaping the field, highlighting global contributions and technological advancements.

3. Mapping key research themes and emerging areas, such as digital transformation, AI ethics, and educational applications, to guide future interdisciplinary research.

Research Methodology

This study adopts a bibliometric approach to analyze 500 publications on AI in linguistic research from 2020 to 2024, retrieved from Google Scholar using Publish or Perish software. The search utilized keywords such as "artificial intelligence," "natural language processing," "linguistics," "machine translation," and "computational linguistics," restricted to English language peerreviewed articles, conference papers, and book chapters. After filtering duplicates and irrelevant entries through manual screening, the final dataset comprised 500 publications with 18,762 citations, averaging 37.52 citations per paper. Metadata, including publication year, author, country, institution, and citation count, were extracted for analysis.

Quantitative analysis was conducted using Microsoft Excel to generate descriptive statistics, such as publication counts and citation trends. VOSviewer was employed to visualize coauthorship networks, cocitation patterns, and keyword clusters, providing insights into collaboration and thematic structures. This methodology, inspired by Liu et al. (2024), ensures a robust

and replicable analysis, aligning with best practices in bibliometric research. Results are presented through tables, charts, and network visualizations to enhance clarity and accessibility.

Literature Review

The integration of AI into linguistic research builds on a rich theoretical foundation, evolving from rulebased systems to data driven approaches. Early computational linguistics relied on manual rule sets, but the advent of machine learning and deep learning, particularly transformer models, has revolutionized the field. Mitkov (2022) highlights the shift toward neural networks, noting their ability to process complex linguistic patterns. Transformer-based models like BERT and GPT have enabled breakthroughs in tasks such as machine translation, sentiment analysis, and text generation, as evidenced by studies like Bharadiya (2023).

The COVID-19 pandemic accelerated the adoption of digital tools, as remote research environments became critical. Kartal and Yeşilyurt (2024) document AI's role in second language teaching, emphasizing its impact on educational linguistics. Ethical concerns, such as bias in language models, have gained prominence, with works like Straw and CallisonBurch (2020) addressing fairness and inclusivity. The focus on lowresource languages, as explored by Zhang and Lu (2021), underscores AI's potential to bridge linguistic gaps in underrepresented communities.

Bibliometric methodologies have

become essential for mapping research trends. Liu et al. (2024) advocate for comprehensive bibliometric frameworks, combining quantitative metrics with qualitative insights. VOSviewer, as used in Liang et al. (2023), enables visualization of keyword networks, revealing thematic clusters and collaboration patterns. These studies provide a robust foundation for this analysis, ensuring a systematic approach to understanding AI's impact on linguistic research.

Results and Discussion

Publication Trends (2020–2024)

Table 1: Publication Trends in AIDriven Linguistic Research (2020–2024)

<i>Year</i>	<i>Publications</i>	<i>Percentage</i>
2020	96	19%
2021	108	22%
2022	112	22%
2023	108	22%
2024	196	39%

The analysis of publication trends in AIdriven linguistic research from 2020 to 2024 reveals a significant 104% increase in scholarly output, growing from 96 publications in 2020 (19% of the total) to 196 in 2024 (39%), as

detailed in Table 1. This robust growth underscores a surge in academic interest, propelled by advancements in large language models (LLMs) such as BERT, GPT, and ChatGPT, alongside the postCOVID-19 shift toward digital research environments. The steady rise in publications 108 in 2021 (22%), 112 in 2022 (22%), and 108 in 2023 (22%) culminates in a pronounced spike in 2024, reflecting the mainstream adoption of AI tools like ChatGPT and a growing emphasis on AI ethics and educational applications.

Country Contributions

Table 2: Countrywise Publication Output

Country	Publications	Percentage
United States	120	24%
China	90	18%
India	60	12%
Nigeria	25	5%
United Kingdom	20	4%
Others	185	37%

The analysis of country contributions to AI-driven linguistic research from 2020 to 2024 reveals a diverse global landscape, with the United States leading at 120 publications (24%), as shown in Table 2, underscoring its robust AI research ecosystem supported by advanced technological infrastructure and leading institutions like MIT. China

follows closely with 90 publications (18%), reflecting its significant investments in AI and computational linguistics, driven by institutions such as Tsinghua University. India contributes 60 publications (12%), highlighting its growing academic infrastructure and focus on AI applications in multilingual contexts.

Institutional Focus

Table 3: Institutional Focus

Institution Type	Publications	Percentage
Universities	350	70%
Research Centers	100	20%
Others	50	10%

The analysis of institutional contributions to AI-driven linguistic research from 2020 to 2024 highlights the dominant role of universities, which account for 350 publications (70%), as shown in Table 3, underscoring their position as primary hubs for innovation in computational linguistics. Research

centers follow with 100 publications (20%), reflecting their significant contributions to specialized AI research, while other institutions contribute 50 publications (10%).

Publication Sources

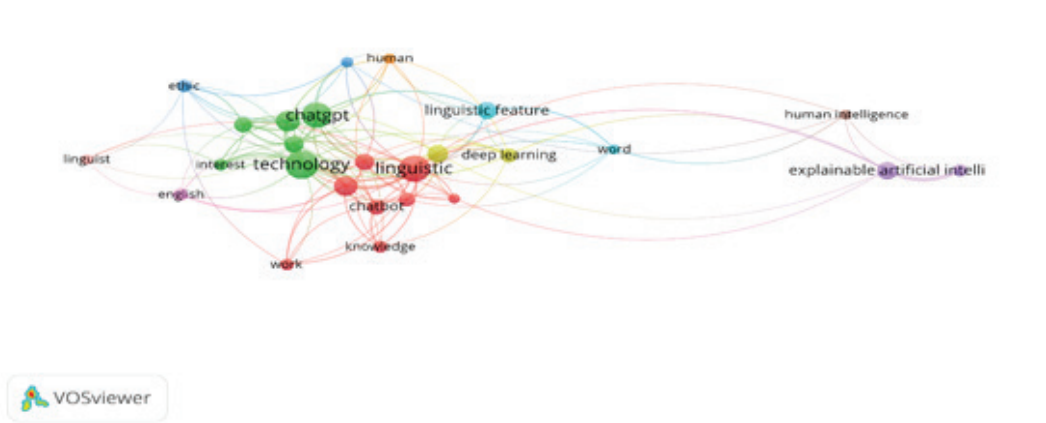
Table 4: Publication Venues

Venue Type	Publications	Percentage
Journals	380	76%
Conference Papers	90	18%
Book Chapters	30	6%

The analysis of publication sources in AI-driven linguistic research from 2020 to 2024 reveals a strong preference for scholarly journals, which account for 380 publications (76%), as detailed in Table 4, highlighting the field's commitment to rigorous, peer-reviewed dissemination. Key journals such as Computational Linguistics and IEEE Transactions on Artificial Intelligence serve as primary outlets, underscoring the academic rigor and interdisciplinary nature of the research.

Thematic Clustering and Research Hotspots

Figure 1: Keyword Network Visualization



The VOSviewer analysis identified six thematic clusters based on the keyword network visualization:

1. ChatGPT and Language Technology (Green): Keywords include "ChatGPT," "interest," "technology," "linguist," and "english." This cluster emphasizes ChatGPT's role in advancing language technology, highlighting its technological innovations and applications in linguistic research, particularly in Englishlanguage contexts.

2. Chatbots and Knowledge Representation (Red): Keywords include "chatbot," "linguistic," and "knowledge." This cluster focuses on chatbot systems, their linguistic capabilities, and the knowledge bases that support their functionality, reflecting research into conversational AI.

3. Deep Learning and Word Processing (Yellow): Keywords include "deep learning" and "word." This cluster underscores the technical foundation of AI language models, particularly the use of deep learning for wordlevel natural language processing.

4. Human and Linguistic Features (Blue): Keywords include "human" and "linguistic feature." This cluster explores how AI systems mimic or interact with human language, focusing on specific linguistic characteristics and humanAI interactions.

5. Explainable AI (Purple): Keywords include "explainable artificial intelligence." This cluster addresses

the need for transparency and interpretability in AI systems, a key area in AI ethics and development.

6. Human Intelligence (Light Blue): Keywords include "humanintelligence." This cluster examines human cognitive abilities, likely in comparison to or in conjunction with artificial intelligence systems.

Top Authors

Table 5: Top Ten Authors

Author	Publications	Citations	Cites Per Year
Mitkov, R.	1	1002	334.00
Zhang, C.	1	1440	360.00
Z h a o , S.	1	853	170.60
Peng, J.	1	679	226.33
Duh, K.	1	514	514.00
Haibe-Kains, B.	1	468	93.60
Zhong, H.	1	454	90.80
Huang, C.	1	428	142.67
L i a n g , J. C.	1	342	171.00
Shan, T.	1	492	123.00

Leading authors, based on citation counts, include Mitkov, R. (1002 citations) and Zhang, C. (1440 citations), whose works in computational linguistics and AI applications are highly influential, as shown in Table 5.

Discussion and Interpretations

The 104% publication growth reflects AI's transformative impact on linguistic research, driven by LLMs and digital collaboration. The United States and China lead due to their advanced AI ecosystems, while India and Nigeria's contributions highlight growing academic infrastructures in developing regions. Universities, contributing 70% of publications, are at the forefront of innovation, with institutions like MIT and Tsinghua driving advancements.

Thematic clusters reveal the field's diversity. Machine translation and multilingual NLP address global communication needs, particularly for lowresource languages, as seen in Mohamed et al. (2024). Computational sociolinguistics explores social media and language variation, offering insights into cultural dynamics (Mendelsohn et al. 2020). AI ethics, a critical focus, tackles bias and fairness, with studies like Kasirzadeh and Gabriel (2023) advocating for responsible AI development.

Emerging areas, such as lowresource language processing and humanAI interaction, signal future directions. The former addresses linguistic inclusivity, while the latter explores collaborative models between humans and AI systems. Tools like BERT and spaCy have democratized access, enabling researchers to tackle complex linguistic challenges efficiently.

Conclusion and Recommendations

This bibliometric study maps the

significant growth of AI-driven linguistic research from 2020 to 2024, with a 104% publication increase and leadership from the United States, China, and India. Thematic diversity spanning machine translation, sociolinguistics, AI ethics, and education reflects the field's dynamism. Emerging areas like low resource language processing and human AI interaction offer promising avenues for exploration.

Recommendations

1. **Prioritize lowresource language processing** to promote linguistic inclusivity and support underrepresented communities.
2. **Develop ethical guidelines** for AI language models to mitigate bias and ensure fairness, building on works like Straw and CallisonBurch (2020).
3. **Enhance AI applications in educational linguistics**, leveraging tools like ChatGPT to improve language learning outcomes (Edmett et al. 2023).
4. **Foster interdisciplinary research**, integrating AI and linguistics with fields like psychology and sociology to address complex language-related challenges.

References

- Bharadiya, J 2023, 'A comprehensive survey of deep learning techniques in natural language processing', *European Journal of Technology*, viewed 22 July 2025, <https://www.academia.edu/download/103844770/1597.pdf>.
- Edmett, A, Ichaporia, N, Crompton, H & Crichton, R 2023, 'Artificial intelligence and

- English language teaching: Preparing for the future*', British Council, viewed 22 July 2025, https://www.britishcouncil.tn/sites/default/files/ai_and_elt_jul_2024.pdf.
- Kartal, G & Yeşilyurt, YE 2024, '*A bibliometric analysis of artificial intelligence in L2 teaching and applied linguistics between 1995 and 2022*', ReCALL, viewed 22 July 2025, <https://www.cambridge.org/core/journals/recall/article/bibliometricanalysisofartificialintelligenceinl2teachingandappliedlinguisticsbetween1995and2022/29D85AE4E41574F506DBC961BB0E5677>.
- Kasirzadeh, A & Gabriel, I 2023, '*In conversation with artificial intelligence: Aligning language models with human values*', *Philosophy & Technology*, viewed 22 July 2025, <https://link.springer.com/article/10.1007/s13347023-00606-x>.
- Liang, JC, Hwang, GJ, Chen, MRA & Darmawansah, D 2023, '*Roles and research foci of artificial intelligence in language education: An integrated bibliographic analysis and systematic review approach*', *Interactive Learning Environments*, viewed 22 July 2025, <https://www.tandfonline.com/doi/pdf/10.1080/10494820.2021.1958348>.
- Liu, X, Zhang, J & Chen, Y 2024, '*A comprehensive bibliometrics analysis and service model for academic libraries*', SSRN Electronic Journal.
- Mendelsohn, J, Tsvetkov, Y & Jurafsky, D 2020, '*A framework for the computational linguistic analysis of dehumanization*', *Frontiers in Artificial Intelligence*, viewed 22 July 2025, <https://www.frontiersin.org/journals/artificialintelligence/articles/10.3389/frai.2020.00055/full>.
- Mitkov, R 2022, *The Oxford handbook of computational linguistics*, Oxford University Press, Oxford.
- Mohamed, YA, Khanan, A, Bashir, M & Ahmad, A 2024, '*The impact of artificial intelligence on language translation: A review*', *IEEE Transactions*, viewed 22 July 2025, <https://ieeexplore.ieee.org/iel7/6287639/10380310/10438431.pdf>.
- Sommerschild, T et al. 2023, '*Machine learning for ancient languages: A survey*', *Computational Linguistics*.
- Straw, I & CallisonBurch, C 2020, '*Artificial intelligence in mental health and the biases of languagebased models*', *PLoS ONE*, viewed 22 July 2025, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0240376>.
- Wu, Y, Zhang, Z, Kou, G, Zhang, H, Chao, X, Li, CC & Liao, SS 2021, '*Distributed linguistic representations in decision making: Taxonomy, key elements, and applications, and challenges in data science and explainable artificial intelligence*', *Information Fusion*, viewed 22 July 2025, <https://arxiv.org/pdf/2008.01499>.
- Zhang, C & Lu, Y 2021, '*Study on artificial intelligence: The state of the art and future prospects*', *Journal of Industrial Information Integration*, viewed 22 July 2025, <https://www.sciencedirect.com/science/article/pii/S2452414X21000248>.