Exploring Citation Dynamics in Computational Linguistics Research: A Citation Context Analysis Approach

Sandra M K and Rupesh Kumar A
Department of Library and Information Science
Tumkur University
Tumakuru. India
sandrasaj98@gmail.com,a.rupeshkumar@gmail.com
ORCID: https://orcid.org/0009-0001-2492-4406,
https://orcid.org/0000-0001-8787-6416

ABSTRACT

Purpose: *The present study aimed to explore the contextual and sectional relevance of citations gathered by the scholarly writings on computational linguistics research using Scite. Further, it investigates the correlation between supporting, contrasting and mentioning citations with Scopus citations.*

Methodology: *The study employed conventional and contextual citation analysis tools and techniques to gather the data. Scientific writings on computational linguistics were exported from Scopus database. Then the contextual citation data was gathered from smart citation index Scite.ai. Scite is an emerging classified citation database developed using machine learning, artificial intelligence and natural language processing. A total of 15607 documents with classified citation statements were included in the present study. The data was further analysed using MS Excel and statistical tests were conducted with the help of SPSS.*

Results: *The study found a gradual increase in the number of publications and citations. 15607 scholarly writings together garnered 3872 supporting, 375 contrasting and 343855 mentioning citations. That is 1.09% agreement, 0.11% disagreement and 97.23% research engagement is recorded on computational linguistic research. There is a clear domination of mentioning citations across all the years, indicating the attitude of scholars in neutrally citing a paper without supporting or contrasting the existing research. Contrasting citations were less compared to supporting and mentioning citations. But still they are relevant as it represents critical examination and inconstancies associated published research. Among the top ten journals with highest number of publications, Transactions of the Association for Computational Linguistics exhibits a lead in terms of supporting and contrasting citations. The Spearman's Rank Correlation test revealed a strong positive correlation between Mentioning citations and Scopus citations with Sperman's rho .839 and p-value .000. Further Scopus citations exhibit a weak positive correlation with supporting (rho value .336 and p-value .000) and contrasting citations (rho=.153, p-value=.000). The sectional citation analysis revealed accumulation of large number of citations (6031) in the "Other" section indicating the diversity followed in scientific communication.*

Conclusion: *The present investigation illustrated the contextual citation analysis using computational linguistics research. The study can be extended to author, journal and institutional level to unleash the real research impact. Contextual citation analysis is a novel approach that redefines the "all that glitters is gold" approach followed in the traditional citation analysis.*

Keywords: Citation Context Analysis, Computational Linguistics, Scite.ai

Received: 30 September 2025, Revised 21 October 2025, Accepted 23 October 2025

DOI: https://doi.org/10.6025/stm/2025/6/79-85

1. Introduction

Citation, the custom of quoting others' ideas, thoughts and concepts in scholarly writing practice. They acknowledge

an individual's work and help the readers to trace the source of information. A proper citation provides scientific prestige to the original work (Bahadoran et al., 2020). The purpose of citation ranges from negotiating the researcher's decision related to the inclusion of already published research ideas in their own investigation to measuring the research impact of a journal/ author/institution (Kostoff, 1998). Scientific community used citation as an effective and simplest way to trace the scholarly influence of research. Thus, citation analysis got popularised as a method of research impact assessment. Now there are a good number of citation-based metrics at author level (such as h-index, g-index etc) and journal level (such as CiteScore, Journal impact factor etc) that works purely on quantifying citations to analyse the overall impact of published research.

Zunde (1971) consider citation analysis as a qualitative and quantitative evaluation of authors, sources and institutions. Later the qualitative side of the citation analysis was obliterated and dominated by the quantitative approach followed in the conventional methods. Researchers pointed out this drawback of citation analysis and found that all citations do not share equal importance (Wang et al., 2020). This understanding gained the attention of researchers in classifying citations based on citation background (Dong & Schäfer, 2011), citing behaviour (Bornmann & Daniel, 2008), and context of citation (Wang et al., 2019). Considering all citations as rewards ignores the actual purpose of citing a paper. The author who cites a paper can express his agreement as well as disagreement with the published research. Thus, simply counting the number of citations received by a publication is giving an aerated image of research impact. Further if the impact is calculated simply based on the count of citations, it leads to the promotion of inconsistent, inaccurate and tendentious content (Hernández-Alvarez & Gomez, 2016).

A study conducted by Worrall & Cohn (2023) says that a citation can be favorable, or they can be critical. Thus, there exist a need for differentiating citations based on the context of citing a paper. This novel distinguished citation approach is popularly known as citation context analysis (CCA). CCA is not an alternative method for the traditional citation analysis. But it's a complementary technique to enhance the accuracy of the traditional citation-based metrics. Bornmann et al., (2019) conducted a citation concept analysis to understand the usefulness of concepts for other researchers illustrated by the case studies of classic books by Thomas S. Kuhn and Karl R. Popper. The study found the flow of knowledge by tracking citations. One of the major difficulties faced by the scholars in conducting citation context analysis is the classification of citations. This limitation was resolved by the invention of AI-based smart citation index Scite.ai (Scite, 2024). Scite is a database of 25 million full-text scientific articles and 880 million classified citation statements developed using machine learning and traditional document ingestion methods (Nicholson et al., 2021). Jangid & Rupak (2025) made an attempt to understand the distribution of supporting, contrasting and mentioning citations accrued by the Elsevier journals using Scite.ai. Another study by Sandra & Rupesh (2024) revealed high concentration of mentioning citations across publications on Meniere's disease.

1.1 Present Study

Computational linguistics is an amalgamation of computer technologies to analyse, synthesize and interpret written and spoken languages (Coursera, 2025). The aim of computational linguistics is to transcribe the natural transmission of information by modelling the speechmaker's production and listener's interpretation using a computer (Hausser, 2014). The application of computational linguistics includes Indexing and retrieval in textual databases Textual databases, Machine translation, Automatic text production etc. Hence, it's a highly influential research field with great social impact. Scholars examined the growth of linguistic computing using the traditional bibliometric and scientometric techniques (Alduais et al., 2025; Radev et al., 2016; Yaseen et al., 2025). Still there exist a need to identify the real research impact created by the scholarly literature by analysing the agreements and discourses associated with this field. Hence, the present study aims to execute a context-based citation analysis on computational linguistics research using Scite.ai.

2. Objectives

1. To understand the distribution of publications and citations across computational linguistics research articles

2. To analyse the context-based citation received by the scientific writings

3. To explore the correlation between supporting, contrasting and mentioning citations with Scopus citation

4.To examine the sectional relevance of computational linguistics research

3. Methdology

The data for the present investigation was collected using Scopus and Scite.ai. Firstly, research output on computational linguistics was exported from the Scopus database using the advanced query "AUTHKEY (computational linguistics) OR INDEXTERMS (computational linguistics) OR AUTHKEY (linguistic computing) OR INDEXTERMS (linguistic computing)". Further the document type and language were limited to Article and English respectively. A total of 16956 documents from

1968 to 2025 was exported with all the relevant bibliographic information. The data was screened and articles without digital object identifiers were excluded. The AI powered citation index, Scite.ai was used in the second stage to obtain contextual citation data. The article wise classified citation counts were gathered from Scite.ai using doi search. Scite.ai tracked a total of 15607 research outputs with classified citation statements. The sectional citation data was obtained from Scite using the "Section" filter. The data was further analysed using MS Excel and statistical tests were conducted with the help of SPSS.

4. Results and Discussion

4.1 Chronological distribution of Publications and Citations

Scientific writings on Computational linguistics from 1968 to 2025 was exported from Scopus database and have been tracked using Scite.ai for analysing the research trend. Table 1 displays the Chronological distribution of Publications and Citations. 15607 publications together accrued 353641 Scite citations and 519698 Scopus citations. The decadal analysis of scholarly literature exhibits a gradual growth in terms of publications and citations. Highest number of publications were produced during the period 2018-2025 indicating the recent and emerging trend on the topic. Further, the period 2008-2017 ranked top with 175180 Scite citations and 203659 Scopus citations. The decade, 1978-1987 displays an inactive decade of computational linguistic research with least number of published articles and accrued citations.

Year	Number of Publication	Scite Citations	Scopus Citations
1968-1977	15	415	1159
1978-1987	29	239	775
1988-1997	1885	41196	72308
1998-2007	3052	77551	121154
2008-2017	4754	175180	203659
2018-2025	5872	59060	120643
Total	**15607**	**353641**	**519698**

Table 1. Chronological distribution of Publications and Citations

4.2 Polarity based distribution of Citations across publications

Scite.ai classifies citations into supporting, contrasting and mentioning cites based on the purpose/context of citing a paper. Table 2 displays the decadal analysis of classified citations based on Scite data. 15607 scholarly writings together garnered 3872 supporting, 375 contrasting and 343855 mentioning citations. That is 1.09% agreement, 0.11% disagreement and 97.23% research engagement is recorded on computational linguistic research. There exists a stagnant growth in the number of supporting, contrasting and mentioning citations from 1988 to 2017. The decade 2008-2017 marked highest number of supporting (2163), contrasting (78) and mentioning (170408) citations. Thus, this period represents the decade of intense research engagement on computational linguistics. Supporting citations indicates the consensus among researchers; while contrasting citations points out the disagreement or discourses associated with the research output. Further, mentioning citations showcases the overall research visibility and scientific engagement on published research. There is a clear domination of mentioning citations across all the years, indicating the attitude of scholars in neutrally citing a paper without supporting or contrasting the existing research. The decline in the number of citations during 2018-2025 is may be because of the time taken for the accrual of citations and partial data.

4.3 Journal-wise citation context analysis

Computational linguistics research was disseminated through 1740 journals. Table 3 illustrates the top 10 journals with highest number of publications on computational linguistics. *Lecture notes in Computer Science* ranked first with 1242 articles followed by *IEE Access* (428 publications). Mentioning cites supremacy is visible across all the ten journals. Highest number of mentioning citations is obtained by the journal *Lecture notes in Computer Science (14907 mentioning citations)*. *Transactions of the Association for Computational Linguistics* exhibits a lead in terms of supporting and contrasting citations. Thus, it affirms that research is always subject to agreements and disagreements. The number of contrasting citations is negligible compared to supporting and mentioning cites. But still, they are important as they demonstrate the difference in opinion among the scholars. The journals *ACM Transactions on Asian and Low-Resource Language Information Processing* and *IEEE/ACM Transactions on Audio Speech and Language Processing* did not gain any contrasting citations affirming that the researchers' views accorded with the literature published through the journals.

Year	Supporting Citations	Contrasting Citations	Mentioning Citations
1968-1977	2	1	397
1978-1987	0	0	234
1988-1997	206	23	40027
1998-2007	763	80	75107
2008-2017	2163	193	170408
2018-2025	738	78	57682
Total	**3872**	**375**	**343855**
Percentage	**1.09**	**0.11**	**97.23**

Table 2. Distribution of supporting, contrasting and mentioning citations

Journal	No. of Titles	Supporting Citations	Contrasting Citations	Mentioning Citations
Lecture Notes in Computer Science	1242	86	9	14907
IEEE Access	428	10	2	2227
Transactions of the Association for Computational Linguistics	388	159	11	14632
Theoretical Computer Science	315	47	2	4320
Expert Systems with Applications	239	17	3	3534
ACM Transactions on Asian and Low-Resource Language Information Processing	219	5	0	677
IEEE/ACM Transactions on Audio Speech and Language Processing	180	22	0	2463
Physical Review A - Atomic, Molecular, and Optical Physics	176	131	10	9527
Computer Speech and Language	172	46	7	3290
Computational Linguistics	170	62	7	4805

Table 3. Journal wise distribution of publications and citations

4.4 Scite Citation v/s Scopus Citation

One of the objectives of the study was to explore the interrelationship between different contextual citations and Scopus citation counts. Hence, Correlation test was conducted. Kolmogorov-Smirnov normality test was performed on all the distributions of data, namely, supporting citations, contrasting citations, Mentioning citations and Scopus citation counts. The distributions deviated from the normal distribution. Thus, Spearman's Rank Correlation test was conducted. The results are presented in Table 4.

	Correlation between Scopus Citations and Contextual Citations		
	Supporting	**Contrasting**	**Mentioning**
Correlation Coefficient (rho)	336.	153.	839
Sig. value	.000	.000	.000

Table 4. Spearman's Rank Correlation

The test results revealed that a strong positive correlation exists between Mentioning citations and Scopus citations with Sperman's *rho =.839* (*p-value .000)*. A weak positive correlation exists between Supporting citations and Scopus citations (*rho=.336* and *p-value .000)*. Contrasting citations also exhibits a weak positive correlation (*rho=.153, p-value=.000*). So, it is evident that research visibility, agreement as well as discourse enhances the overall impact of research outputs and ensures future citations.

4.5 Sectional Citation Analysis

Scite.ai indexes citations in different sections of the scientific literature. Citation in each section of a research output indicates the sectional relevance of that paper. That is citation in the Introduction part specifies the strength of the study background or the theoretical structuring of the research. Citation in the methodology part is the indicator of well-built research design that can be followed for the future research. The use of existing findings for strengthening the upcoming results and discussion is evident with the citations in the analayis/discussion/results part. Hence, rather than simply quantifying citations, they have more insights to share.

Scite classifies citations in major five sections of a research article, namely Introduction, Methods, Results, Discussion and Other. Usually, scientific writings follow IMRaD format. Scite classifies all the citations that diverge from the standard structure as "Other". From figure 1, it may be observed that 6031 publications accrued citation in the "Other" section. Hence it proclaims that the structure of scientific investigation has obtained new dimensions and it deviates highly from the standard IMRaD structure. The high concentration of publications garnering citation in the "Other" section affirms the interdisciplinary approach followed in scholarly investigations in the field of Computational Linguistics. Further, it points out the diversity followed in structuring scholarly writings. Even though IMRaD is considered as a standard structure for the scientific writings, still scholarly communication channels accept diverse compositions of articles. Least number of publications (998) gained citation in the "Methods" section. This is a clear indication of less emphasize given to the published research design. Linguist's infrequently use the existing established methodologies followed in the computational linguistics research. Accordingly sectional analysis of citations shed light in to the highly interdisciplinary research area that focuses on conceptual frameworks.

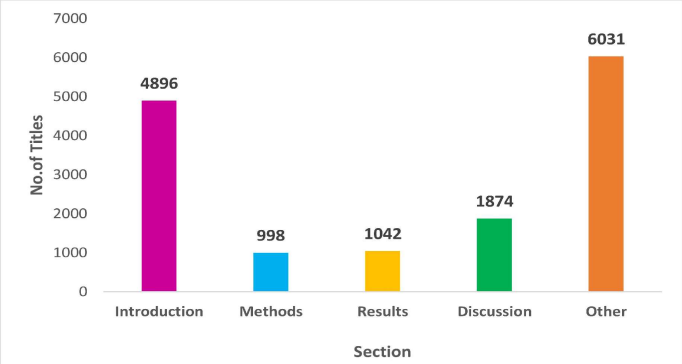

Figure 1. Sectional Citation Analysis

5. Conclusion

Contextual citation analysis discloses the hidden connection between cited and citing paper by analysing the authors view in adducing a scientific writing. Hence it goes beyond traditional citation counting metrics and unleashes the hidden citation potential. The gradual growth in the number of supporting, contrasting and mentioning citations gathered by the computational linguistics research points out the growing agreements and discourses on the topic. High volume of mentioning citations indicates the researcher's non-discriminatory approach in citing a paper. The number of contrasting citations were less compared to supporting and mentioning citations. But still its relevant as it represents critical examination and inconstancies associated published research. Further it leads to refining the existing knowledge. The sectional analysis points out dense accumulation of citations in the "Other" sections. It is because of the divergence of research writings from the standard IMRaD structure. Further this deviation indicates the diversified approach in communicating research findings. Also, the present investigation found statistically significant positive correlation between Supporting, Contrasting and Mentioning citations with Scopus citations. Hence agreements, discourses and engagement on published research impacts the accumulation of future citations.

6. Areas Of Further Research

Contextual citation analysis is a complementary tool that helps to enhance the efficiency of traditional citation-based metrics. The complex intellectual interconnections between scientific communication can be decoded by analysing its context, hence adding a qualitative touch to quantitative research impact assessment techniques. The present study solely focused on Computational linguistics research. Thus, it can be further extended to different subjects/disciplines, Journals, Institutions and Authors. Classified citations, such as supporting, contrasting and mentioning, helps the scholars to trace relevant study in a more efficient and effective manner. So there exist a need for understanding the effectiveness of contextual citation analysis in impactful literature reviewing. Further, the present investigation made use of AI based classified citation database Scite.ai. Our investigation tries to put forth the idea of incorporating classified citations in various citation databases to understand the real impact of research.

References

[1] Alduais, A., Yassin, A. A., Allegretta, S. (2025). Computational linguistics: A scientometric review. *Quality & Quantity*, *59*(5), 4097–4136. https://doi.org/10.1007/s11135-025-02138-2.

[2] Bahadoran, Z., Mirmiran, P., Kashfi, K., Ghasemi, A. (2020). The Principles of Biomedical Scientific Writing: Citation. *International Journal of Endocrinology and Metabolism*, *18*(2). https://doi.org/10.5812/ijem.102622.

[3] Bornmann, L., Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80. https://doi.org/10.1108/00220410810844150.

[4] Bornmann, L., Wray, B. K., Haunschild, R. (2019). Citation concept analysis (CCA): A new form of citation analysis revealing the usefulness of concepts for other researchers illustrated by exemplary case studies including classic books by Thomas S. Kuhn and Karl R. Popper. *Scientometrics*, *122*, 1051–1074. https://doi.org/10.1007/s11192-019-03326-2.

[5] Coursera. (2025, March 11). *What Is Computational Linguistics?* Coursera. https://www.coursera.org/in/articles/computational-linguistics.

[6] Dong, C., Schäfer, U. (2011). Ensemble-style Self-training on Citation Classification. In H. Wang & D. Yarowsky (Eds.), *Proceedings of 5th International Joint Conference on Natural Language Processing* (pp. 623–631). Asian Federation of Natural Language Processing. https://aclanthology.org/I11-1070.

[7] Hausser, R. (2014). *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41431-2.

[8] Hernández-Alvarez, M., Gomez, J. M. (2016). Survey about citation context analysis: Tasks, techniques, and resources. *Natural Language Engineering*, *22*(3), 327–349. https://doi.org/10.1017/S1351324915000388.

[9] Jangid, S., Rupak. (2025). AI-Powered Citation Classification: Understanding Contrasting and Supporting Citations in Scholarly Research. *Annals of Library and Information Studies*, *72*(3). https://doi.org/10.56042/alis.v72i3.19939.

[10] Kostoff, R. N. (1998). The use and misuse of citation analysis in research evaluation. *Scientometrics*, *43*(1), 27–43.

https://doi.org/10.1007/BF02458392.

[11] Nicholson, J. M., Mordaunt, M., Lopez, P., Uppala, A., Rosati, D., Rodrigues, N. P., Grabitz, P., & Rife, S. C. (2021). scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. *Quantitative Science Studies*, *2*(3), 882–898. https://doi.org/10.1162/qss_a_00146.

[12] Radev, D. R., Joseph, M. T., Gibson, B., Muthukrishnan, P. (2016). A bibliometric and network analysis of the field of computational linguistics. *Journal of the Association for Information Science and Technology*, *67*(3), 683–706. https://doi.org/10.1002/asi.23394.

[13] Sandra, M. K., Rupesh, K. (2024, January 10). *(PDF) Citation Context Analysis: A Case Study of Scholarly Literature on Meniere's Disease*. LISACON 2024:5th International Conference on Intelligent Libraries.
Scite. (2024). *Scite.ai*. Scite.Ai. https://scite.ai.

[14] Wang, M., Ren, J., Li, S., Chen, G. (2019). Quantifying a Paper's Academic Impact by Distinguishing the Unequal Intensities and Contributions of Citations. *IEEE Access*, *7*, 96198–96214. IEEE Access. https://doi.org/10.1109/ACCE- SS.201- 9.2927016.

[15] Wang, M., Zhang, J., Jiao, S., Zhang, X., Zhu, N., Chen, G. (2020). Important citation identification by exploiting the syntactic and contextual information of citations. *Scientometrics*, *125*(3), 2109–2129. https://doi.org/10.1007/s11192-020-03677-1.

[16] Worrall, J. L., Cohn, E. G. (2023). Citation Data and Analysis: Limitations and Shortcomings. *Journal of Contemporary Criminal Justice*, *39*(3), 327–340. https://doi.org/10.1177/10439862231170972.

[17] Yaseen, M. G., Alkattan, H., Farhan, L. (2025). The Evolution of Computational Linguistics: A Bibliometric Analysis of Research Trends from 1966 to 2023. *Applied Data Science and Analysis*, *2025*, 83–93.

[18] Zunde, P. (1971). Structural models of complex information sources. *Information Storage and Retrieval*, *7*(1), 1–18. https://doi.org/10.1016/0020-0271(71)90023-4.