SCRIPT JOURNAL
Journal of Linguistic and English Teaching
http://jurnal.fkip-uwgm.ac.id/index.php/Script
P-ISSN: 2477-1880; E-ISSN: 2502-6623
October 2024, Vol. 9 No. 2

# Quantitative Analysis within Language Studies: An Analytical View Based on the Bibliometrics Method

**Danang Satria Nugraha**[1]
Sanata Dharma University, Indonesia [1]
Correspondence email: d.s.nugraha@usd.ac.id

**Abstract**

***Background:***
While qualitative research methods have provided insights into linguistic structures and processes, the recent decades have witnessed a growing emphasis on quantitative approaches.

***Methodology:***
This study utilizes a bibliometric analysis based on the PRISMA guidelines to explore research trends in "Quantitative Analysis within Language Studies" from 2014 to 2024 by analyzing the metadata of 53,575 journal articles retrieved from a reputable database, i.e., Dimensions.

***Findings:***
The findings show that (1) there was a steady rise in publications throughout the period; (2) the total citations increased by over 113,000, representing a nearly 100-fold growth (1,141 citations in 2014 compared to 41,203 citations in 2024); (3) according to the ANZSRC 2020 classification, the field of Language, Communication and Culture has the highest number of publications (53,575), followed by Linguistics (32,121) and Language Studies (18,272); (4) there was the top 10 journals from 2014 to 2024; (5) there was a detail into the publication productivity and citation impact of 20 leading researchers; and (6) it has been extracted 1,028 terms concerning the examined query that classified into eight clusters.

***Conclusion:***
By integrating the insights gleaned from publication trends, citation analysis, leading authors, journal distribution, map-based visualizations, and cluster analysis of terms, this study provides a comprehensive and detailed picture of research trends within the search query.

***Originality:***
While current several research in quantitative language studies tends to delve into specific subfields, offering piecemeal understandings of the overall landscape, this study takes a distinct approach. By examining 53,575 articles content and metadata, this study aims to provide a more comprehensive perspective compare to previous fragmented insights.

| | | |
|---|---|---|
| **Keywords** | : | analytical views; bibliometrics; language studies; quantitative analysis |
| **DOI** | : | 10.24903/sj.v9i2.1695 |
| **Received** | : | May 2024 |
| **Accepted** | : | August 2024 |
| **Published** | : | October 2024 |
| **How to cite this article (APA)** | : | Nugraha, D. S. (2024). Quantitative analysis within language studies: An analytical view based on the bibliometrics method. Script Journal: Journal of Linguistic and English Teaching, 9(2),16-34. https://doi.org/10.24903/sj.v9i2.1695 |

## 1. INTRODUCTION

Language, a complex and dynamic system, has long been the subject of rigorous investigation. While traditional research methods have provided insights into linguistic structures and processes, recent decades have witnessed a growing emphasis on quantitative approaches (cf., Chen & Liu, 2014; Kortmann, 2021; Plonsky, 2014, 2017). These methods leverage the functionality of statistics and data analysis to offer a more objective and measurable understanding of language phenomena (e.g., Jiang & Liu., 2019a; Paquot & Plonsky, 2017; Plonsky & Oswald, 2014). Quantitative methods in language studies encompass a diverse array of techniques, from corpus linguistics and computational modeling to statistical analysis of survey data and psycholinguistic experiments (cf., Lei, 2012; List & Moran, 2013; Mersbergen & Patrick, 2022). These approaches enable researchers to move beyond qualitative descriptions and delve into the quantifiable aspects of language. For instance, corpus linguistics allows researchers to analyze massive datasets of text, revealing patterns in vocabulary use, grammatical structures, and stylistic variations across different genres and registers (Biber & Jones, 2009). Similarly, statistical analysis of survey data can shed light on the relationship between language attitudes, demographic factors, and language proficiency (Mizumoto & Plonsky, 2016).

Previous studies have documented a significant rise in the application of quantitative methods across various subfields of language studies. Studies by Chen & Liu (2016) and Hou & Huang (2020) highlight the growing use of quantitative analysis in areas such as speech production, language acquisition, and language assessment. These investigations underscore the potential of quantitative methods to reveal intricate patterns in language use and to provide valuable empirical data for informing language-related theories and practices. Studies by Kato et al. (2020), Lu (2017), and Zhang & Lu (2019) demonstrate the effectiveness of quantitative analysis in assessing language proficiency, evaluating the impact of instructional interventions, and exploring the relationship between language and cognition. Additionally, research by Rietveld & van Hout (2010) and Shcherbakova et al. (2023) highlights the growing use of computational methods for large-scale language analysis, opening new avenues for understanding language patterns and variation.

However, a comprehensive understanding of current research trends in quantitative language studies remains somewhat limited. Given the paucity of comprehensive overviews on the contemporary trajectory of quantitative language studies, this research endeavors to fill this scholarly gap. Existing literature has primarily focused on specific subfields or applications (e.g., Crosthwaite et al., 2023; Liao & Lei, 2017; Plonsky, 2014) providing fragmented insights into the broader landscape. This study aims to address this gap by employing a systematic and data-driven approach. Employing a bibliometric methodology grounded in PRISMA reporting standards (Page et al., 2021), this study delves into the evolving trajectory of "Quantitative Analysis within Language Studies" from 2014 to 2024. Through a structured examination of metadata derived from a substantial corpus of journal articles indexed within the Dimensions database, the present study seeks to identify recurrent terms and latent thematic structures within the field. Admittedly, this study seeks to accomplish the following research questions:

1) To what extent does the corpus of this study represent the diverse research landscape within quantitative language studies?
2) How have research interests and methodologies within quantitative language studies evolved from 2014 to 2024?
3) What are the primary thematic clusters within the field of quantitative language studies as revealed through co-occurrence analysis?

## 2. LITERATURE REVIEW

Quantitative analysis has emerged as a useful tool in language studies, offering researchers a robust framework for investigating diverse linguistic phenomena. This growing emphasis on data-driven approaches reflects a shift towards a more objective and measurable understanding of language (Egbert et al., 2020). While traditional qualitative methods like discourse analysis and in-depth interviews have provided rich insights into language use, quantitative methods offer the ability to analyze large datasets and identify patterns that might not be readily apparent through qualitative exploration alone (Biber, 2012; Biber et al., 2012; Miller & Biber, 2015).

Previous research has documented the successful application of quantitative methods in various subfields of language studies. For instance, corpus linguistics, a staple of quantitative language research, allows for the analysis of massive textual corpora, revealing patterns in vocabulary use (Yoon et al., 2019), grammatical structures (Chen et al., 2015), and stylistic variations across different genresClick here to enter text.. Similarly, quantitative analysis has proven valuable in second language acquisition research, with studies examining the relationship between instructional practices (Dewaele, 2018), learner exposure (Dewaele et al., 2022), and language proficiency gains (Jiang & Liu, 2019b). In psycholinguistics, quantitative methods have been instrumental in investigating the cognitive processes underlying language production and comprehension through experiments and surveys that measure reaction times, accuracy rates, and brain activity. The appeal of quantitative methods lies in their ability to provide statistically significant results that can be generalized to broader populations (Larson-Hall, 2012). This stands in contrast to qualitative research, which can sometimes be susceptible to researcher bias and limitations in generalizability (Gries, 2015b). Additionally, the inherent replicability of quantitative methods enables other researchers to verify findings and build upon existing knowledge.

However, a growing body of scholarship emphasizes the value of a methodological pluralism that combines quantitative and qualitative approaches (e.g., Gries, 2015a; Köhler, 2013; Rietveld & van Hout, 2011; Sheng, 2023). Qualitative research can provide context and nuance to quantitative findings (Eatough & Tomkins, 2022), helping researchers delve deeper into the "why" behind the "what" revealed by statistical analysis. For instance, a quantitative study might identify a correlation between classroom vocabulary instruction and improved student performance on standardized tests (Plonsky & Gass, 2011). However, a qualitative study observing classroom interactions could shed light on the specific instructional practices that contribute to those gains.

Bibliometric methods, which involve the quantitative analysis of large collections of publications (Donthu et al., 2021), offer a valuable tool for exploring research trends within specific domains. The present study has employed bibliometrics to analyze publication patterns in language studies, identifying emerging areas of research and prominent research communities (Mukherjee et al., 2022). This article builds upon the existing body of research by using a bibliometric approach to investigate the specific field. Accordingly, by analyzing a large corpus of journal articles retrieved from a comprehensive database, this work aims to identify prominent research themes, emerging methodologies, and potential areas for future exploration (Lim & Kumar, 2024).

## 3. METHODOLOGY

This study employed a detailed data possession process by employing a bibliometrics method based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021) to ensure the reliability and representativeness of the analyzed corpus (*see* Figure 1). The initial retrieval stage utilized the Dimensions (app.dimensions.ai) database, a comprehensive source of scholarly publications. The selection

of the database is based on (1) unparalleled breadth of coverage, (2) rigorous data linkage, and (3) advanced search functionality. By employing the search phrase "Quantitative Analysis within Language Studies" within the full text of publications, more than 2 million titles have been identified. However, to ensure the retrieved articles aligned precisely with the research focus, a four-pronged filtering approach was implemented.

The first stratum of refinement involved restricting the publication type to "article (of journal)". This decision excluded book chapters, conference proceedings, and other publication formats that might encompass a wider range of content styles and potentially less rigorous peer-review processes. focusing on journal articles ensured a focus on scholarly research contributions that have undergone rigorous peer-review, enhancing the overall quality and credibility of the analyzed data. Based on this filtration, more than 1 million titles have been screened.

The second strainer narrowed the publication year range to 2014 – 2024. This timeframe provided a focused ten-year snapshot of recent advancements in quantitative language studies. Selecting a recent timeframe ensured the capture data reflected the most current trends and methodologies in this rapidly evolving field. Additionally, a ten-year window offered a sufficient period to observe meaningful trends and capture a substantial body of research. Based on this filtration, only 959,192 titles were matched to the criterion as eligible for the next stage.

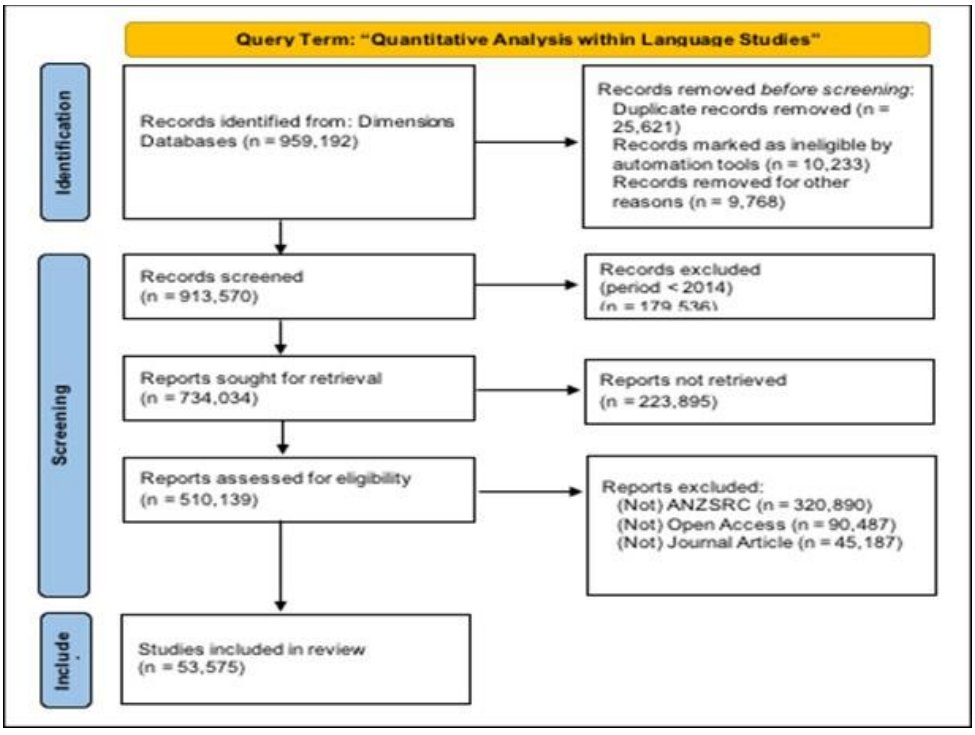



**Figure 1. Flow Diagram of the bibliometric study on "Quantitative Analysis within Language Studies".**

The third strainer specified the field of research as "Language, Communication, and Culture" according to the Australian and New Zealand Standard Research (ANZSRC 2020) framework. This categorization ensured the relevance of the retrieved articles to the target domain. The ANZSRC 2020 classification is a well-established framework for categorizing research fields and utilizing it provided a standardized approach to selecting relevant publications. By focusing on "Language, Communication, and Culture" the analysis excluded potentially tangential articles from other disciplines that might employ quantitative methods in

a different context. As the final data, a vast corpus of 53,575 titles was included for further analysis.

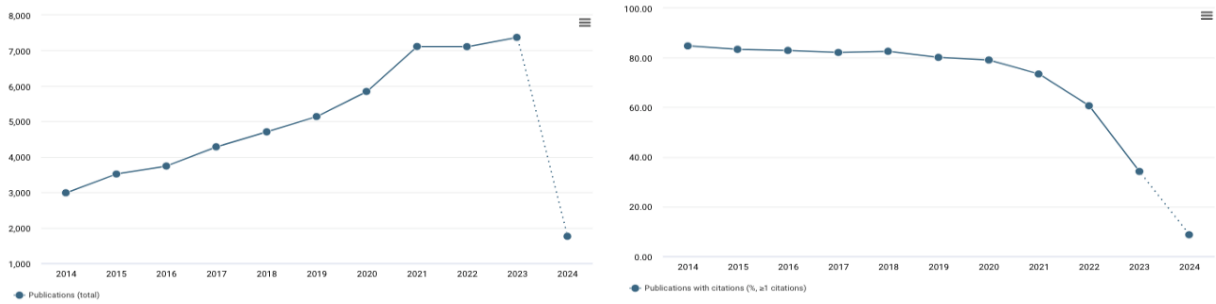
Finally, a term frequency analysis was employed to identify high-occurrence terms that represent the core themes within the corpus by utilizing the VOSviewer (1.6.20) (van Eck & Waltman, 2023). A threshold of minimum occurrences within ten publications was set, resulting in 1,713 terms for further exploration. Following rigorous double screening based on relevance, a final set of 1,028 terms emerged as the foundation for the subsequent analysis. This meticulous selection process ensured the focus remained on the most pertinent terminology characterizing quantitative research in language studies. Based on these selected terms, the map-based visualization has been created.

## 4. FINDINGS

### 4.1 Annual Publication

Figure 2(a) presents a decade-long trend (2014-2024) in the number of publications related to "quantitative analysis in language studies". The data of 53,575 publications, gleaned through a bibliometric analysis, reveals a steady rise in publications throughout the period. This upward trajectory starts with 2,987 publications in 2014 and culminates 7,371 publications in 2023. Interestingly, there appears to be a slight dip in publications between 2022 and 2023, with 7,109 publications recorded in 2022. The year 2024, with data potentially incomplete, shows a significant decrease (1,761 publications) compared to the previous year. This finding offers a springboard for further investigation into the reasons behind the increasing scholarly interest in quantitative approaches to language studies.

Upon closer examination of the data, it becomes apparent that there is no direct correlation between the rise in publication count and the number of citations (≥1 citations) received. Figure 2(b) presents a diachronic analysis (analysis through time) of publication impact in quantitative language studies. The finding suggests a declining trend in citation rates over the past decade. From a high over 80% in 2014-2018, the percentage of cited publications dips below 80% in 2019 and continues a steady decline, reaching a provisional low of 8.69% for publications in 2024 (as the data collection likely occurred recently).




(a) (b)

**Figure 2. The yearly publication covering "Quantitative Analysis within Language Studies" from 2014 to 2024 (Provider of data retrieval services: https://app.dimensions.ai/).**

This figure warrants further investigation into the potential causes behind this trend, such as a shift in publication practices or a change in the research focus within the field of quantitative language studies. It is worth considering that potential technical factors contributing to the lack of diversity in cited sources could be the over-reliance on a handful of popular publications. This means that certain publications, for instance titles under the leading authors, may be cited more frequently than others, leading to a centered representation of the available research and ideas.

## 4.2 The Amount of Citations

Figure 3 reveals a significant upward trend in the number of citations over time. From 2014 to 2024, the total citations increased by over 113,000, representing a nearly 100-fold growth (1,141 citations in 2014 compared to 41,203 citations in 2024). This exponential growth suggests a burgeoning interest in quantitative analysis within language studies. Further analysis could involve calculating the annual growth rate and investigating potential factors driving this trend. These factors might include the increasing availability of digital language data, the development of new quantitative methods, or a growing recognition of the value of quantitative approaches in language research. Note that citation counts can be influenced by value factors beyond inherent quality or impact of the research. Therefore, additional matrices or qualitative analysis might be necessary for a more comprehensive evaluation of research trends in this field, such as the FCR (Field Citation Ratio) and RCR (Relative Citation Ratio) indicators.

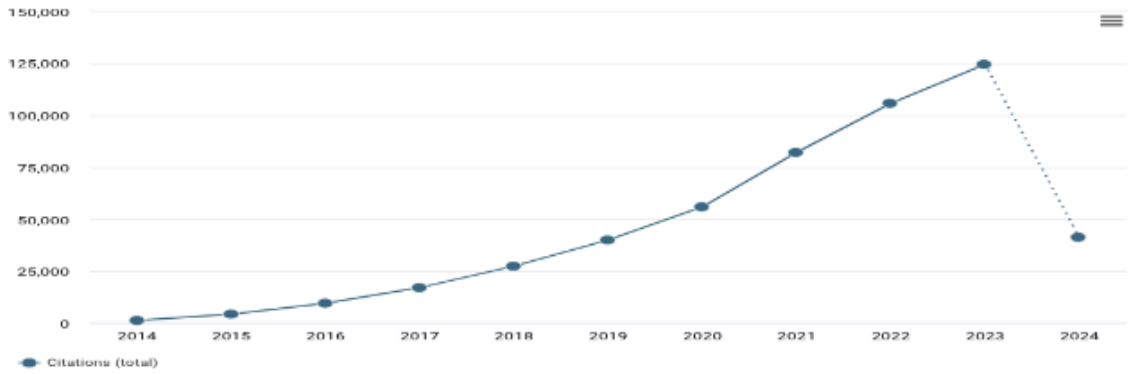
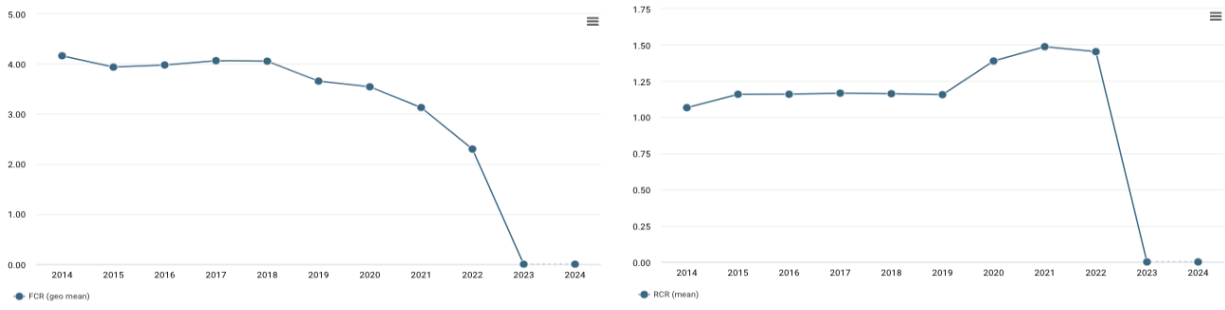



**Figure 3. The yearly scholarly impact from 2014 to 2024 as indicated by the number of citations for "Quantitative Analysis within Language Studies" (Provider of data retrieval services: https://app.dimensions.ai/).**

On the one hand, FCR is a metric used to assess the relative citation performance of a publication when compared to articles of similar age in the same field. A higher FCR signifies a more influential publication. Figure 4(a) reveals the average FCR across all publication years is 3.69. The FCR appears to be fluctuating over time. While there is general decrease from 2014 to 2022, there are also fluctuations throughout the period. Publications published between 2014 and 2018 have a consistently high FCR, ranging from 4.05 to 4.16. There is a slight decrease in FCR in 2019 (3.65) and a more substantial decrease in 2020 (3.54) and 2021 (3.13). This trend continues with a significant drop in FCR in 2022 (2.29) and no citations for publications in 2023 and 2024 (FCR of 0). Because publications usually take several years to receive citations, it is crucial to highlight that the figures for 2023 and 2024 probably reflect a citation lag.




(a) (b)

**Figure 4. The FCR (presented by (a) in left panel) and RCR (presented by (b) in right panel) examinations on publication impact from 2014 to 2024 as indicated by the number of citations for "Quantitative Analysis within Language Studies" (Provider of data retrieval services: https://app.dimensions.ai/).**

However, RCR is a metric used to assess a publication's relative citation performance when compared to other publications in the same field. Figure 4(b) reveals a slight upward trend

in the RCR over the past decade. Publications published between 2014 and 2019 have an RCR that is close to the average (1.02), with some variation from year to year. There is a more substantial increase in RCR starting in 2020 (1.39) and continuing through 2021 (1.49) and 2022 (1.45). It is important to note that the data for 2023 and 2024 likely reflects a citation lag, as it typically takes several years for publications to accrue citations.

Overall, the findings suggest that the relative citation impact of research in this field has remained stable over the past decade, with a possible increase in recent years. Further analysis would be needed to confirm this trend and to explore the reasons behind it. Potential explanations for the increase in RCR could include a growing interest in language studies or changes in publications practices leading to more highly cited articles.

### 4.3 Research Subject Areas

In respect to this object of studies, i.e. the "quantitative analysis within language studies", Table 1 provides information about the number of publications, citations, and average citations per publication in various fields of research. Here is a breakdown of the key findings. Firstly, the number of publications. The field of Language, Communication and Culture has the highest number of publications (53,575), followed by Linguistics (32,121) and Language Studies (18,272). Education (7,527) and Curriculum and Pedagogy (6,322) are among the fields with a lower number of publications.

Secondly, citations. Language, Communication and Culture also has the highest number of citations (508,702), followed by Linguistics (277,925) and Communication and Media Studies (161,084). Fields like Literary Studies (9,311) have a substantially lower number of citations. Thirdly, citations per publications (mean). Communication and Media Studies has the highest average citations per publication (15.00), followed by Human Society (10.79) and Creative Arts and Writing (10.47). Literary Studies (3.11) have the lowest average citations per publication.

**Table 1. The ten most popular study area associated with the "Quantitative Analysis within Language Studies" search query from 2014 to 2024 (Provider of data retrieval services: https://app.dimensions.ai/).**

| Field of Research | Code (ANZSRC 2020) | Publications | Citations | Citations (mean) |
|---|---|---|---|---|
| Language, Communication and Culture | 47 | 53,575 | 508,702 | 9.50 |
| Linguistics | 4704 | 32,121 | 277,925 | 8.65 |
| Language Studies | 4703 | 18,272 | 140,131 | 7.67 |
| Communication and Media Studies | 4701 | 10,735 | 161,084 | 15.00 |
| Education | 39 | 7,527 | 58,003 | 7.71 |
| Curriculum and Pedagogy | 3901 | 6,322 | 47,952 | 7.58 |
| Cultural Studies | 4702 | 4,425 | 35,071 | 7.93 |
| Creative Arts and Writing | 36 | 4,339 | 45,451 | 10.47 |
| Human Society | 44 | 4,091 | 44,143 | 10.79 |
| Literary Studies | 4705 | 2,991 | 9,311 | 3.11 |

It is important to consider that citation patterns can vary across fields. Fields that are more theoretical or foundational may tend to have lower citation counts compared to fields with more applied applications. Additionally, newer fields may have lower citation counts simply because there has been less time for them to be cited. Overall, Table 1 provides a snapshot of publication and citation activity across a range of research fields (based on the code of ANZSRC 2020). More in-depth analysis would be required to draw further conclusions about citation patterns and trends within these fields.

### 4.4 The Journal's Identification

The present study also analyzes the leading journal for the query "quantitative analysis within language studies". Table 2 provides information about the top 10 journals from 2014 to 2024 along with their citations for academic publication. It lists 10 publications, along with the number of publications (articles) each has produced, the total number of citations their publications have received, and the average number of citations per publication (citation means).

Firstly, high-impact journals. The table reveals a substantial variation in citation counts across the listed publications. Modern Language Journal has the highest average citation count (25.35), followed by PLOS ONE (23.26) and System (21.39). These publications focus on areas of high current interest in language studies. High-impact journals, exemplified by the aforementioned names, serve as intellectual barometers, establishing research agendas and influencing scholarly discourse through their rigorous publication standards and focus on cutting-edge research within their respective fields.

Secondly, low-impact journals. Several publications on the list have a substantially lower average number of citations per publication. These include Arab World English Journal (1.12), Languages (3.61), and Frontiers in Communication (5.96). It is important to consider that citation patterns can vary across fields within language studies. Publications that focus on more specialized or niche areas may naturally have lower citation counts compared to those that address broader topics.

**Table 2. The top ten journals from 2014 to 2024 that are connected to the search query "Quantitative Analysis within Language Studies" (Provider of data retrieval services: https://app.dimensions.ai/).**

| Publications Name | Publications | Citations | Citations (mean) |
|---|---|---|---|
| SAGE Open | 1,012 | 10,718 | 10.59 |
| System | 639 | 13,668 | 21.39 |
| Journal of Pragmatics | 595 | 6,198 | 11.63 |
| PLOS ONE | 539 | 12,537 | 23.26 |
| Languages | 353 | 1,276 | 3.61 |
| Arab World English Journal | 290 | 326 | 1.12 |
| Modern Language Journal | 274 | 6,946 | 25.35 |
| Frontiers in Communication | 269 | 1,602 | 5.96 |
| Lingua | 269 | 2,309 | 8.58 |
| Journal of English for Academic Purposes | 265 | 4,506 | 17.00 |

Overall, Table 2 provides a snapshot of the citation landscape for a variety of journals in language studies in respect to the query "quantitative analysis within language studies". The findings suggest that there is a significant disparity in citation counts between different publications. This highlights the importance of considering the reputation and reach of a publication when making decisions about where to publish research.

### 4.5 Leading Authors

Table 3 shows the publication productivity and citation impact of 20 researchers, likely in language studies. It includes the following information for each researcher: name, country, publications, citations and mean value of citations. Accordingly, there are several key aspects. Firstly, highly cited researchers. Table 3 reveals significant variation in citation impact among the researchers. Luke Plonsky (United States) stands out with the highest average citations per publication (80.33), followed by Jean-Marc Dewaele (United Kingdom) at 43.23 and Kazuya Saito (United Kingdom) at 40.90. This researchers' work likely addresses topics of high current interest within the field.

Secondly, geographic distribution. The researchers come from a variety of countries, with the United States, China, and the United Kingdom being the most well-represented. It is important to acknowledge that citation practices can vary by region, so raw citation counts might not fully capture a researcher's global impact. However, there is a positive correlation between the number of publications and total citations, but not necessarily average citations per publication. For instance, Haitao Liu (China) has published nearly as many papers as Plonsky but has a lower average citation count. This suggests Liu's work may be more focused on specific areas with less broad appeal. Also, it is interesting to note the variation in citation impact within some countries. For instance, China has researchers on both ends of spectrum, with Haitao Liu (17.62) and Barry lee Reynolds (6.88) showcasing this disparity. This suggests that factors beyond national affiliation, such as specific research focus and publication choices, may play a significant role in citation counts.

**Table 3. The twenty prominent authors between 2014 and 2024 who are associated with the search term "Quantitative Analysis within Language Studies" (Provider of data retrieval services: https://app.dimensions.ai/).**

| Name | Country | Publications | Citations | Citations (mean) |
|---|---|---|---|---|
| Jean-Marc Dewaele | United Kingdom | 62 | 2,680 | 43.23 |
| Haitao Liu | China | 47 | 828 | 17.62 |
| Luke Plonsky | United States | 45 | 3,615 | 80.33 |
| Lawrence Jun Zhang | New Zealand | 45 | 1,222 | 27.16 |
| Sali A. Tagliamonte | Canada | 34 | 493 | 14.50 |
| Jesse Egbert | United States | 30 | 519 | 17.30 |
| Johann-Mattis List | Germany | 30 | 840 | 28.00 |
| Kazuya Saito | United Kingdom | 29 | 1,186 | 40.90 |
| Miroslaw Pawlak | Poland | 28 | 677 | 24.18 |
| Barry Lee Reynolds | Macau (China) | 26 | 179 | 6.88 |
| Douglas E. Biber | United States | 26 | 598 | 23.00 |
| Mariusz Kruk | Poland | 26 | 548 | 21.08 |
| Xiaofei Lu | United States | 25 | 408 | 16.32 |
| Shigeto Kawahara | Japan | 24 | 277 | 11.54 |
| Roeland W.N.M. Van Hout | Netherlands | 24 | 429 | 17.88 |
| Simon J. Greenhill | New Zealand | 24 | 956 | 39.83 |
| Chu-Ren Huang | China | 22 | 160 | 7.27 |
| Lei Lei | China | 22 | 195 | 8.86 |
| Xuesong Andy Gao | Australia | 22 | 437 | 19.86 |
| Pavel Trofimovich | Canada | 21 | 291 | 13.86 |

Lastly, further considerations. There are potential explanations for citation disparity. Researchers who focus on high-demand or emerging areas within language studies may receive more citations. It is evident that publishing in high-impact journals often leads to wider readership and potentially more citations. Admittedly, collaborative research can lead to broader dissemination and potentially higher citation counts. Overall, Table 3 offers a starting point for exploring scholarly productivity and citation impact within the query "quantitative analysis within language studies".

### 4.6 Map-based Visualization

Based on the selection of threshold for a term analysis, by applying the minimum number of occurrences of a term in a ten, this study found 1,713 terms meet the threshold of 52,243 identified in all publications examined in the VOSviewer. The double screening has been applied by following the standard for the most relevant terms (60%) to the selected terms that met the threshold. From the second screening, it has been found in relation to the query "quantitative analysis within language studies". Based on these selected terms, the map-based

visualization has been created, i.e., network (*see* Figure 5), overlay (*see* Figure 6), and density (*see* Figure 7).

Figure 5 displays eight distinct clusters of interrelated terms. These clusters represent the search query "Quantitative Analysis within Language Studies" from 2014 to 2024. Each cluster has a different color marking which is the result of terms analysis based on all the data that was successfully extracted during that period. In addition, Table 4 provides a comprehensive list of the specific terms included within each of these eight clusters. A complete description and analysis of the cluster findings is presented in the following section.

Furthermore, Table 4 offers a fascinating glimpse into the research landscape of "Quantitative Analysis within Language Studies" from 2014 to 2024. By analyzing the eight term clusters and their constituent terms, one can identify prominent research themes and potential emerging areas within this field. First, cluster 1: speech production and acoustic analysis. This cluster, dominated by terms like 'acoustic analysis', articulation rate', 'fundamental frequency', and 'nasalance score', delves into the quantitative measurement of speech production. Researchers here likely employ tools like speech analysis software to explore aspects like speaking rate, voice quality, and pronunciation. The presence of "classification accuracy" might suggest investigations into automated speech recognition or speaker identification systems.

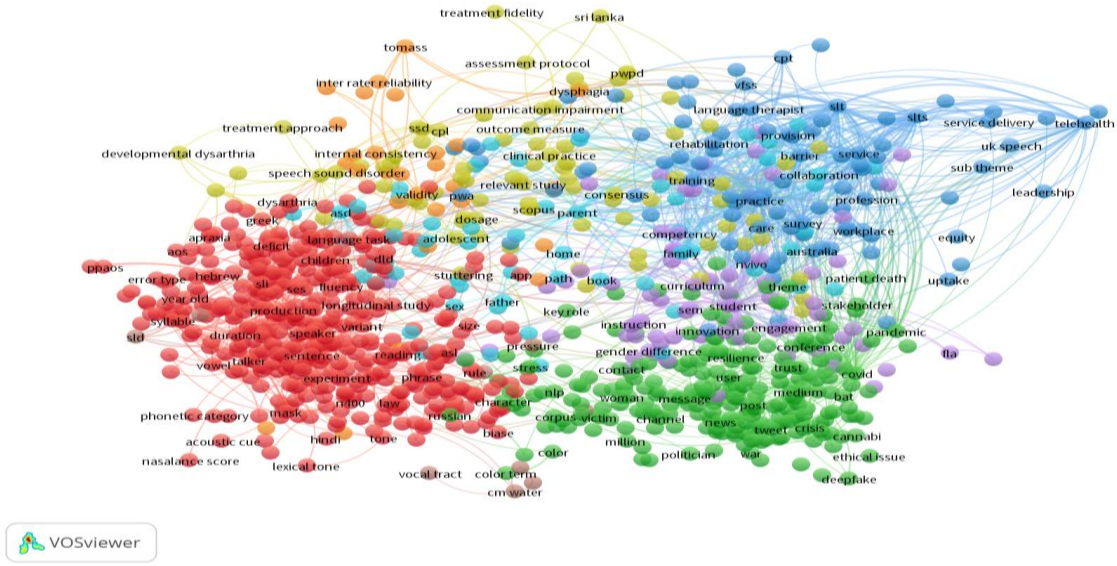



**Figure 5. The terms related to the query "Quantitative Analysis within Language Studies" are visualized through network visualization.**

Second, cluster 2: computational methods and language analysis. This cluster highlights the growing emphasis on computational approaches in language studies. Terms like "content analysis", "corpus analysis", "large scale", and "machine learning" point towards the use of big data and statistical techniques for analyzing language patterns. Research here could involve analyzing large textual datasets to identify trends in vocabulary use, sentiment analysis, or exploring language variation across different domains. Third, cluster 3: quantitative methods in applied linguistics. This cluster focuses on applying quantitative methods in areas like language therapy and education. Terms like "descriptive statistic", "quantitative data", "language therapy", and "pediatric speech" suggest research that uses quantitative data to assess language skills, evaluate therapeutic interventions, or examine language development in children.

Fourth, cluster 4: quantitative methods in clinical communication assessment. This cluster aligns with cluster 3, but with a specific focus on clinical settings. Terms like "assessment protocol", "brain injury", "cleft palate", and "outcome measure" indicate research

that develops and validates quantitative tools for assessing communication skills in individuals with various disorders. Fifth, cluster 5: quantitative research in language education. This cluster explores the use of quantitative methods in language learning and teaching. Terms like "academic writing", "blended learning", "competency", "curriculum", and "experimental group" suggest research that investigates the effectiveness of different teaching methods, assesses learner proficiency, or explores the relationship between learning strategies and outcomes.

**Table 4. The eight term clusters from 2014 to 2024 that belong to the category "Quantitative Analysis within Language Studies" in search results.**

| Cluster | Items | Cluster Color | Term (10 notable sample) |
|---|---|---|---|
| 1. | 421 | Red | accent; acoustic analysis; articulation rate; classification accuracy; computational model; dependency distance; fundamental frequency; mean length; nasalance score |
| 2. | 255 | Green | affordance; computational method; content analysis; corpus analysis; critical discourse analysis; cultural transmission; digital medium; large scale; language diversity; machine learning |
| 3. | 96 | Blue | descriptive statistic; equity; key theme; quantitative data; research literature; adherence; clinical experience; functional communication; language therapy; pediatric speech |
| 4. | 86 | Yellow | accordance; applicability; assessment protocol; brain injury; cleft palate; clinical assessment; communication skill; generalization; outcome measure; quantitative study |
| 5. | 73 | Purple | academic writing; blended learning; competency; curriculum; educator; empirical research; experimental group; proficiency level; self-efficacy; significant relationship |
| 6. | 58 | Turquoise | audio recording; conversational turn; contextual factor; environmental factor; language skill; literacy development; mix methods approach; precision; socioeconomic status; vocalization |
| 7. | 29 | Orange | aspiration; communicative function; confidence interval; content validity; factor analysis; panel; pitch; reliability; psychometric property; validity |
| 8. | 10 | Brown | vocal tract; pressure; disfluency; decrease; exercise; further investigation; stuttering |

Sixth, cluster 6: mixed methods research in language. This cluster showcases the integration of quantitative and qualitative approaches. Terms like "mix methods approach", "contextual factor", "environmental factor", "language skill", and "literacy development" suggest research that uses both quantitative data (e.g., surveys) and qualitative data (e.g., interviews) to gain a more comprehensive understanding of language learning and use in various context. Seventh, cluster 7: psychometrics and language assessment. This cluster delves into the quantitative evaluation of language tests and measures. Terms like "confidence interval", "content validity", "factor analysis", and "psychometric property" indicate research that examines the reliability, validity, and fairness of language assessments used in various settings. Lastly, cluster 8: speech disfluency and quantitative analysis. This cluster, with the smallest number terms, focuses on the quantitative analysis of stuttering. Terms like "aspiration", "disfluency", "decrease", and "stuttering" suggest research that investigates the physiological and acoustic characteristics of stuttering, potentially exploring the effectiveness of interventions aimed at reducing disfluency.

Overall, Table 4 reveals a vibrant field with diverse research foci. The prominence of computational methods and large-scale data analysis suggests a growing trend towards using technology to explore language patterns. Additionally, the integration of quantitative and qualitative approaches (cluster 6) highlights a move towards a more holistic understanding of language phenomena. In short, by analyzing trends in quantitative language studies, one can

gain valuable insights into how language is used, learned, and processed. This knowledge can inform the development of new research designs.

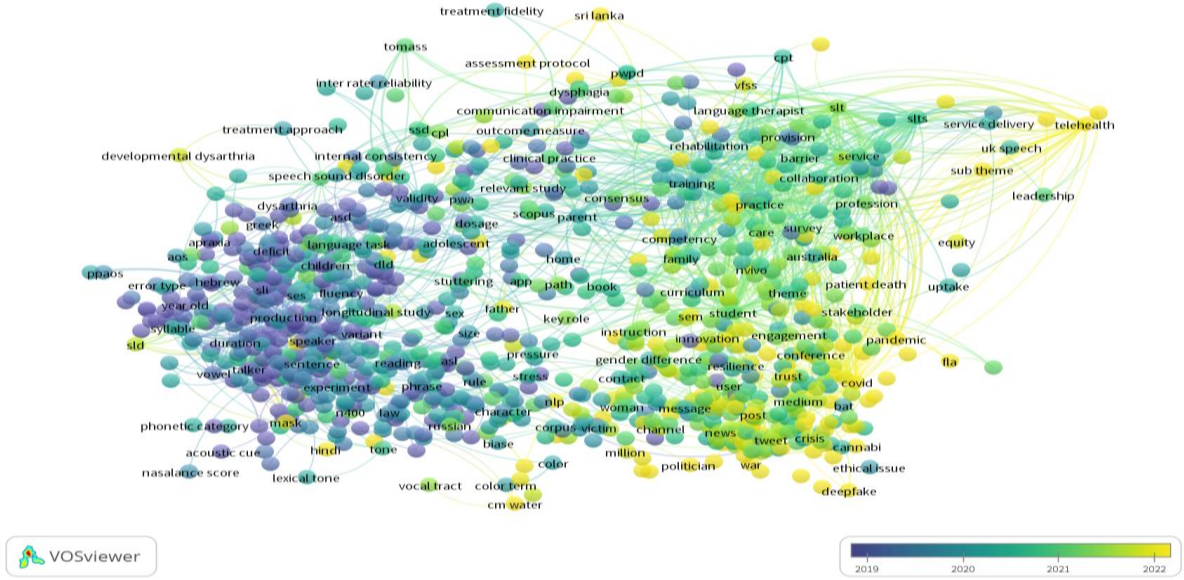



**Figure 6. Overlay visualization is used to illustrate the terms associated with the query "Quantitative Analysis within Language Studies".**

## 5. DISCUSSION

This study delves into the burgeoning field of "Quantitative Analysis within Language Studies" by employing a bibliometric approach. Focusing on publications from 2014 to 2024, the research leverages the Dimensions database to uncover key trends and thematic clusters within this domain. According to Page et al. (2021) and Donthu et al. (2021), by adhering to the PRISMA guidelines for data collection and analysis, the investigation offers an objective and systematic perspective on the evolving landscape of quantitative research in language studies.

While a detailed analysis of annual publication volume is not provided in the current discussion (*see* Section 4.1), the sheer quantity of retrieved titles (53,575) underscores the substantial scholarly interest in quantitative approaches within language studies over the past decade. This suggests a growing recognition of the value of quantitative methods in exploring intricate linguistic phenomena and fostering a data-driven understanding of language use, development, and acquisition. In other words, the growing focus on quantitative methods in language studies signifies a paradigm move within the field (Larsson et al., 2022). Researchers are increasingly recognizing the value of leveraging statistical analysis and data-driven approaches to explore the complexities of language. By employing quantitative methodologies (e.g., Liu, 2011; Liu & Xu, 2012; Norris et al., 2015; Plonsky, 2015a), researchers can move beyond misleading observations and subjective interpretations, instead uncovering statistically significant patterns and relationships within language use. This data-driven approach, to some extent, fosters a more objective and replicable understanding of these intricate linguistic phenomena, ultimately leading to a more comprehensive picture of how language functions across various contexts (Comanaru & Dewaele, 2015).

Unfortunately, the information regarding citation data is not explicitly included in the current description (*see* Section 4.2). However, further analysis of citation patterns could offer valuable insights into the impact and reach of different research areas within quantitative language studies. Articles with higher citation counts could represent particularly influential contributions that have shaped subsequent research directions. These highly cited works often

introduce novel methodologies, groundbreaking theoretical frameworks, or significant empirical findings that resonate with a wide audience of researchers. For instance, the new methodologies are including the machine learning algorithms, i.e., K-Means, Hierarchical Clustering, Naïve Bayes, amongst others. The use of this type of method directs language analysis which tends to be qualitative towards a new quantitative paradigm. By analyzing the citation patterns of these influential publications, we can gain valuable insights within a specific field. This analysis can illuminate the intellectual landscape of quantitative language studies (Plonsky, 2015b), revealing the topics and approaches that have garnered the most attention and potentially guided the course of future research endeavors.

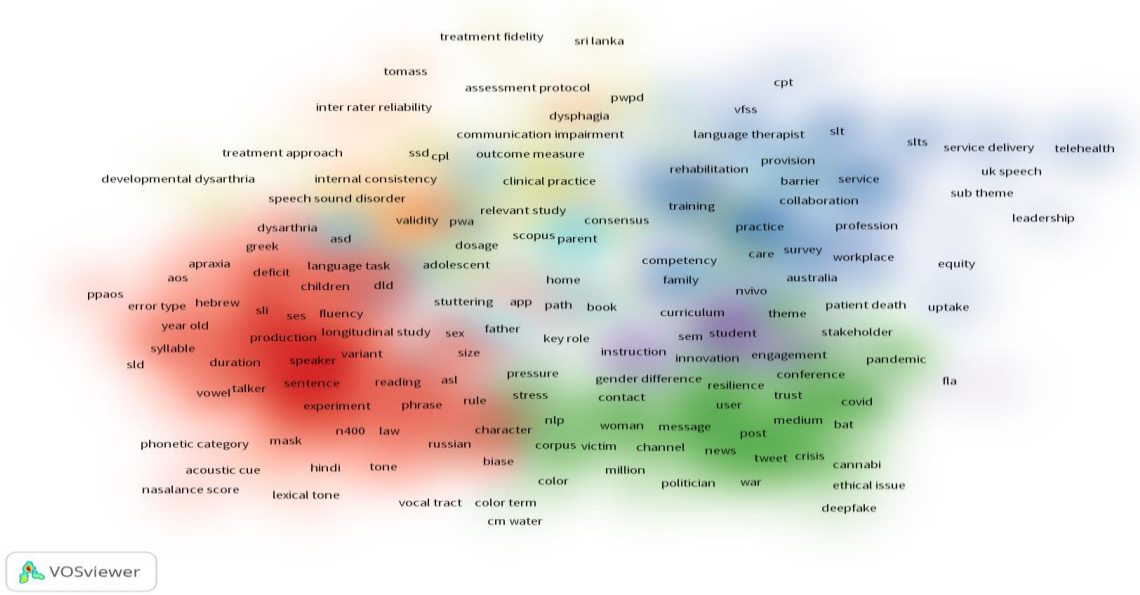



**Figure 6. The concepts associated with the subject "Quantitative Analysis within Language Studies" are shown using density visualization.**

Furthermore, the identification of leading authors and journals within this field could provide valuable insights into the key players driving the research agenda (*see* Section 4.5). By examining the publication records of authors with the highest number of publications or those contributing to highly cited articles, the discussion could highlight prominent researchers shaping the field, for example Jean-Marc Dewaele's prolific output in bilingualism or Luke Plonsky's influential research on second language acquisition. In other words, by identifying researchers with a substantial publication output or whose work garners significant citations, we can highlight prominent figures who are actively shaping the field's trajectory. Examining their research foci, methodological preferences, and collaborative networks can shed light on prevailing research themes, emerging methodologies, and potential areas of future inquiry. This analysis can inform future research directions by identifying key players and areas of intellectual influence within quantitative language studies (e.g., Egbert & LaFlair, 2018; LaFlair et al., 2015; Mystkowska-Wiertelak & Pawlak, 2016).

Similarly, pinpointing journals that publish a significant proportion of quantitative language studies research would reveal the preferred venues for disseminating scholarly findings in this domain (*see* Section 4.4). These journals can be considered the preferred venues for scholars seeking to share their findings with a targeted audience of experts. Analyzing publication patterns within these journals can reveal several key aspects of the field. On the one hand, it can shed light on the established research communities within quantitative language studies. By pinpointing the journals where scholars frequently publish (e.g., Kawahara & Kumagai, 2023; Kruk et al., 2022; Rietveld & van Hout, 2015; Xue et al., 2023), we can identify

research groups and prominent figures who are shaping the field's trajectory. This information can be particularly valuable for early-career researchers seeking to connect with established scholars and find suitable outlets for their own work. On the other hand, analyzing the thematic focus of these high-publishing journals can reveal the dominant research trends within quantitative language studies. Each journal typically cultivates a specific editorial focus, attracting submissions that align with its thematic interests. By examining the content published in these journals, we can gain a comprehensive understanding of the current areas of inquiry that are capturing the attention of researchers (e.g., Biber, 2023; Gray, 2013; Staples & Biber, 2015).

Lastly, the visual representation could depict the network of co-occurring terms within the 1,028 identified terms. By analyzing the size and proximity of nodes (terms) and the strength of connections (co-occurrences), the map (*see* Figure 5 – 7) could reveal thematic clusters and dominant research area within quantitative language studies. For instance, a cluster of terms encompassing "corpus analysis", "statistical modeling", and "vocabulary development" might signify a prominent research focus on applying quantitative techniques to analyze large text corpora and investigate vocabulary acquisition patterns. Similarly, another cluster with terms like "pronunciation assessment", "acoustic analysis", and "speech perception" could represent a distinct research area centered on quantitative methods for evaluating and understanding spoken language features. This visual representation offers a tool for navigating the intellectual landscape of quantitative language studies and identifying potential areas for future exploration. By analyzing this network map, we can identify thematic clusters where terms with strong co-occurrence relationships reside. These clusters likely represent dominant research areas within quantitative language studies. Examining the proximity of nodes and the strength of connections within these clusters can reveal the key themes and subthemes that are shaping the current research agenda (e.g., Botes et al., 2021; Dewaele, 2019; Dewaele et al., 2023; Greenhill et al., 2020; Plonsky, 2012).

## 6. CONCLUSION

By integrating the insights assembled from publication trends, citation analysis, leading authors, journal distribution, map-based visualizations, and cluster analysis of terms, a comprehensive picture of research trends within "Quantitative Analysis within Language Studies" can be constructed. This study can highlight established research areas, identify emerging themes, and pinpoint potential areas for future investigation. The prominent of terms related to speech production, computational methods, and language assessment highlights the importance of quantitative methods in understanding various aspects of language. Additionally, the emergence of mixed methods research underscores a shift towards a more holistic approach to language study. However, this study has limitations. Firstly, relying solely on text metadata may have excluded relevant publications with less explicit terminology. Secondly, the chosen timeframe (2014-2024) and single database might limit the capture of very recent trends. Finally, the analysis focused on high-frequency terms, potentially overlooking the emergence of niche research areas. It is acknowledged that this approach may have inadvertently overlooked the emergence of nascent and potentially groundbreaking research areas characterized by low-frequency, yet innovative, terminology.

Despite these limitations, the findings provide valuable insights for future research. Here are some potential directions. Firstly, in-depth analysis of specific term clusters. A closer examination of the identified term clusters could reveal more holistic research themes within the field. Secondly, author co-citation analysis. Investigating co-citation patterns could identify prominent researchers and intellectual communities within quantitative language studies. Thirdly, geographical analysis of research activity. Exploring the geographical distribution of publications could reveal research hotspots and potential areas for international collaboration.

Lastly, investigating the impact of emerging technologies. Analyzing how advancements in artificial intelligence and natural language processing are shaping quantitative language research is crucial. By exploring these directions, researchers can gain a deeper understanding of the evolving landscape of quantitative language studies and its potential to inform various areas like language pedagogy, assessment, and communication interventions. Future research that incorporates additional data sources and methodologies can further refine the understanding of this dynamic field.

**Acknowledgement**
The author acknowledges the assistance of Grammarly (https://app.grammarly.com/) in enhancing the stylistic and grammatical accuracy of this manuscript.

## 7. REFERENCES

Biber, D. (2012). Corpus-Based and Corpus-driven Analyses of Language Variation and Use. In *The Oxford Handbook of Linguistic Analysis*. https://doi.org/10.1093/oxfordhb/9780199544004.013.0008

Biber, D. (2023). Representativeness in Corpus Design. In *Practical Lexicography: A Reader*. https://doi.org/10.1093/oso/9780199292332.003.0005

Biber, D., Reppen, R., & Friginal, E. (2012). Research in Corpus Linguistics. In *The Oxford Handbook of Applied Linguistics, (2 Ed.)*. https://doi.org/10.1093/oxfordhb/9780195384253.013.0038

Biber, D., & Jones, J. K. (2009). Quantitative methods in corpus linguistics. In *Corpus Linguistics: An International Handbook* (Vol. 2, pp. 1286-1304). De Gruyter Mouton.

Botes, E., Dewaele, J., & Greiff, S. (2021). The Development and Validation of the Short Form of the Foreign Language Enjoyment Scale. *Modern Language Journal*, *105*(4), 858–876. https://doi.org/10.1111/modl.12741

Chen, H., & Liu, H. (2014). Quantitative Aspects of *Journal of Quantitative Linguistics*. *Journal of Quantitative Linguistics*, *21*(4), 299–340. https://doi.org/10.1080/09296174.2014.944327

Chen, H., Liang, J., & Liu, H. (2015). How does word length evolve in written Chinese?. *PloS one*, 10(9), e0138567. https://doi.org/10.1371/journal.pone.0138567

Chen, H., & Liu, H. (2016). How to Measure Word Length in Spoken and Written Chinese. *Journal of Quantitative Linguistics*, *23*(1), 5–29. https://doi.org/10.1080/09296174.2015.1071147

Comanaru, R.-S., & Dewaele, J.-M. (2015). A bright future for interdisciplinary multilingualism research. *International Journal of Multilingualism*, *12*(4), 404–418. https://doi.org/10.1080/14790718.2015.1071016

Crosthwaite, P., Ningrum, S., & Schweinberger, M. (2023). Research trends in corpus linguistics. *International Journal of Corpus Linguistics*, *28*(3), 344–377. https://doi.org/10.1075/ijcl.21072.cro

Dewaele, J.-M. (2019). When Elephants Fly: The Lift-Off of Emotion Research in Applied Linguistics. *Modern Language Journal*, *103*(2), 533–536. https://doi.org/10.1111/modl.12576

Dewaele, J.-M. (2018). Online Questionnaires. In *The Palgrave Handbook of Applied Linguistics Research Methodology*. https://doi.org/10.1057/978-1-137-59900-1_13

Dewaele, J.-M., Rolland, L., Cook, S., & Costa, B. (2023). Mixed Methods When Researching Sensitive Topics. In *Supporting Research in Counselling and Psychotherapy: Qualitative, Quantitative, and Mixed Methods Research*. https://doi.org/10.1007/978-3-031-13942-0_13

Dewaele, J.-M., Saito, K., & Halimi, F. (2022). How teacher behaviour shapes foreign language learners' enjoyment, anxiety and attitudes/motivation: A mixed modelling longitudinal investigation. *Language Teaching Research*. https://doi.org/10.1177/13621688221089601

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Eatough, V., & Tomkins, L. (2022). 8 Qualitative methods. In G. Schiewer, J. Altarriba & B. Ng (Ed.), *Volume 1* (pp. 163-182). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110347524-008

Egbert, J., & LaFlair, G. T. (2018). Statistics for Categorical, Nonparametric, and Distribution-Free Data. In *The Palgrave Handbook of Applied Linguistics Research Methodology*. https://doi.org/10.1057/978-1-137-59900-1_23

Egbert, J., Larsson, T., & Biber, D. (2020). Doing linguistics with a corpus: Methodological considerations for the everyday user. In *Doing Linguistics with a Corpus: Methodological Considerations for the Everyday User*. https://doi.org/10.1017/9781108888790

Gray, B. (2013). Interview with Douglas Biber. *Journal of English Linguistics*, *41*(4), 359–379. https://doi.org/10.1177/0075424213502237

Greenhill, S. J., Heggarty, P., & Gray, R. D. (2020). Bayesian Phylolinguistics. In *The Handbook of Historical Linguistics, Volume II*. https://doi.org/10.1002/9781118732168.ch11

Gries, S. T. (2015a). Quantitative designs and statistical techniques. In *The Cambridge Handbook of English Corpus Linguistics*. https://doi.org/10.1007/9781139764377.004

Gries, S. T. (2015b). Statistics for learner corpus research. In *The Cambridge Handbook of Learner Corpus Research*. https://doi.org/10.1017/CBO9781139649414.008

Hou, R., & Huang, C.-R. (2020). Robust stylometric analysis and author attribution based on tones and rimes. *Natural Language Engineering*, *26*(1), 49–71. https://doi.org/10.1017/S135132491900010X

Jiang, J., & Liu, H. (2019a). Interlanguage: a perspective of quantitative linguistic typology. *Language Sciences*, *74*, 85–97. https://doi.org/10.1016/j.langsci.2019.04.004

Jiang, J., & Liu, H. (2019b). Does Scale-Free Syntactic Network Emerge in Second Language Learning? *Frontiers in Psychology*, *10*, *925*. https://doi.org/10.3389/fpsyg.2019.00925

Kato, M., Kawahara, S., & Idemaru, K. (2020). Speaking rate normalization across different talkers in the perception of Japanese stop and vowel length contrasts. *Proceedings of the International Conference on Speech Prosody*, *2020-May*, 61–65. https://doi.org/10.21437/SpeechProsody.2020-13

Kawahara, S., & Kumagai, G. (2023). Lyman's Law can count only up to two. *Laboratory Phonology*, *14*(1). https://doi.org/10.16995/labphon.9335

Köhler, R. (2013). Statistical comparability: Methodological caveats. In *Building and Using Comparable Corpora*. https://doi.org/10.1007/978-3-642-20128-8_4

Kortmann, B. (2021). Reflecting on the quantitative turn in linguistics. *Linguistics*, *59*(5), 1207–1226. https://doi.org/10.1515/ling-2019-0046

Kruk, M., Pawlak, M., Elahi Shirvan, M., Taherian, T., & Yazdanmehr, E. (2022). A longitudinal study of foreign language enjoyment and boredom: A latent growth curve modeling. *Language Teaching Research*. https://doi.org/10.1177/13621688221082303

LaFlair, G. T., Egbert, J., & Plonsky, L. (2015). LaFlair, G. T., Egbert, J., & Plonsky, L. (2015). A practical guide to bootstrapping descriptive statistics, correlations, t tests, and ANOVAs. In *Advancing quantitative methods in second language research* (pp. 46-77). Routledge.

Larson-Hall, J. (2012). Our statistical intuitions may be misleading us: Why we need robust statistics. *Language Teaching*, *45*(4), 460–474. https://doi.org/10.1017/S0261444811000127

Larsson, T., Egbert, J., & Biber, D. (2022). On the status of statistical reporting versus linguistic description in corpus linguistics: a ten-year perspective. *Corpora*, *17*(1), 137–157. https://doi.org/10.3366/cor.2022.0238

Lei, L. (2012). Quantitative Linguistic Computing with Perl [Review of the book Quantitative Linguistic Computing with Perl, by Fengxiang Fan, Yaochen Deng]. *Literary and Linguistic Computing*, *27*(2), 227–230. https://doi.org/10.1093/llc/fqs012

Liao, S., & Lei, L. (2017). What we talk about when we talk about corpus: A bibliometric analysis of corpus-related research in linguistics (2000-2015). *Glottometrics*, *38*, 1–20.

Lim, W. M., & Kumar, S. (2024). Guidelines for interpreting the results of bibliometric analysis: A sensemaking approach. *Global Business and Organizational Excellence*, *43*(2), 17–26. https://doi.org/10.1002/joe.22229

List, J.-M., & Moran, S. (2013). An open source toolkit for quantitative historical linguistics. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, *2013-Augus*, 13–18.

Liu, H. (2011). Quantitative Properties of English Verb Valency. *Journal of Quantitative Linguistics*, *18*(3), 207–233. https://doi.org/10.1080/09296174.2011.581849

Liu, H., & Xu, C. (2012). Quantitative typological analysis of Romance languages. *Poznań Studies in Contemporary Linguistics*, 48(4), 597-625. https://doi.org/10.1515/psicl-2012-0027

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, *34*(4), 493–511. https://doi.org/10.1177/0265532217710675

Mersbergen, M. van, & Patrick, C. J. (2022). 10 Quantitative and physiological measures. In G. Schiewer, J. Altarriba & B. Ng (Ed.), *Volume 1* (pp. 201-225). Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110347524-010

Miller, D., & Biber, D. (2015). Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition. *International Journal of Corpus Linguistics*, *20*(1), 30–53. https://doi.org/10.1075/ijcl.20.1.02mil

Mizumoto, A., & Plonsky, L. (2016). R as a Lingua Franca: Advantages of using r for quantitative research in applied linguistics. *Applied Linguistics*, *37*(2), 284–291. https://doi.org/10.1093/applin/amv025

Mizumoto, A., Plonsky, L., & Egbert, J. (2020). Meta-analyzing Corpus Linguistic Research. In *A Practical Handbook of Corpus Linguistics* (pp. 663–688). Springer International Publishing. https://doi.org/10.1007/978-3-030-46216-1_27

Mukherjee, D., Lim, W. M., Kumar, S., & Donthu, N. (2022). Guidelines for advancing theory and practice through bibliometric research. *Journal of Business Research*, *148*, 101–115. https://doi.org/10.1016/j.jbusres.2022.04.042

Mystkowska-Wiertelak, A., & Pawlak, M. (2016). Designing a Tool for Measuring the Interrelationships between L2 WTC, Confidence, Beliefs, Motivation, and Context. In *Second Language Learning and Teaching*. https://doi.org/10.1007/978-3-319-30373-4_2

Norris, J. M., Plonsky, L., Ross, S. J., & Schoonen, R. (2015). Guidelines for reporting quantitative methods and results in primary research. *Language Learning*, *65*(2), 470–476. https://doi.org/10.1111/lang.12104

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting

systematic reviews. *PLOS Medicine*, *18*(3), e1003583. https://doi.org/10.1371/journal.pmed.1003583

Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research*, *3*(1), 61–94. https://doi.org/10.1075/ijlcr.3.1.03paq

Plonsky, L. (2012). Effect size. In *The Routledge Encyclopedia of Second Language Acquisition*.

Plonsky, L. (2014). Study quality in quantitative l2 research (1990-2010): A methodological synthesis and call for reform. *Modern Language Journal*, *98*(1), 450–470. https://doi.org/10.1111/j.1540-4781.2014.12058.x

Plonsky, L. (2015a). Advancing quantitative methods in second language research. In *Advancing Quantitative Methods in Second Language Research*. https://doi.org/10.4324/9781315870908

Plonsky, L. (2015b). Quantitative considerations for improving replicability in CALL and applied linguistics. *CALICO Journal*, *32*(2), 232–244. https://doi.org/10.1558/cj.v32i2.26857

Plonsky, L. (2017). Quantitative research methods. In *The Routledge Handbook of Instructed Second Language Acquisition*. https://doi.org/10.4324/9781315676968

Plonsky, L., & Gass, S. (2011). Quantitative Research Methods, Study Quality, and Outcomes: The Case of Interaction Research. *Language Learning*, *61*(2), 325–366. https://doi.org/10.1111/j.1467-9922.2011.00640.x

Plonsky, L., & Oswald, F. L. (2014). How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Rietveld, T., & van Hout, R. (2010). Statistics in Language Research: Analysis of Variance. In *Statistics in Language Research: Analysis of Variance*. https://doi.org/10.1515/9783110877809

Rietveld, T., & van Hout, R. (2011). Statistical techniques for the study of language and language behaviour. In *Statistical Techniques for the Study of Language and Language Behaviour*.

Rietveld, T., & van Hout, R. (2015). The t test and beyond: Recommendations for testing the central tendencies of two independent samples in research on speech, language and hearing pathology. *Journal of Communication Disorders*, *58*, 158–168. https://doi.org/10.1016/j.jcomdis.2015.08.002

Shcherbakova, O., Gast, V., Blasi, D. E., Skirgård, H., Gray, R. D., & Greenhill, S. J. (2023). A quantitative global test of the complexity trade-off hypothesis: The case of nominal and verbal grammatical marking. *Linguistics Vanguard*, *9*(s1), 155–167. https://doi.org/10.1515/lingvan-2021-0011

Sheng, D. (2023). Statistics in corpus linguistics: a practical guide. *Social Semiotics*, *33*(4), 909–912. https://doi.org/10.1080/10350330.2021.1969215

Staples, S., & Biber, D. (2015). Cluster analysis. In *Advancing Quantitative Methods in Second Language Research*. https://doi.org/10.4324/9781315870908-11

van Eck, N. J., & Waltman, L. (2023). *VOSviewer* (1.6.20). Universiteit Leiden.

Xue, W., Cucchiarini, C., van Hout, R., & Strik, H. (2023). Measuring the intelligibility of dysarthric speech through automatic speech recognition in a pluricentric language. *Speech Communication*, *148*, 23–30. https://doi.org/10.1016/j.specom.2023.02.004

Yoon, S.-Y., Lu, X., & Zechner, K. (2019). Features measuring vocabulary and grammar. In *Automated Speaking Assessment: Using Language Technologies to Score Spontaneous Speech*. https://doi.org/10.4324/9781315165103-8

Zhang, J., & Lu, X. (2019). Measuring and supporting second language development using computerized dynamic assessment. *Language and Sociocultural Theory*, *6*(1), 92–115. https://doi.org/10.1558/lst.31710