



Supporting, Contrasting, and Mentioning Citations in Computational Linguistics: Insights from Scite and Scopus

Sandra M K, Rupesh Kumar A

Department of Library and Information Science

Tumkur University, Tumakuru- 572103

sandrasaj98@gmail.com, a.rupeshkumar@gmail.com

ORCID: <https://orcid.org/0009-0001-2492-4406>, <https://orcid.org/0000-0001-8787-6416>

ABSTRACT

Purpose: This study examines the contextual and sectional relevance of citations received by scholarly publications in computational linguistics using Scite. It further investigates the relationships between supporting, contrasting and mentioning citations and the corresponding Scopus citation counts.

Methodology: Both conventional and contextual citation-analysis techniques were employed. Scientific publications on computational linguistics were retrieved from the Scopus database. Contextual citation data were subsequently obtained from the smart citation index Scite.ai, an emerging classified citation database built with machine learning, artificial intelligence and natural language processing. A total of 15,607 documents containing classified citation statements were included. Data were analysed in Microsoft Excel and statistical tests were performed with SPSS.

Results: The analysis reveals a steady rise in both publications and citations. The 15,607 scholarly works collectively received 3,872 supporting, 375 contrasting and 343,855 mentioning citations corresponding to 1.09 % agreement, 0.11 % disagreement and 97.23 % neutral research engagement. Mentioning citations dominate every year, reflecting a tendency among scholars to cite neutrally without explicit endorsement or critique. Although contrasting citations are far fewer than supporting or mentioning citations, they remain significant because they signal critical scrutiny and inconsistencies in the published literature. Among the ten journals with the highest publication counts, Transactions of the Association for Computational Linguistics leads in both supporting and contrasting citations. Spearman's rank-correlation tests show a strong positive association between mentioning citations and Scopus citations ($\rho = .839$, $p = .000$), a weak positive association with supporting citations ($\rho = .336$, $p = .000$) and a still weaker positive association with contrasting citations ($\rho = .153$, $p = .000$). Sectional citation analysis indicates that the largest number of citations (6,031) fall into the "Other" category, underscoring the diversity of structural conventions in scientific communication.

Conclusion: The investigation demonstrates the value of contextual citation analysis applied to computational linguistics research. The approach can be extended to author-, journal- and institutional-level assessments to reveal genuine research impact. Contextual citation analysis offers a novel perspective that moves beyond the traditional assumption that "all that glitters is gold."

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

Received: 20 April 2026, **Revised:** 29 June 2026, **Accepted:** 14 July 2026

Copyright: DLINE

1. Introduction

Citation is the established practice of acknowledging the ideas, thoughts and concepts of others within scholarly writing. Proper citation confers scientific prestige on the original work (Bahadoran et al., 2020) and enables readers to locate the source of information. The purposes of citation range from justifying a researcher's decision to incorporate previously published ideas into a new investigation to measuring the research impact of a journal, author or institution (Kostoff, 1998). The scientific community has long regarded citation counts as a simple and effective proxy for scholarly influence. Consequently, citation analysis has become a widely adopted method of research-impact assessment. Numerous citation based metrics now exist at the author level (for example, the h-index and g-index) and at the journal level (for example, CiteScore and the Journal Impact Factor), all of which rely exclusively on the quantification of citations.

Zunde (1971) characterised citation analysis as both a qualitative and a quantitative evaluation of authors, sources and institutions. Over time, however, the qualitative dimension was overshadowed by a predominantly quantitative approach. Subsequent researchers have drawn attention to this limitation, noting that not all citations carry equal weight (Wang et al., 2020). This realisation has stimulated efforts to classify citations according to background (Dong & Schäfer, 2011), citing behaviour (Bornmann & Daniel, 2008) and citation context (Wang et al., 2019). Treating every citation as an unqualified reward ignores the actual purpose for which a paper is cited. An author may express agreement or disagreement with the cited work; therefore, simply tallying citations can produce an inflated picture of research impact. When impact is calculated solely on the basis of raw citation counts, the result may inadvertently promote inconsistent, inaccurate or tendentious content (Hernández-Alvarez & Gomez, 2016).

Worrall and Cohn (2023) observe that a citation may be favourable or critical. Differentiating citations according to the context in which they appear is therefore essential. This more nuanced approach is commonly termed citation context analysis (CCA). CCA is not intended to replace traditional citation analysis; rather, it complements existing metrics and improves their accuracy. Bornmann et al. (2019) applied citation concept analysis to classic works by Thomas S. Kuhn and Karl R. Popper, demonstrating how the flow of knowledge can be traced through citations. One of the principal obstacles to large scale CCA has been the labour intensive classification of citation statements. That barrier has been substantially lowered by the AI-powered smart citation index Scite.ai (Scite, 2024). Scite currently indexes more than 25 million full text scientific articles and 880 million classified citation statements, employing machine learning and conventional document-ingestion methods (Nicholson et al., 2021). Jangid and Rupak (2025) examined the distribution of supporting, contrasting and mentioning citations among Elsevier journals using Scite.ai. In a related study, Sandra and Rupesh (2024) found a high concentration of mentioning citations in the literature on Ménière's disease.

1.1 Present Study

Computational linguistics combines computer technologies with the analysis, synthesis and interpretation of written and spoken language (Coursera, 2025). Its central aim is to model the production of speech by a speaker and the interpretation of that speech by a listener through computational means (Hausser, 2014). Applications include indexing and retrieval in textual databases, machine translation and automatic text production. Because of these applications, computational linguistics is a highly influential research domain with considerable social impact. Previous studies have mapped the growth of the field with conventional bibliometric and scientometric techniques (Alduais et al., 2025; Radev et al., 2016; Yaseen et al., 2025). Nevertheless, a systematic examination of the agreements and disagreements surrounding published research is still lacking. The present study therefore undertakes a context-based citation analysis of computational

linguistics literature using Scite.ai.

2. Objectives

The study pursues four principal objectives:

1. To map the chronological distribution of publications and citations in computational linguistics research.
2. To analyse the context-based citations received by these scientific writings.
3. To explore the correlations between supporting, contrasting and mentioning citations and Scopus citation counts.
4. To examine the sectional relevance of computational linguistics research.

3. Methodology

Data were collected from two complementary sources: Scopus and Scite.ai. Research output on computational linguistics was first retrieved from Scopus by means of the advanced query “AUTHKEY (computational linguistics) OR INDEXTERMS (computational linguistics) OR AUTHKEY (linguistic computing) OR INDEXTERMS (linguistic computing)”. Document type was restricted to Article and language to English. The search yielded 16,956 documents published between 1968 and 2025, together with complete bibliographic metadata. Records lacking digital object identifiers were subsequently excluded. In the second stage, the AI-powered citation index Scite.ai was used to obtain contextual citation data. Article-level classified citation counts were retrieved via DOI search. Scite.ai located 15,607 research outputs that possessed classified citation statements. Sectional citation data were extracted with Scite’s “Section” filter. All data were processed in Microsoft Excel; statistical tests were conducted with SPSS.

4. Results and Discussion

4.1 Chronological Distribution of Publications and Citations

Scientific writings on computational linguistics spanning 1968–2025 were exported from Scopus and subsequently tracked in Scite.ai. Table 1 and Figure 1 present the chronological distribution of publications and citations. The 15,607 publications accumulated 353,641 Scite citations and 519,698 Scopus citations. Decadal analysis reveals a gradual increase in both publications and citations. The period 2018–2025 produced the largest number of publications, reflecting the recent expansion of the field. The decade 2008–2017 recorded the highest citation totals (175,180 Scite citations and 203,659 Scopus citations). By contrast, 1978–1987 constitutes an inactive decade, characterised by the fewest publications and the lowest citation counts.

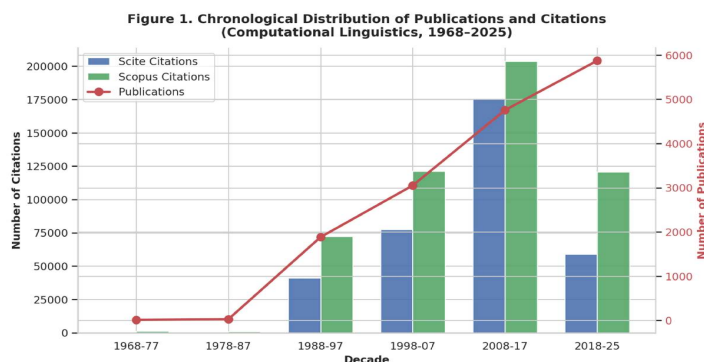


Figure 1. Chronological Distribution of Publications and Citations

Year	Number of Publications	Scite Citation	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 as supporting, contrasting or mentioning according to the purpose and context of the citation. Table 2 and Figure 2 display the decadal breakdown of these classified citations. The 15,607 scholarly writings together received 3,872 supporting, 375 contrasting and 343,855 mentioning citations, corresponding to 1.09 % agreement, 0.11 % disagreement and 97.23 % neutral research engagement. Supporting, contrasting and mentioning citations all increased steadily from 1988 to 2017. The decade 2008–2017 registered the highest numbers in every category (2,163 supporting, 193 contrasting and 170,408 mentioning). This period therefore represents a phase of intensive research engagement. Supporting citations indicate consensus among researchers, while contrasting citations highlight disagreement or discourse. Mentioning citations reflect overall research visibility and scientific engagement. Across every decade, mentioning citations predominate, suggesting that scholars frequently cite papers neutrally without explicitly endorsing or challenging the cited work. The decline observed for 2018–2025 is most likely attributable to the time lag required for citation accrual and to the incomplete nature of the most recent data.

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

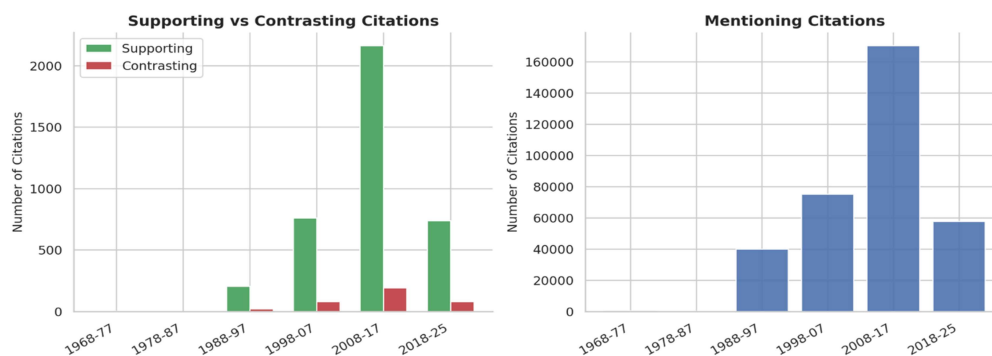


Figure 2. Distribution of Supporting, Contrasting and Mentioning Citations by Decade

4.3 Journal-Wise Citation Context Analysis

Computational linguistics research appeared in 1,740 journals. Table 3 lists the ten journals that published the largest numbers of articles. Lecture Notes in Computer Science ranks first with 1,242 articles, followed by IEEE Access (428 publications). Mentioning citations predominate in every journal. Lecture Notes in Computer Science also leads in absolute mentioning citations (14,907). Transactions of the Association for Computational Linguistics records the highest counts of both supporting and contrasting citations, confirming that research in the field is subject to both agreement and disagreement. Although the number of contrasting citations is modest relative to supporting and mentioning citations, they remain important indicators of divergent scholarly opinion. Two journals ACM Transactions on Asian and Low Resource Language Information Processing and IEEE/ACM Transactions on Audio, Speech, and Language Processing received no contrasting citations, implying broad consensus with the literature they published.

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

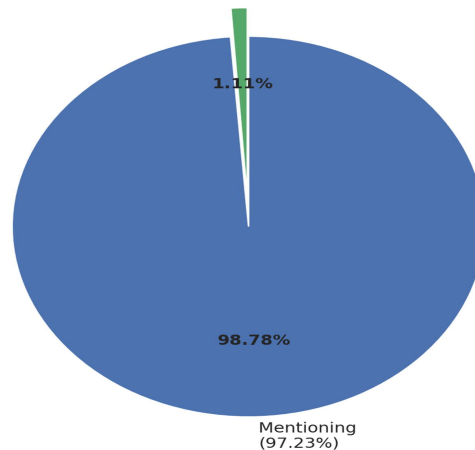


Figure 3. Overall Citation Polarity Distribution

(Total Scite citations 348102)

Computational linguistics research appeared in 1,740 journals. Table 3 lists the ten journals that published the largest numbers of articles. Lecture Notes in Computer Science ranks first with 1,242 articles, followed by IEEE Access (428 publications). Mentioning citations predominate in every journal. Lecture Notes in Computer Science also leads in absolute mentioning citations (14,907). Transactions of the Association for Computational Linguistics records the highest counts of both supporting and contrasting citations, confirming that research in the field is subject to both agreement and disagreement. Although the number of contrasting citations is modest relative to supporting and mentioning citations, they remain important indicators of divergent scholarly opinion. Two journals ACM Transactions on Asian and Low-Resource Language Information Processing and IEEE/ACM Transactions on Audio, Speech, and Language Processing received no contrasting citations, implying broad consensus with the literature they published.

4.4 Scite Citations versus Scopus Citations

One objective of the study was to examine the interrelationships between the three categories of contextual citations and Scopus citation counts. Because the Kolmogorov Smirnov test indicated that none of the distributions (supporting citations, contrasting citations, mentioning citations or Scopus citations) followed a normal distribution, Spearman's rank-correlation coefficients were calculated. The results appear 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

A strong positive correlation exists between mentioning citations and Scopus citations ($\rho = .839$, $p = .000$). Supporting citations display a weak positive correlation with Scopus citations ($\rho = .336$, $p = .000$), while contrasting citations exhibit an even weaker positive association ($\rho = .153$, $p = .000$). These findings indicate that research visibility, scholarly agreement and scholarly discourse all contribute to the overall impact of research outputs and influence the accumulation of future citations.

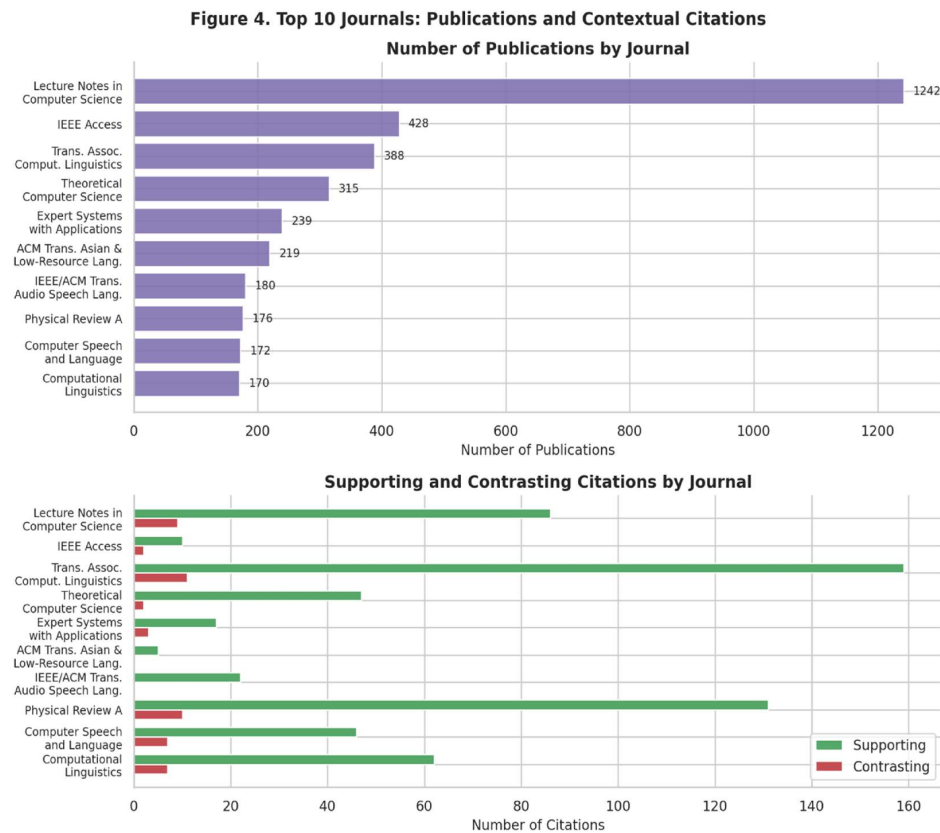


Figure 4. Top ten journals

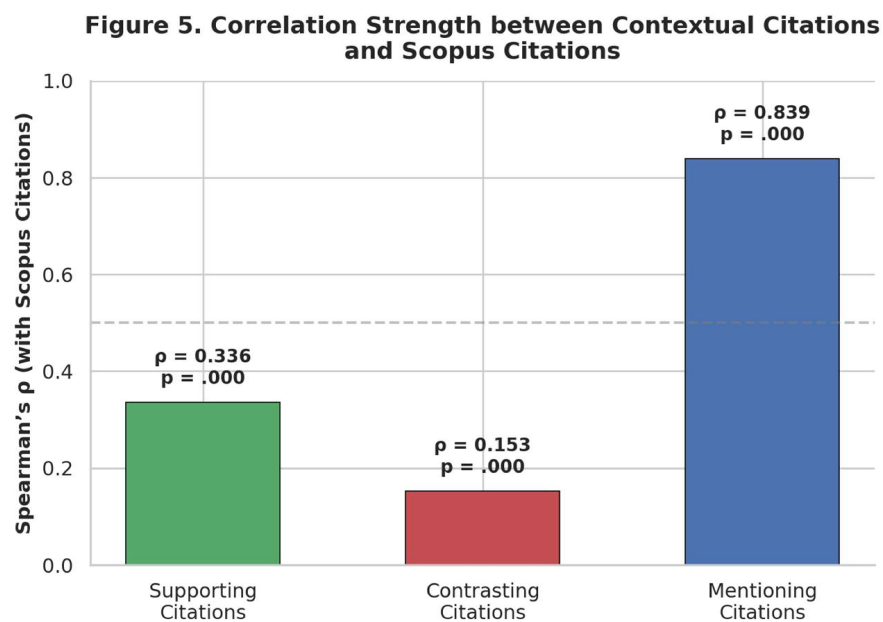


Figure 5. Spearman Rank Correlation

4.5 Sectional Citation Analysis

Scite.ai indexes citations according to the section of the citing article in which they appear. The location of a citation supplies additional information about the relevance of the cited work. A citation in the Introduction signals the strength of the theoretical background; a citation in the Methods section indicates a reusable research design; citations in the Results or Discussion sections demonstrate the use of existing findings to strengthen new analyses. Consequently, sectional analysis yields insights that simple citation counts cannot provide.

Scite classifies citations into five principal sections: Introduction, Methods, Results, Discussion and Other. Most scientific articles follow the IMRaD structure; citations that fall outside this standard framework are assigned to the “Other” category. As shown in Figure 1, 6,031 publications received citations in the “Other” section. This concentration indicates that the organisation of scientific writing in computational linguistics frequently departs from the conventional IMRaD format and reflects the interdisciplinary character of the field. The lowest number of publications (998) received citations in the Methods section, suggesting that established research designs are less frequently reused by subsequent authors. Overall, the sectional analysis highlights a highly interdisciplinary domain that places greater emphasis on conceptual frameworks than on standardised methodological templates.

Section	No. of Titles
Introduction	4896
Methods	998
Results	1042
Discussion	1874
Other	6031

Figure 1. Sectional citation analysis

Figure 6. Sectional Citation Analysis of Computational Linguistics Research

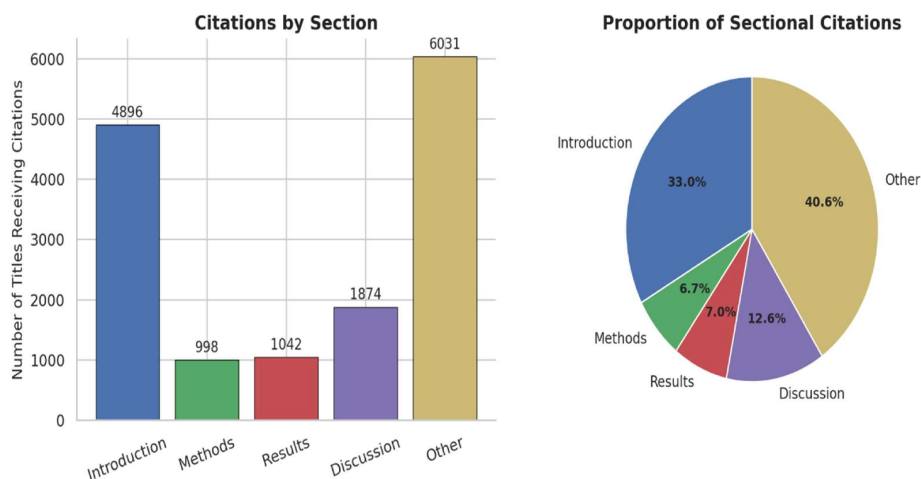


Figure 6. Section Citation Analysis

Figure 6 presents the distribution of citations according to the section of the citing paper in which they appear. Scite.ai classifies citations into five categories: Introduction, Methods, Results, Discussion, and Other.

The left panel is a bar chart showing the absolute number of titles that received citations in each section:

Section	Number of Titles
Introduction	4,896
Methods	998
Results	1,042
Discussion	1,874
Other	6,031

The right panel is a pie chart showing the relative proportions:

- Other – 40.6 %
- Introduction – 33.0 %
- Discussion – 12.6 %
- Results – 7.0 %
- Methods – 6.7 %

1. Dominance of the “Other” category (6,031 titles / 40.6 %)

The largest share of citations falls outside the conventional IMRaD structure. This indicates that a substantial portion of computational linguistics papers do not follow the traditional Introduction Methods Results Discussion format. The high concentration in “Other” reflects the interdisciplinary nature of the field and the diversity of article structures accepted in scholarly communication (e.g., system descriptions, resource papers, short communications, or hybrid formats common in computer science and NLP venues).

2. Strong presence in the Introduction (4,896 titles / 33.0%)

Citations appearing in the Introduction are the second most frequent. This suggests that computational linguistics research is frequently used to establish theoretical background, define the research problem, or position new work within the existing literature.

3. Moderate use in the Discussion (1,874 titles / 12.6 %)

Citations in the Discussion section indicate that prior work is often invoked when authors interpret their own findings or compare results with earlier studies.

4. Low representation in Methods and Results (998 and 1,042 titles)

The relatively small numbers of citations in the Methods and Results sections are noteworthy. This pattern suggests that researchers in computational linguistics less frequently reuse or directly build upon previously published experimental designs and methodological protocols. Instead, the field appears to place greater emphasis on conceptual framing and theoretical positioning than on the replication or incremental adaptation of specific methods.

Figure 6 reveals that citation practices in computational linguistics are heavily oriented toward conceptual and background engagement (Introduction + Other) rather than methodological reuse. The pronounced deviation from the classic IMRaD structure underscores the field's interdisciplinary character and the flexible article formats that characterise research communication in this domain.

5. Conclusion

Contextual citation analysis reveals the underlying relationship between citing and cited papers by examining the stance an author adopts when referencing prior work. It therefore moves beyond traditional citation-counting metrics and uncovers latent citation potential. The steady growth in supporting, contrasting and mentioning citations received by computational linguistics research documents the expanding agreements and discourses within the field. The high volume of mentioning citations points to a predominantly neutral citation practice. Although contrasting citations are fewer than supporting or mentioning citations, they remain valuable because they represent critical examination and the identification of inconsistencies, thereby contributing to the refinement of existing knowledge. Sectional analysis shows a dense accumulation of citations in the "Other" category, a pattern that arises from the frequent departure of research writing from the standard IMRaD structure and from the diversified modes of communicating research findings. Finally, statistically significant positive correlations were observed between supporting, contrasting and mentioning citations and Scopus citation counts, confirming that agreements, discourses and engagement with published research influence the accumulation of future citations.

6. Areas of Further Research

Contextual citation analysis functions as a complementary tool that enhances the precision of traditional citation based metrics. By examining the intellectual interconnections within scientific communication through the lens of citation context, it adds a qualitative dimension to quantitative impact assessment. The present study concentrated exclusively on computational linguistics; the same approach can be extended to other disciplines, individual journals, institutions and authors. Classified citations supporting, contrasting and mentioning enable scholars to locate relevant literature more efficiently. Consequently, further research is needed to evaluate the effectiveness of contextual citation analysis for impactful literature reviewing. The investigation relied on the AI-based classified citation database Scite.ai; the findings encourage the incorporation of classified citation data into other major citation databases so that the true impact of research can be more fully understood.

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* (p. 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/S1351324915-000388>.
- [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*.
- [14] Scite. (2024). [Scite.ai](https://scite.ai). [Scite.Ai](https://scite.ai). <https://scite.ai>.
- [15] 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. <https://doi.org/10.1109/ACCESS.2019.2927016>.
- [16] 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>.
- [17] 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>.
- [18] 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.
- [19] 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](https://doi.org/10.1016/0020-0271(71)90023-4).