Applied Data Science and Analysis
Vol.2025, **pp**. 83–93
DOI: https://doi.org/10.58496/ADSA/2025/005; ISSN: 3005-317X
https://mesopotamian.press/journals/index.php/ADSA

Review Article

The Evolution of Computational Linguistics: A Bibliometric Analysis of Research Trends from 1966 to 2023

Mohanad G. Yaseen [1,*], Hussein Alkattan [2,3], Laith Farhan [4,5]

[1] *Computer science, college of education, Al-Iraqia University, Iraq.*

[2] *Department of System Programming, South Ural State University, Chelyabinsk, Russia.*

[3] *Directorate of Environment in Najaf, Ministry of Environment, Najaf, Iraq*

[4] *School of Engineering, Manchester Metropolitan University, Manchester, M1 5GD, UK*

[5] *Department of Computer, College of Engineering, University of Diyala, Baquba 32001, Iraq*

ARTICLE INFO

Article History

Received 15 Jan 2025
Revised: 11 Feb 2025
Accepted 07 Mar 2025
Published 31 Mar 2025

Keywords

Computational Linguistics

Natural Language Processing Systems

Linguistics

Semantics

Bibliometric Analysis

ABSTRACT

This comprehensive bibliometric analysis scrutinizes the evolution of computational linguistics from 1966 to 2023, employing Scopus and specialized software. Findings unveil a noticeable surge in scientific output post-mid-2000s, coinciding with heightened citations, indicating a strong correlation between research output and impact. Key conferences and journals significantly disseminate research, while authorship patterns exhibit diverse scholarly contributions, depicting both consistent and sporadic impact. The identification of recurring themes emphasizes interdisciplinary convergence. Furthermore, the collaborative network analysis delineates dominant countries like the United States, the United Kingdom, and Germany, actively engaged due to prolific research output and extensive collaborations. This emphasizes varying country involvement, offering insights into future interdisciplinary collaboration for advancing computational linguistics.

1. INTRODUCTION

Computational linguistics[1], the convergence of linguistics and computer science, represents a multidisciplinary field dedicated to understanding human language through computational 8methods[2]. Evolving significantly over the years, this discipline encompasses various applications, including natural language processing, machine translation, information retrieval, and speech recognition. By employing computational techniques, computational linguistics aims to model and comprehend natural language, enabling machines to process, analyze, and generate human language[3].

Within the realm of scientific research, bibliometric analysis has emerged as a powerful quantitative tool used to assess and map the evolution of academic fields[4]. This methodology employs statistical and mathematical analyses to scrutinize patterns, trends, and relationships within scholarly literature. By examining citation counts, publication trends, collaboration networks, and recurring themes within academic publications, bibliometric analysis offers valuable insights into the dynamics and development of fields such as computational linguistics[5]. This study aims to explore the evolutionary trajectory of computational linguistics from 1966 to 2023 through a comprehensive bibliometric analysis. Utilizing extensive scholarly data available in databases like Scopus and leveraging specialized bibliometric software, this research seeks to uncover pivotal trends, influential sources, prolific authors, thematic foci, and collaborative networks within computational linguistics. In this pursuit, several hypotheses and research questions arise:

1. Is there a discernible correlation between the increase in scientific production over time and subsequent impacts, as indicated by citation rates?
2. What conferences, journals, or publications wield significant influence within computational linguistics?

*Corresponding author. Email: mohanad.yassen@aliraqia.edu.iq

3. How do author productivity and impact manifest over time? Are there specific authors demonstrating consistent impact or intermittent yet influential contributions?
4. What prevailing themes or key terms recur in computational linguistics literature, and how do these themes evolve over time?
5. How do countries and institutions collaborate within computational linguistics research? Are there prominent clusters or intermediary roles in the collaborative network?

The forthcoming sections of this paper will delve into a thorough analysis of these aspects. The methodology section will outline the data collection process, detailing the database, search strategy, and filters applied to assemble the dataset. Subsequently, the results section will comprehensively analyze temporal trends, influential sources, authorship patterns, recurring themes, and collaborative networks within computational linguistics. The discussion section will interrelate these findings, exploring their connections, implications, and potential future directions. Ultimately, the conclusion will summarize key outcomes and provide insights into the future trajectory of computational linguistics research, gleaned from the bibliometric analysis conducted herein. Through these sections, this study aims to offer a comprehensive illustration of the evolution and dynamics of computational linguistics.

2. METHODOLOGY

Database Selection and Justification: Scopus was chosen as the primary database for this bibliometric analysis due to its comprehensive coverage of scholarly literature across multiple disciplines, including computational linguistics. Scopus provides access to a vast collection of high-quality peer-reviewed journals, conference proceedings, and other scholarly publications, ensuring a comprehensive exploration of the evolution of computational linguistics from 1966 to 2023.

2.1. Search Strategy and Keyword Selection

Utilizing Scopus, a search was conducted using the keywords "Computational Linguistics" OR "Computational Linguistic" within the target article titles to retrieve relevant publications. This specific search string was chosen to encompass a broad spectrum of literature directly related to the field, ensuring a comprehensive overview of research trends.

2.2. Data Filtering

Filters were applied to refine the search results. The focus was solely on English-written publications to maintain consistency and eliminate language barriers, resulting in a total of 933 documents.

2.3. Bibliography Analysis and Processing

Upon uploading the bibliography file to the biblioshiny software for further analysis, the initial count of documents reduced to 924 due to encountered issues related to the structure of bibliographic data. RStudio, in conjunction with the R language, was employed to extract figures and tables using the biblioshiny package.

2.4. Challenges and Mitigation Strategies

The analysis revealed discrepancies in metadata completeness. While critical metadata elements such as author, document type, journal, publication year, title, and total citations exhibited satisfactory completeness, certain components, notably abstracts, affiliations, keywords, DOIs, and corresponding author details, displayed varying degrees of incompleteness[6]. Of significant concern were the absence of cited references and science categories for all publications, posing limitations on comprehensive analysis. To mitigate these issues, we utilized supplementary methods, not solely relying on missing metadata, to derive conclusive results despite the incompleteness observed in certain fields.

2.5. Summary of Missing Metadata

A summary table (Table 1) delineates the completeness of metadata elements, indicating critical areas with significant missing data. Notably, abstracts exhibited an acceptably moderate missing percentage (10.93%), whereas crucial elements such as cited references and science categories were entirely absent. This necessitated caution in the interpretation and analysis of the gathered data, emphasizing the importance of supplementary approaches to ensure the robustness of the findings.

TABLE I. COMPLETENESS OF METADATA ELEMENTS

Metadata	Description	Missing Counts	Missing %	Status
AU	Author	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent
SO	Journal	0	0.00	Excellent
LA	Language	0	0.00	Excellent
PY	Publication Year	0	0.00	Excellent
TI	Title	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent
AB	Abstract	101	10.93	Acceptable
C1	Affiliation	562	60.82	Critical
ID	Keywords Plus	592	64.07	Critical
DI	DOI	650	70.35	Critical
DE	Keywords	734	79.44	Critical
RP	Corresponding Author	741	80.19	Critical
CR	Cited References	924	100.00	Completely missing
WC	Science Categories	924	100.00	Completely missing

3. RESULTS

3.1 Main information

The analysis encompassed data spanning from 1966 to 2023, aggregating information from 586 sources including journals, books, and other scholarly publications, resulting in a dataset comprising 924 documents. Figure 1 illustrates an annual growth rate of 7.18%, showcasing the dynamic expansion of research in computational linguistics over time. Remarkably, the average age of the documents stands at 10.9 years, indicating a sustained relevance and ongoing citations within the field. Each document, on average, garnered approximately 5.001 citations, signifying a considerable impact within the computational linguistics domain[6].

Delving into the document contents, it's noteworthy that while references were absent or entirely missing from the dataset, keywords played a crucial role in indexing and understanding the research landscape. Authors contributed a significant corpus of keywords, with a total of 2349 identified through the Keywords plus (ID) category and 789 through Author's Keywords (DE). This robust set of keywords elucidates the breadth and depth of topics and themes explored within computational linguistics research.

In terms of authorship, 928 individuals contributed to the documents, reflecting a collaborative nature within the field. Interestingly, a subset of these individuals, 108 in total, authored single-authored documents, emphasizing the presence of independent research efforts alongside collaborative endeavors. The collaborative nature of the field is further highlighted by an average of 1.73 co-authors per document, showcasing a moderate level of collaboration among authors[7, 8].

Analyzing document types reveals the diverse scholarly output within computational linguistics. The majority of documents were articles (129), conference papers (162), and conference reviews (528), indicating the prevalence of these formats in disseminating research findings. Additionally, books, book chapters, editorials, and other publication types contributed to the multifaceted landscape of computational linguistics literature.

The data also indicates a low percentage of international co-authorships, standing at 7.576%. Despite this, the collaborative network displayed through co-authorships underscores the global engagement and exchange of ideas within the field. Overall, this comprehensive snapshot of the data provides valuable insights into the evolution, collaboration patterns, and scholarly output within computational linguistics over the examined timespan[9].

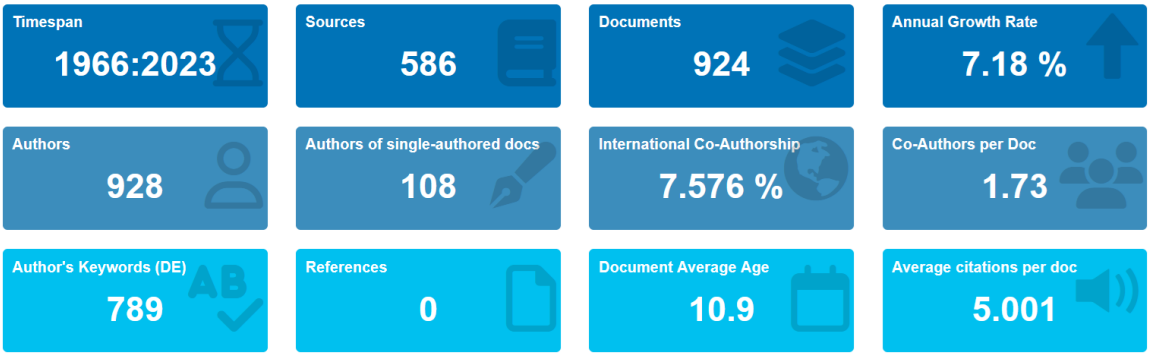

Fig.1. Main information

3.2 Annual Scientific Output and Citation Impact

The annual scientific production within the domain of computational linguistics reveals a dynamic trajectory illustrated in Figure 2. Over the period from 1966 to 2023, the scientific output experienced notable fluctuations. Commencing with minimal contributions in the late 1960s and early 1970s, there was a gradual increase in the number of articles published annually, with occasional fluctuations across subsequent decades. A significant surge in scientific production emerged notably around the mid-2000s, peaking in 2022 with 68 articles published, marking a noteworthy increase from previous years. However, in 2023, a slight decline was observed with 52 articles published[10].

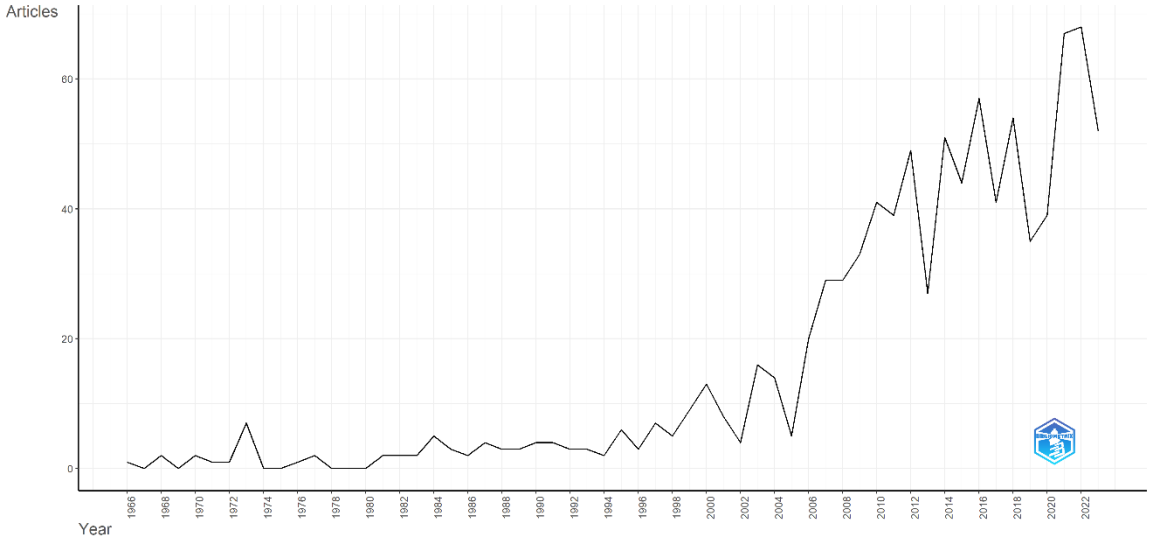

Fig. 2. Annual scientific production

In parallel, Figure 3 demonstrates the average citations per year for articles published within computational linguistics[11]. Initially, citation rates remained relatively low, with sporadic peaks and troughs in the earlier years. However, a discernible pattern emerged from the mid-2000s onwards, indicating an increase in average citations per year. Notably, the average citations per year peaked at 3.58 in 2008, followed by fluctuations ranging between 0.29 to 1.08 until 2023, where a decline in average citations per year was noted.

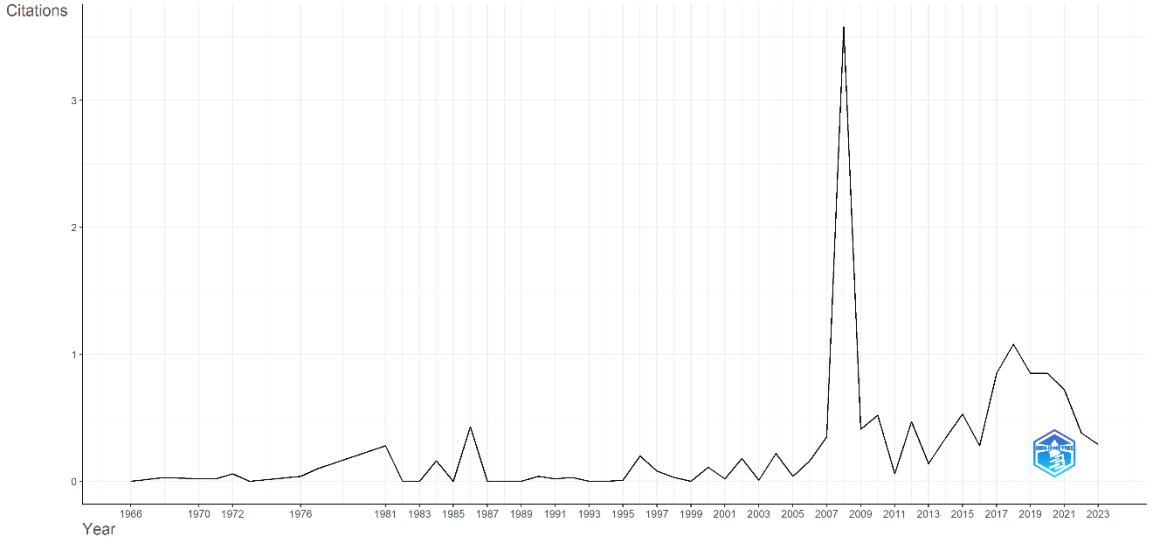

Fig. 3. Average citations per year

Comparing the trends in scientific production with average citations per year reveals an intriguing correlation[12, 13]. The rise in scientific output from the mid-2000s onwards seems to correspond with an increase in citations per year, indicating a potential correlation between increased research output and subsequent impact, as measured by citations. However, it's

essential to note the fluctuations in citation rates in recent years, despite a continued rise in scientific production. This suggests a more complex relationship between the quantity of publications and their subsequent impact within the field of computational linguistics, warranting further investigation into the underlying factors influencing citation patterns[14, 15]. The trends depicted in Figures 2 and 3 underscore the evolving nature of scientific output and its impact within computational linguistics. While a notable increase in publications has been observed, the correlation between scientific production and average citations per year showcases a nuanced relationship, demanding deeper exploration to understand the underlying dynamics influencing the field's scholarly impact.

3.3 Most relevant sources

The analysis of the most relevant sources within the field of computational linguistics, as depicted in Figure 2, sheds light on the prominent platforms contributing to scholarly discourse. PROCEEDINGS OF THE ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS emerges as the leading source, comprising 116 articles, showcasing its pivotal role as a cornerstone for disseminating research findings in the domain. This is followed closely by LECTURE NOTES IN COMPUTER SCIENCE, encompassing various subseries including Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, with a total of 77 articles, further solidifying its significance in the field. Further down the hierarchy of sources, CEUR WORKSHOP PROCEEDINGS stands out with 27 articles, showcasing its notable contribution to the computational linguistics landscape. Additionally, COMPUTATIONAL LINGUISTICS, the journal after which the field is named, holds a considerable presence with 13 articles, indicating its continued relevance and scholarly impact over time. Similarly, COMPUTATIONAL LINGUISTICS IN THE NETHERLANDS JOURNAL holds a notable presence, underscoring the localized contributions within the field.

Beyond journals, conference proceedings also play a substantial role, as evidenced by various platforms such as ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES, COMPUTATIONAL AND MATHEMATICAL LINGUISTICS, and PROCEEDINGS - INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, COLING. These platforms, each with 5 articles, signify the significance of conferences as key forums for disseminating cutting-edge research and fostering academic discourse within computational linguistics.

The distribution of articles across these sources showcases a diverse landscape of scholarly output, with journals, conference proceedings, and specialized publications contributing significantly to the field's knowledge repository. The dominance of specific platforms like the Annual Meeting of the Association for Computational Linguistics and Lecture Notes in Computer Science emphasizes their enduring influence and pivotal role in shaping the discourse and advancement of computational linguistics. Understanding the distribution and impact of articles across these sources provides valuable insights into the hierarchical importance and scholarly influence of various publication platforms within the field.

TABLE II. MOST RELEVANT SOURCES

Sources	Articles
PROCEEDINGS OF THE ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS	116
LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL INTELLIGENCE AND LECTURE NOTES IN BIOINFORMATICS)	77
CEUR WORKSHOP PROCEEDINGS	27
COMPUTATIONAL LINGUISTICS	13
COMPUTATIONAL LINGUISTICS IN THE NETHERLANDS JOURNAL	12
COMPUTERS AND THE HUMANITIES	6
KOMP'JUTERNAJA LINGVISTIKA I INTELLEKTUAL'NYE TEHNOLOGII	6
ACM INTERNATIONAL CONFERENCE PROCEEDING SERIES	5
COMPUTATIONAL AND MATHEMATICAL LINGUISTICS: PROCEEDINGS OF THE 5TH INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, COLING 1973	5
PROCEEDINGS - INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS, COLING	5

3.4 Sources Local Impact

The local impact of sources within the realm of computational linguistics is elucidated through various bibliometric indices, as illustrated in Figure 4. These indices, including h-index, g-index, and m-index, serve as valuable metrics in gauging the influence and prominence of sources within the field.

COMPUTATIONAL LINGUISTICS, a longstanding and seminal journal, emerges with a notable local impact as evidenced by its indices. With an h-index of 7, g-index of 13, and an m-index of 0.333, it demonstrates substantial influence

within the field, underpinned by 231 total citations (TC) across 13 publications (NP) since its start in 2003. This showcases its significant scholarly contribution and sustained influence over the years.
CEUR WORKSHOP PROCEEDINGS, despite having a lower h-index of 5, displays a noteworthy local impact reflected in an h-index, g-index, and m-index of 5, 9, and 0.294, respectively. Despite a lower total citation count of 87 across 27 publications since its initiation in 2007, its impact within computational linguistics is notable, underscoring its relevance in the scholarly discourse.
PROCEEDINGS OF THE ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, a renowned platform, exhibits an h-index, g-index, and m-index of 5, 9, and 0.116, respectively. Despite a higher citation count of 97 across 116 publications since 1981, its indices suggest a moderate local impact, indicating a balance between the volume of publications and their respective influence within the field.
LECTURE NOTES IN COMPUTER SCIENCE, comprising various subseries including Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, showcases an h-index of 4, g-index of 6, and an m-index of 0.111. Despite a substantial total citation count of 58 across 77 publications since its inception in 1988, its impact within computational linguistics is relatively moderate, reflecting a nuanced balance between publication volume and influence.
The analysis also includes non-traditional sources like PLOS ONE, exhibiting a relatively lower total citation count of 35 across 4 publications since 2015. Although its h-index stands at 3, it showcases a higher m-index of 0.333, indicating a notable local impact within the field despite a smaller publication volume.
The diverse range of sources assessed through these indices illuminates their varying local impacts within computational linguistics. While established platforms like COMPUTATIONAL LINGUISTICS wield considerable influence, other sources, despite smaller publication volumes, display noteworthy impacts within the field, emphasizing the multi-dimensional nature of scholarly influence and the importance of considering various metrics in evaluating sources' local impact.

Fig. 4. local impact of sources

3.5 Authors Production over Time

The analysis of authors' production over time within computational linguistics, as depicted in Figure 5, offers insights into the productivity and impact of individual authors across different years. The frequency of publications and their respective total citations (TC) alongside the citations per year (TCpY) provides a comprehensive view of authors' contributions to the field over time[16, 17].
Notably, author DALE R stands out in 2008, demonstrating prolific productivity with two publications amassing a substantial total citation count of 189, resulting in an impressive citations per year of 11.813. Similarly, author SIDOROV G in 2019 exhibited remarkable productivity with three publications, although with a lower total citation count of 12, resulting in a relatively lower citations per year of 2.400.
Comparing these prolific authors to others in various years reveals a range of productivity and impact levels. Authors like MEHLER A, BELL E, and KLAVANS JL contributed sporadically across different years with varying citation impacts. For instance, MEHLER A in 2017 and 2016, and BELL E in 2012, demonstrated lower productivity but relatively higher citations per year compared to their publication frequency.

However, certain authors like GELBUKH A and CHO K showcased variable productivity and impact across different years. GELBUKH A exhibited prolific publication frequency in some years (e.g., 2015, 2001) but with minimal to no citations, resulting in negligible citations per year. Similarly, CHO K exhibited sporadic contributions across years, occasionally garnering moderate citations per year despite a relatively lower frequency of publications.

The analysis of authors' production over time underscores the diversity in productivity and impact within computational linguistics. While some authors exhibited sporadic but impactful contributions, others demonstrated consistent productivity with varying citation impacts. These variations highlight the multidimensional nature of authors' contributions to the field, emphasizing the significance of considering both publication frequency and their corresponding citation impact to comprehensively assess authors' scholarly influence and productivity over time.

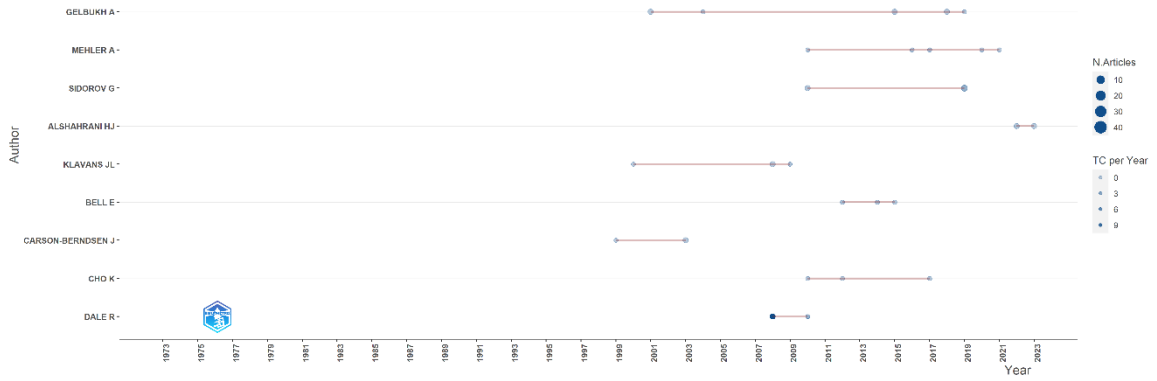

Fig. 5. Authors Production over Time

3.6 Most frequent words

The examination of the most frequent words utilized within the corpus of computational linguistics literature, as outlined in Figure 6, offers valuable insights into the prevalent themes and focal points within the field. These frequent words underscore the key areas of emphasis and recurrent topics shaping research discussions and scholarly discourse.

"Computational linguistics" emerges as the most prevalent phrase, occurring 238 times within the corpus. This prominence signifies the central focus and overarching theme within the literature, reflecting the core subject matter that forms the backbone of research endeavors within this domain.

Following closely, "natural language processing systems" appears 131 times, emphasizing the significant emphasis on systems and methodologies tailored for processing natural language, highlighting the paramount importance of this area within computational linguistics research. Additionally, the term "linguistics" surfaces frequently, underscoring the foundational aspect and interplay of linguistic theories and principles within computational linguistics.

Moreover, terms such as "semantics" and "syntactics" with 94 and 61 occurrences, respectively, reveal the persistent exploration and scrutiny of linguistic structures, meaning, and language representation, further emphasizing the interdisciplinary nature of computational linguistics encompassing linguistic theory and computational methods.

The occurrence of terms like "classification (of information)," "data mining," and "speech recognition" showcases the convergence of computational techniques with linguistic inquiry, highlighting the utilization of computational tools and methodologies for tasks such as information categorization, knowledge extraction, and speech analysis within the realm of natural language processing.

Furthermore, the recurrence of "natural languages" and "natural language processing" as frequent words reaffirms the field's focus on real-world languages and the development of computational methods to interact with and understand these languages, reflecting a continuous effort toward advancing language-related technologies and applications.

The diversity in the most frequent words elucidates the multifaceted nature of computational linguistics, encompassing linguistic theories, computational methodologies, and their intersection to address diverse challenges in understanding, processing, and utilizing natural language. These recurrent terms collectively encapsulate the core themes and focal points within computational linguistics literature, providing a comprehensive view of the prevalent areas of research and scholarly discourse within the field.

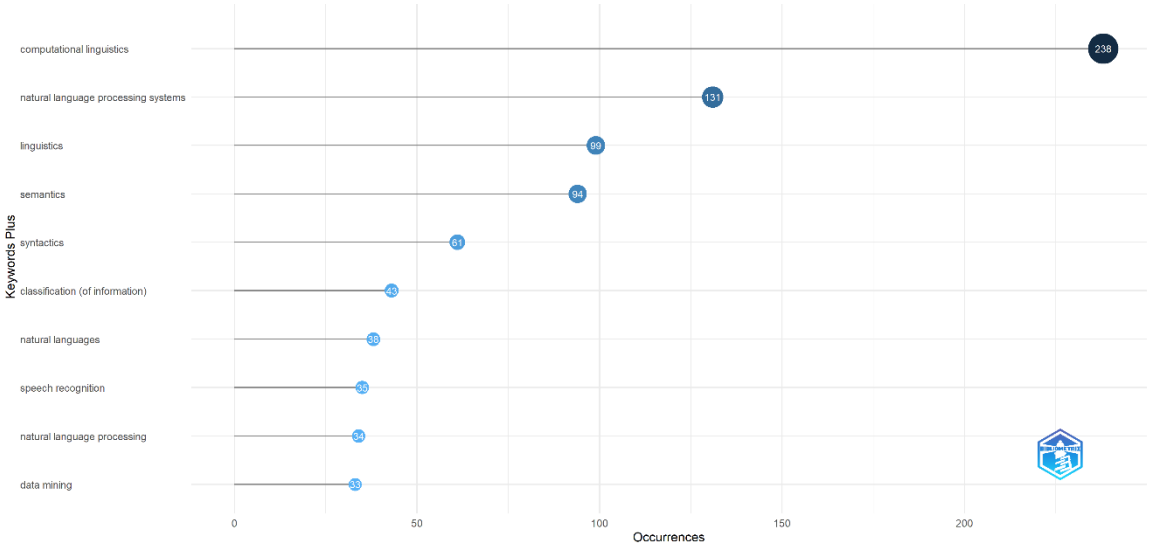

Fig. 6. Most frequent words

3.7 Countries collaboration network

The collaboration network among different countries within computational linguistics, as illustrated in Figure 7, offers valuable insights into the interconnectedness and engagement levels between nations. The clustering of countries based on collaboration patterns and their respective network metrics, including betweenness centrality, closeness centrality, and PageRank, underscores the diverse collaborative relationships and their significance within the field.

Cluster 1, highlighted in red, represents countries such as the USA, Canada, and Korea. These countries demonstrate higher betweenness centrality, indicating their pivotal roles as intermediaries or bridges connecting various research collaborations within computational linguistics. However, in terms of closeness centrality and PageRank, their values are comparatively lower, signifying a relatively lesser degree of closeness and influence within the collaborative network.

Cluster 2, depicted in blue, encompasses the United Kingdom, Germany, Italy, and Japan. These countries exhibit moderate betweenness centrality, suggesting their involvement in facilitating connections between different collaborative groups. The United Kingdom emerges with higher closeness centrality and PageRank, indicating a closer proximity and greater influence within the collaborative network compared to other countries in this cluster.

Cluster 3, represented in green, includes China and Australia. These countries showcase similar network metrics, indicating an equivalent degree of engagement and influence within the collaborative network. While their betweenness centrality remains low, signifying a lesser role as intermediaries, their closeness centrality and PageRank values suggest a considerable degree of proximity and influence within their respective collaborative circles.

Cluster 4, denoted in purple, encompasses France, Spain, the Netherlands, and Finland. These countries display notable betweenness centrality, emphasizing their intermediary roles in connecting diverse collaborative groups within computational linguistics. Additionally, the Netherlands emerges with higher closeness centrality and PageRank, indicating a closer proximity and greater influence compared to other countries in this cluster.

Lastly, cluster 5, highlighted in orange, consists of Saudi Arabia and Egypt. These countries exhibit high closeness centrality and PageRank values, suggesting their significant influence and proximity within their respective collaborative circles despite their lower betweenness centrality, implying a lesser role as intermediaries in the overall collaborative network.

The clustering and analysis of countries based on their collaborative network metrics within computational linguistics unveil a diverse landscape of collaborative relationships. These insights underscore the varying degrees of engagement, influence, and intermediary roles that countries assume within the global network of computational linguistics research and collaborations.

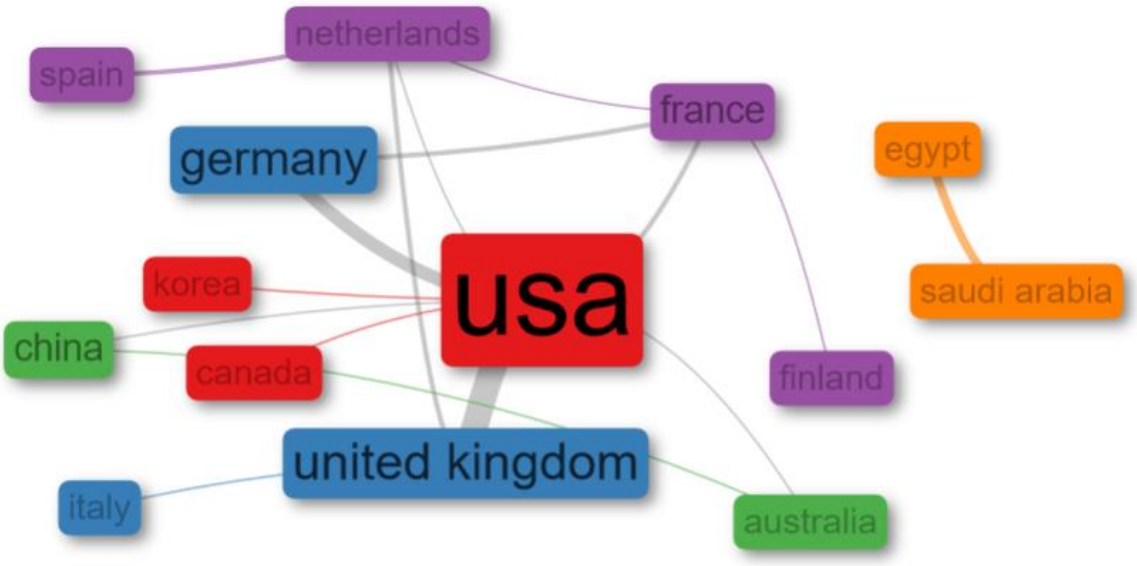

Fig. 7. Countries collaboration network

4. UNRAVELING THE COMPLEX EVOLUTION OF COMPUTATIONAL LINGUISTICS: INSIGHTS FROM MULTIDIMENSIONAL ANALYSIS

Examining the results holistically, the temporal trends in scientific production display a consistent increase in publications, notably from the mid-2000s onwards. This surge aligns with a subsequent rise in average citations per year, implying a potential correlation between heightened research output and subsequent impact within computational linguistics. However, the recent fluctuations in citation rates, despite sustained growth in scientific production, pose intriguing questions regarding the complex relationship between publication quantity and scholarly impact.

In tandem with the temporal trends, the most relevant sources, such as the PROCEEDINGS OF THE ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS and LECTURE NOTES IN COMPUTER SCIENCE, significantly influence the dissemination of research findings. These platforms serve as pivotal arenas for sharing advancements, reflecting their enduring influence and role in shaping the discourse within computational linguistics.

Authors' production over time elucidates varied levels of productivity and impact. Some authors demonstrate intermittent yet impactful contributions, while others exhibit consistent productivity with varying citation impacts. This variance underscores the diverse scholarly influences within computational linguistics, emphasizing the necessity of considering both publication frequency and citation impact to comprehensively evaluate authors' contributions.

Moreover, the analysis of the most frequent words uncovers recurring themes like "computational linguistics," "natural language processing systems," and "linguistics," highlighting core areas of focus within the field. These recurrent themes signify the convergence of linguistic theories and computational methodologies, showcasing computational linguistics' interdisciplinary nature in addressing language processing challenges.

Additionally, the collaboration network analysis unveils distinct clusters of countries based on their collaborative relationships. These clusters delineate varying levels of engagement and intermediary roles assumed by countries within the global network of computational linguistics research collaborations. The prominence of certain countries as intermediaries suggests their critical role in facilitating connections among diverse collaborative groups.

The correlations between these sections reveal a nuanced interplay between research output, influential sources, authors' contributions, recurring themes, and collaborative networks within computational linguistics. This holistic view accentuates the interdisciplinary evolution of the field, propelled by collaborations, recurrent research themes, influential publication platforms, and diverse scholarly influences.

Looking ahead, future research might delve deeper into the influence of prolific authors on prominent sources, investigate emerging interdisciplinary intersections, and explore the impact of collaborative networks on research output and citation patterns. By fostering interdisciplinary collaborations and studying evolving research themes, computational linguistics can continue to innovate and redefine language processing paradigms across various domains.

The correlations between these sections illustrate a dynamic and interconnected landscape within computational linguistics, portraying the collective contributions of researchers, influential platforms, recurring themes, and collaborative efforts in shaping the field's evolution.

5. CONCLUSION

The comprehensive bibliometric analysis conducted through this study illuminated several crucial facets within the realm of computational linguistics. Employing a meticulous methodology utilizing Scopus database, targeted keyword searches, and subsequent analysis through Biblioshiny software, this study delved into various dimensions, unveiling insightful outcomes. The temporal analysis revealed a discernible surge in scientific production, particularly from the mid-2000s, aligning with increased average citations per year, signifying a potential correlation between heightened research output and subsequent scholarly impact. However, the fluctuations in recent citation rates, despite sustained publication growth, present a noteworthy area for further exploration. The examination of relevant sources highlighted the enduring influence of prominent platforms such as conferences like the PROCEEDINGS OF THE ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS and established journals like LECTURE NOTES IN COMPUTER SCIENCE, showcasing their pivotal role in shaping discourse and disseminating research findings. Authors' production over time showcased diverse patterns, emphasizing the multifaceted scholarly contributions within the field, from sporadic yet impactful authors to consistently productive ones with varying citation impacts. The recurring themes in most frequent words underscored the interdisciplinary nature of computational linguistics, encapsulating linguistic theories and computational methodologies. Additionally, the collaboration network analysis unveiled clusters of countries exhibiting varying degrees of engagement, influence, and intermediary roles within the global research network. Looking ahead, future directions involve deeper investigations into the nuanced relationships between research output and impact, exploring emerging interdisciplinary intersections, and understanding the dynamics of collaborative networks. This comprehensive analysis offers a holistic understanding of the field's evolution, emphasizing the need for continued interdisciplinary collaboration and exploration to further propel computational linguistics into new frontiers of innovation and application.

Data Source

The data used for this bibliometric analysis was sourced from Scopus, comprising 924 documents obtained through a keyword search ("Computational Linguistics" OR "Computational Linguistic") in the target article title field. Metadata discrepancies were resolved using Biblioshiny software, resulting in the final dataset. The BibTeX file containing the bibliographic information used in this study is available for reference and verification at this link.

Funding

No financial support was received from any institutions or sponsors for the completion of this study.

Conflicts of Interest

The authors affirm that there are no conflicts of interest related to this research.

Acknowledgment

The authors extend sincere gratitude to the institution for offering unwavering moral support and encouragement during the course of this research.

References

[1] D. R. Radev, M. T. Joseph, B. Gibson, and P. Muthukrishnan, "A bibliometric and network analysis of the field of computational linguistics," *Journal of the Association for Information Science and Technology*, vol. 67, no. 3, pp. 683-706, 2016.

[2] M. CheshmehSohrabi and A. Mashhadi, "Using data mining, text mining, and bibliometric techniques to the research trends and gaps in the field of language and linguistics," *Journal of Psycholinguistic Research*, vol. 52, no. 2, pp. 607-630, 2023.

[3] G. Kartal and Y. E. Yeşilyurt, "A bibliometric analysis of artificial intelligence in L2 teaching and applied linguistics between 1995 and 2022," *ReCALL*, vol. 36, no. 3, pp. 359-375, 2024.

[4] A. Benseng and S. Tam, "Mapping the Evolution of Arabic Language Research: A Bibliometric Approach," *Mesopotamian Journal of Arabic Language Studies*, no. 2024, pp. 1-15, 2024.

[5] B. Mengliye, S. Hamroyeva, and O. Abdullayeva, "Scopus-based bibliometric analysis on corpus linguistics for the period of 2017-2021," in *E3S Web of Conferences*, 2023, vol. 413, p. 03008: EDP Sciences.

[6] X. Chen, H. Xie, G. Cheng, and Z. Li, "A decade of sentic computing: topic modeling and bibliometric analysis," *Cognitive computation*, vol. 14, no. 1, pp. 24-47, 2022.

[7] S. Yan and L. Zhang, "Trends and hot topics in linguistics studies from 2011 to 2021: A bibliometric analysis of highly cited papers," *Frontiers in Psychology*, vol. 13, p. 1052586, 2023.

[8] P. Crosthwaite, S. Ningrum, and M. Schweinberger, "Research trends in corpus linguistics: A bibliometric analysis of two decades of Scopus-indexed corpus linguistics research in arts and humanities," *International Journal of Corpus Linguistics*, vol. 28, no. 3, pp. 344-377, 2023.

[9] E. Arik, "A bibliometric analysis of linguistics in web of science," *J. Sci. Res.*, vol. 4, no. 1, pp. 20-28, 2015.

[10] M. A. Mohsen, H.-Z. Fu, and Y.-S. Ho, "A Bibliometric Analysis of Linguistics Publications in the Web of Science," *J. Sci. Res.*, vol. 6, no. 2, pp. 109-118, 2017.

[11] A. Perez-de-Arriluzea-Madariaga, "Computational Sociolinguistics: An Emerging Multidisciplinary Research Area," *International Journal of Applied Linguistics*, 2025.

[12] I. V. Savin and N. S. Teplyakov, "Using computational linguistics to analyse main research directions in Economy of Regions," *Экономика региона. 2022. Том 18, выпуск 2,* vol. 18, no. 2, pp. 338-352, 2022.

[13] X. Chen, H. Xie, F. L. Wang, Z. Liu, J. Xu, and T. Hao, "A bibliometric analysis of natural language processing in medical research," *BMC medical informatics and decision making*, vol. 18, pp. 1-14, 2018.

[14] M. Sharnin, "A Bibliometric Analysis of Trending Topics in the English Language Linguistics Research Literature," *Russian Linguistic Bulletin*, no. 12 (48), 2023.

[15] A. K. Kaleel, A. Assad, and M. Fyadh, "Quranic Studies Evolution: A Bibliometric Analysis from 1880 to 2023," *Mesopotamian Journal of Quran Studies*, vol. 2024, pp. 1-12, 2024.

[16] G. Chen, B. Tan, N. Laham, T. J. Tracey, S. Lapinski, and Y. Liu, "A Bibliometric Review of Natural Language Processing Applications in Psychology from 1991 to 2023," *Basic and Applied Social Psychology*, vol. 47, no. 2, pp. 105-119, 2025.

[17] X. Chen, H. Xie, X. Tao, F. L. Wang, D. Zhang, and H.-N. Dai, "A computational analysis of aspect-based sentiment analysis research through bibliometric mapping and topic modeling," *Journal of Big Data*, vol. 12, no. 1, p. 40, 2025.