Vocabulary Research Trends in Applied Linguistics through Factorial Analysis and Thematic Analysis

Hamisu Hamisu Haruna
Department of English, Al-Qalam University Katsina , Katsina, Nigeria

Azza Jauhar Ahmad Tajuddin*
English Language Learning Center, Center of Fundamental and Continuing Education, Universiti Malaysia Terengganu–Malaysia
*Corresponding author

Sulaiman Muhammad Isa
Department of Arts and Sciences, Kampala University, Uganda

Abstract

The study examines vocabulary research trends in applied linguistics from 2004 to 2023, providing a fine-grained understanding of the emerging contours of vocabulary research in the field. This study utilised a mixed methods design, which combined bibliometric analysis, thematic analysis and factorial analysis to capture the evolution of vocabulary research. A dataset of 136 journal articles was composed from the Clarivate Analytics Web of Science journal database. It was found that, while the "learners" cluster showed a central influence, the clusters of "acquisition," "children," "model," and "adjustment" showed varying degree of centrality and interconnectedness, through inductive thematic analysis of the articles, by categorising them into research themes. The triangulation of these qualitatively-developed themes and the statistically-configured themes offered a more comprehensive account of vocabulary research. Thus, the resulting timeline showed the transitions in the themes from 2004 to 2023, and the trends indicated by these transitions: from "acquisition" to "knowledge", "children" to "frequency" and from "learners" to "size". The authors concluded that the study makes a substantive contribution to the field in that it presents a multidimensional view of recent trends in vocabulary research. Although this review may not have included all pertinent literature in the subject as it is restricted to the papers indexed in the Clarivate Analytics Web of Science journal database, it offers a multifaceted perspective on current trends in vocabulary research, which significantly advances the subject.

Keywords

Vocabulary Research, Bibliometric Analysis, Bibliometrix R-package, Biblioshiny Software, Factorial Analysis, Thematic Analysis

1. Introduction

Vocabulary knowledge is critical for learners of a second language (L2) in developing proficiency in the target language. With a robust vocabulary, student can understand and make sense of input, express their ideas and reach academic and professional goals (Beck, McKeown & Kucan, 2013; Schmitt, 2019). Consequently, vocabulary has been one of the main areas of investigation in the field of applied linguistics for many decades, as evidenced by the publication of hundreds of studies exploring various facets of vocabulary acquisition, teaching and assessment. Despite the extensive body of literature in this area, vocabulary acquisition continues to be one of the most challenging dimensions of L2 mastery for many learners (Milton 2013; Xodabande et al., 2022) and inadequate vocabulary knowledge has been cited as a persistent barrier to proficiency for students across contexts (Benzie, 2010). The need for on-going research on effective treatments to improve lexical outcomes is apparent in the analysis.

Bibliometric analysis is a research method that uses publication data to quantify patterns and trends over a specific period in an academic field, and across various key measures such as citations, authorship, and so on (Ellegaard & Wallin, 2015). A number of bibliometric studies have reported on the focus and trajectory of subfields within applied linguistics, including second language writing (Farsani & Jamali, 2023; Liu, 2022), classroom discourse (Rymes, 2015) and language learning strategies (Lestari & Wahyudin, 2020). These reviews offer insights into the growth and development of research

topics, productivity of influential scholars and institutions and research gaps that might be ripe for further exploration (Martínez-López et al., 2018). However, to date, no bibliometric analysis has investigated disciplinary trends specifically concerning L2 vocabulary. In this light, a comprehensive overview of vocabulary research in applied linguistics would help to identify the state of the field, influential works and active scholars as well as areas in need of further investigation.

On this note, while a bibliometric mapping of vocabulary research provides a basis for understanding the intellectual structure and conceptual foci of research, a thematic and factorial analysis of key directions within this literature can deepen our understanding. Thematic analysis groups findings into common topics or themes to understand what a body of work is conceptually about (Guest et al., 2011). Factorial analysis groups variables together statistically using data reduction methods so that relationships between topics can be identified (Yong & Pearce, 2013). By applying both thematic and factorial analysis, both the conceptual and statistical relationships between topics can be ascertained to understand critical research strands. Drawing on a mixed-methods approach that incorporates bibliometric review, thematic analysis and factorial analysis as part of the systematic review and full-text analysis, we can gain a deep understanding of scholarly activity, influential works and scholars, disciplinary themes and trends, and research gaps related to L2 vocabulary acquisition, teaching and assessment. The specific study aims are three-fold:

1. Conduct a bibliometric analysis of vocabulary research in applied linguistics journals from 2004-2023 to identify publication trends.
2. Carry out a thematic analysis using excerpts from the most highly cited articles to uncover the predominant themes and sub-themes across vocabulary research.
3. Implement a factorial analysis to identify underlying statistical relationships between applied linguistics vocabulary research topics in order to determine connections and clusters among themes.

2. Research Methodology

2.1 Data Collection

This study employs a mixed methods design incorporating bibliometric analysis, thematic analysis and factorial analysis to examine vocabulary research trends in applied linguistics over a 20-year period from 2004-2023. The publication dataset was accessed through the Clarivate Analytics Web of Science database given its extensive indexing of over 1.9 billion cited references across the social sciences, arts and humanities (Clarivate Analytics, 2024). Built-in text mining, mapping and visualization tools were leveraged to generate descriptive statistics including annual growth of publications and citations, authorship and collaboration analytics, as well as topical co-occurrence mapping. Emerging trends over time and predominance of specific topics within vocabulary literature were determined quantitatively through bibliographic mining.

2.2 Data Analysis

Data analysis for this systematic literature review was carried out using the R-based Biblioshiny software package which facilitates bibliometric analysis, scientific mapping, and visualization of research topics and trends (Aria & Cuccurullo, 2017). Specifically, we leveraged Biblioshiny tools for correspondence analysis (CA), multiple correspondence analysis (MCA) and multidimensional scaling (MDS) to examine conceptual connections and distances between themes arising from the vocabulary research literature. These interrelated techniques provide a means for statistically grouping variables to uncover relationships within a dataset (Greenacre, 2018). We first utilized CA which enables examination of associations between categorical variables through a factorial statistical approach similar to principal component analysis. However, CA provides more flexibility for analyzing count data from not-normally distributed textual corpora (Nenadić & Greenacre, 2007). In our case, CA reduced the complex lexical terminology from abstracts into simplified two-dimensional maps with proximate distance signifying strong connections between vocabulary themes.

MCA was then used to extend analysis across multiple categorical variables related to author keywords, countries, journals, etc. Finally, MDS configured the lexical data into a statistical space to determine the dimensionality and clustering tendency of research topics based on their co-occurrence and similarity (Foroudi, 2021; Yan and Ding, 2012). Alongside these factorial approaches, we also conducted an inductive thematic analysis manually through open, axial and selective coding of the vocabulary literature (Kuckartz, 2014). This allowed us to group articles into research themes and attendant subtopics that were then integrated with the statistical output maps noted above. Comparing conceptual proximities from the textual analysis with the CA, MCA, and MDS outputs enabled triangulation between the qualitatively developed themes and statistically configured factorial groupings and clusters.

2.3 Framework

Using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (O'Dea et al., 2021), an initial search of databases yielded 3,346 results. The eligibility criteria limited the publication types to open access English-language articles published between 2004 and 2023. Sources were restricted to those catalogued in the Web of Science Categories "Linguistics" and "Language Linguistics" and originating from the top ten productive countries in vocabulary research. Abstract and title screening led to the exclusion of irrelevant publications and sources lacking full text availability. This screening process resulted in a final corpus of 40 publications eligible for bibliometric analysis examining vocabulary research trends and productivity over a 20-year period. The systematic identification and

selection of materials is depicted via the PRISMA flow diagram in Figure 1. Applying inclusion and exclusion filters ensured an analytically balanced representation of contemporary vocabulary scholarship within applied linguistics upon which data mining could provide both qualitative thematic and robust statistical insights.
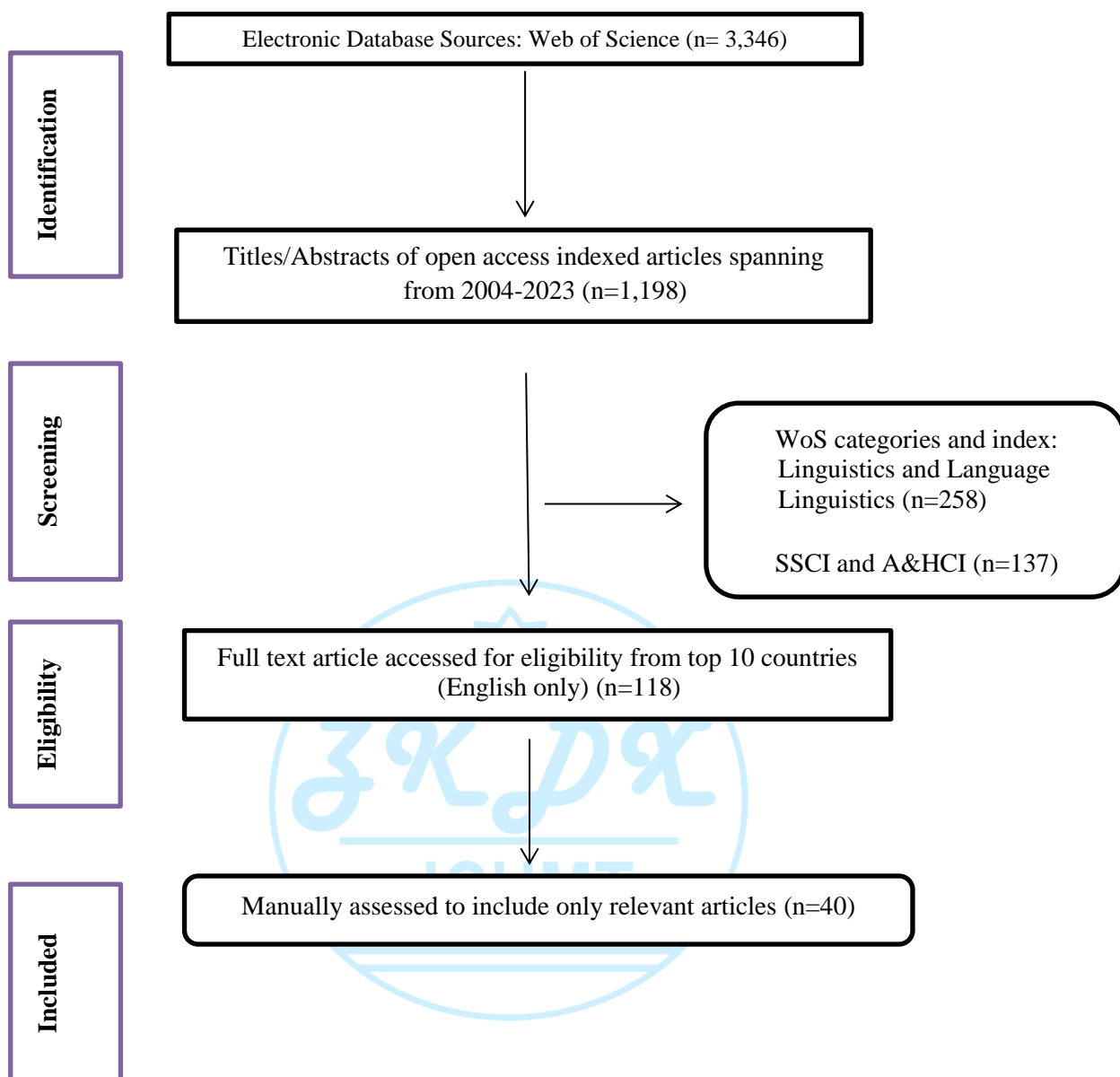

Fig.1 PRISMA Framework

3. Results and Discussion

3.1 Annual Scientific Production and Average Citations per Year

Table 1 presents an annual breakdown of scientific production and average citations per year, offering a temporal perspective on research output and impact. The data spans from 2004 to 2023 and provides insights into the trends and dynamics of scholarly activity within this timeframe. The year 2004 starts off the study with one article gathering 28 citations over its existence and stayed citable for 20 years with average citations per year (Mean TC per Year). It appears to have a sustained impact over a good period of time. The following years, 2005 and 2006, have no scientific production, so there could be gaps in research during these periods. Then from 2007 on we get resurgence in the dataset, with a noticeable increase in articles and average citations per year. It shows that 2007 was a large spike with two articles and 153 citations. They had a high concentration of impactful research output. The years 2009, 2011 and 2013 show a consistent march of research activity with one article per each and varying citation rates. On the whole, 2009 and 2013 have notably higher average citations per year, indicating the sustained influence of these works. The next few years are characterized by relatively low and fluctuating levels of scientific production. 2014 has four such articles and an impressive 44 in collective citations. It appears the field was more productive in this year, with a moderate citation influence. The reverse appears to be true for 2015. There are no such articles in 2016. However 2017 and 2018 are marked by a return to this resurgence, with two articles each. In addition these come with higher average citations per year, indicating a renewed focus on impactful research output.

Table 1 Annual Scientific Production and Average Citations per Year

Year	Articles	N	Mean TCperArt	Mean TCperYear	Citable Years
2004	1	1	28	1.4	20
2005	0	0	0	0	0
2006	0	0	0	0	0
2007	2	2	153	9	17
2008	0	0	0	0	0
2009	1	1	89	5.933333	15
2010	0	0	0	0	0
2011	1	1	56	4.307692	13
2012	1	1	45	3.75	12
2013	1	1	62	5.636364	11
2014	4	4	44	4.4	10
2015	1	1	21	2.333333	9
2016	0	0	0	0	0
2017	2	2	37.5	5.357143	7
2018	2	2	44	7.333333	6
2019	4	4	21.75	4.35	5
2020	2	2	14.5	3.625	4
2021	7	7	4.285714	1.428571	3
2022	7	6	9	4.5	2
2023	5	5	1.6	1.6	1

However, this table shows that between a period of 2021 and 2022 two articles are published, however their average citations per year decrease. The most recent year, 2023, shows a decline in both the number of articles and average citations per year, potentially reflecting a preliminary or evolving stage of research within that period.

3.2 Author Production over Time Documents

Table 2 presents information on Author Top Production Per Year and Production Over Time Documents, detailing the frequency (Freq), total citations (TC), and average citations per year (TCpY) for various authors and their respective articles. AIZAWA had a single publication in 2020, contributing 23 total citations that year, suggesting a focused output. On the other hand, DANG TNY stands out as a highly productive author, with a substantial increase in annual production over the years (from 1 in 2014 to 3 in 2022). The total citations (TC) and TC per year (TCpY) metrics reflect DANG TNY's impact, reaching 49 total citations in 2022, translating to an impressive 16.333 TCpY. BRANUM-MARTIN L, BROOKS G, BROWN KJ, and BRYBAERT M each made a single contribution in their respective years, with varying degrees of total citations and TCpY. A spectrum of research productivity and impact is evident among these authors.

Table 2 Author Top Production per Year and Production Over Time Documents

Author Top Production Per Year					Author Production Over Time Documents					
Author	Year	Freq	TC	TCpY	Author	Year	TI	SO	DOI	TC	TCpY
Aizawa I	2020	1	23	4.6	Dang TNY	2022	Measuring depth of academic vocabulary knowledge	Language teaching research	10.1177/13621688221105913	4	1.333
Branum-Martin L	2012	1	45	3.462	Dang TNY	2014	The lexical profile of academic spoken english	English for specific purposes	10.1016/j.esp.2013.08.001	92	8.364
Brooks G	2021	1	4	1	Dang TNY	2023	Online news as a resource for incidental learning of core academic words, academic formulas, and general formulas	Tesol quarterly	10.1002/tesq.3208	0	0
Brown KJ	2019	1	14	2.333	Dang TNY	2022	Incidental learning of single words and collocations through viewing an academic lecture	Studies in second language acquisition	10.1017/S0272263121000474	21	7
Brysbaert M	2018	1	71	10.143	Dang TNY	2017	The academic spoken word list	Language learning	10.1111/lang.12253	74	9.25
Capotosto L	2012	1	45	3.462	Dang TNY	2023	Open access academic lectures as sources for incidental vocabulary learning: examining the role of input mode, frequency, type of vocabulary, and elaboration	Applied linguistics	10.1093/applin/amac044	4	2
Clanton J	2021	1	4	1	Dang TNY	2022	Evaluating lists of high-frequency words: teachers' and learners' perspectives	Language teaching research	10.1177/1362168820911189	24	8
Coxhead A	2017	1	74	9.25	Webb S	2014	The lexical profile of academic spoken english	English for specific purposes	10.1016/j.esp.2013.08.001	92	8.364
Coxhead A	2022	1	24	8	Webb S	2022	Incidental learning of single words and collocations through viewing an academic lecture	Studies in second language acquisition	10.1017/S0272263121000474	21	7
Dang TNY	2014	1	92	8.364	Webb S	2017	The academic spoken word list	Language learning	10.1111/lang.12253	74	9.25
Dang TNY	2017	1	74	9.25	Webb S	2023	Open access academic lectures as sources for incidental vocabulary learning: examining the role of input mode, frequency, type of vocabulary, and elaboration	Applied linguistics	10.1093/applin/amac044	4	2
Dang TNY	2022	3	49	16.333	Webb S	2022	Evaluating lists of high-frequency words: teachers' and learners' perspectives	Language teaching research	10.1177/1362168820911189	24	8
Dang TNY	2023	2	4	2	Lawrence JF	2015	Differential effects of a systematic vocabulary intervention on adolescent language minority students with varying levels of english proficiency	International journal of bilingualism	10.1177/1367006914521698	21	2.1
Durrant P	2009	1	89	5.563	Lawrence JF	2012	Language proficiency, home-language status, and english vocabulary development: a longitudinal follow-up of the word generation program	Bilingualism-language and cognition	10.1017/S1366728911000393	45	3.462

| Durrant P	2014	1	52	4.727	Lawrence JF	2017	Vocabulary and reading performances of redesignated fluent english proficient students	Tesol quarterly	10.1002/tesq.346	1	0.125
Hwang JK	2015	1	21	2.1	Skoufaki S	2022	Spanish l1 efl learners' recognition knowledge of english academic vocabulary: the role of cognateness, word frequency and length	Applied linguistics review	10.1515/applirev-2018-0109	3	1
Hwang JK	2017	1	1	0.125	Skoufaki S	2021	Academic vocabulary in an eap course: opportunities for incidental learning from printed teaching materials developed in-house	English for specific purposes	10.1016/j.esp.2021.03.002	4	1
Lawrence JF	2012	1	45	3.462	Skoufaki S	2021	Exploring polysemy in the academic vocabulary list: a lexicographic approach	Journal of english for academic purposes	10.1016/j.jeap.2021.101038	1	0.25
Lawrence JF	2015	1	21	2.1	Warnby M	2023	Academic vocabulary knowledge among adolescents in university preparatory programmes	Journal of english for academic purposes	10.1016/j.jeap.2022.101203	1	0.5
Lawrence JF	2017	1	1	0.125	Warnby M	2023	Linking scores from two written receptive english academic vocabulary tests-the vlt-ac and the avt	Language testing	10.1177/02655322221145643	3	1.5
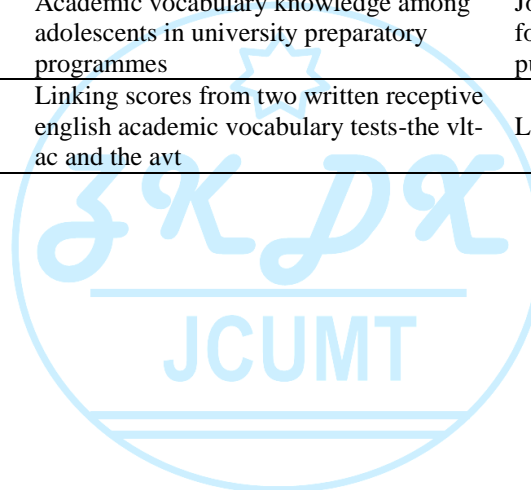
Table 2 also presents one notable trend, which is the sustained productivity of Coxhead A over the years. Her TC shows; the TCpY is consistently > 8 and > 24, respectively, it is clear that hers is a sustained and substantial impact on the field. Dang TNY's work also stands out, reflecting a diversity of publications and a TC that is in the same range as Coxhead A's. Her research spans incidental learning, high-frequency words, and vocabulary interventions (Teng, 2023). This diversity supports a comprehensive inquisition of academic vocabulary acquisition. Evidence of collaboration is also plentiful between Dang TNY and Webb S; they share several common themes and citations across multiple years. Such collaboration might at first appear possible between partners; however, they could well represent shared research interests, adding to the cumulative considerable value of their words. The survey also identifies longitudinal contributions from authors such as Lawrence JF and Durrant P, signaling a continuing interest in academic vocabulary research. These authors approached a variety of study topics, including language proficiency and home-language status, vocabulary and reading performances, students with learning disabilities.

3.3 Clusters by Authors Coupling

The Figure 2 depicts 5 clusters formed based on the coupling of documents or concepts within the documents. Each cluster contains keywords that are frequently mentioned together in the documents, indicating an association or common theme. The largest cluster (labelled 1) relates to language learning and teaching, with key terms like "language", "english", "learners", and "conf" which likely refer to conference papers where these concepts are discussed. This cluster has high frequency and centrality scores (Berahmand et al., 2018), meaning the concepts appear often and are interconnected. The high impact score also shows these are important themes. Cluster 2 connects ideas around children's language acquisition, with terms like "learners", "children", "language", and "acquisition". The group has moderately high frequency and centrality, but a lower impact score compared to cluster 1, suggesting child language acquisition specifically may be a narrower sub-theme. Cluster 3 unites concepts of language "acquisition", "context", and "repetition".
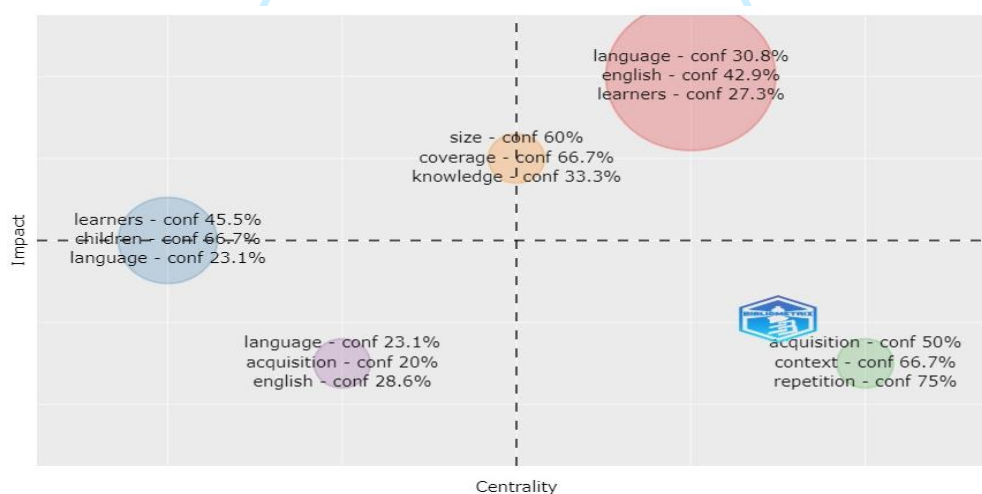

Fig. 2 Clusters by Authors Coupling

This point to research on contextual and repetitive learning mechanisms in acquisition (Williams & Rebuschat, 2016). The high centrality and percentages indicate these concepts are closely related in the documents. Cluster 4 reconnects terms from cluster 1 related to language, English, and acquisition more broadly. The centrality is moderate while frequency and impact are lower, implying more dispersed discussion of these general themes. Finally, cluster 5 links ideas around knowledge, coverage, and size, perhaps indicating study of linguistic knowledge breadth and depth. Overall, language teaching and learning is the dominant theme, with sub-topics like child development, acquisition mechanisms, English specifically, and knowledge measurement emerging as secondary concepts. The clustering and scores give a quantitative view of which ideas are most central and impactful in the document set (Rosas & Kane, 2012). Additional context on the specific research areas would aid deeper understanding, but the figure provides an overview of key associations in the language and linguistic research.

3.4 Thematic Analysis

Figure 3 provides a structured view of thematic transition and emphasis variation within the time frame of 2004 to 2023. Themes are indicated by grid position within a 2-dimensional space (Barrera et al., 2010), "From" theme and "To" theme,

with metrics from the source paper data — "Words"; "Weighted Inclusion Index"; "Inclusion Index"; "Occurrences"; "Stability Index". A thematic shift from "acquisition" (2004-2020) to "knowledge" (2021-2023) captures a shift in focus. A Weighted Inclusion Index of 0.33 indicates moderate relatedness where it would appear "From" theme is not entirely subsumed by "To" theme but where they are related. An Inclusion Index of 0.50 indicates half-inclusion, in which the "knowledge" theme is a substantial but not exclusive part of a morphing topography. Similarly, a Stability Index of 0.13 with two instances suggests a relatively stable thematic transition. This transition from acquisition to language is significant and indicates a shift in emphasis. A Weighted Inclusion Index of 0.29 indicates a moderate connection, which means that language is related to acquisition, but it is not on its periphery. An Inclusion Index of 0.33 indicates that language, within the broader theme of acquisition, is only partially included. In the same vein, a Stability Index of 0.11 with six instances suggests a relatively stable thematic trajectory. This transition from acquisition to repetition in the context represented a thematic diversification. However, a Weighted Inclusion Index of 0.20 indicates a moderate connection, which means that repetition is related to acquisition, but it is not incorporated into the theme to a large extent. The Inclusion Index is 0.25 to indicate partial inclusion of repetition within the thematic domain. At 4 occurrences, the Stability Index is 0.10 to indicate a stable thematic transition.

The thematic domain is remarkably stable; the thematic evolution from children (2004-2020) to knowledge (2021-2023) is an example of the extraordinary thematic persistence. A Weighted Inclusion Index of 0.50 for the connection signifies strong involvement. It indicates that the theme of "children" is a significant component within the theme of knowledge. Similarly, an Inclusion Index of 1.00 indicates full involvement, suggesting that "knowledge" completely composes the evolving theme. With four occurrences, the high Stability Index of 0.50 suggests a stable thematic progression. The thematic transition from "learners" (2004-2020) to "frequency" (2021-2023) further evidences an emphasis shift. A Weighted Inclusion Index of 1.00 for the relationship indicates strong involvement, showing that "frequency" is an essential part of the ongoing theme of "learners." An Inclusion Index of 0.50 for the relation shows a partial involvement, indicating that "frequency" is substantial, but not all there is of the theme of "learners." A Stability Index of 0.10 (8 occurrences) suggests a relatively stable thematic transformation. Also, the thematic evolution from "learners" to "language" signals an evolving shift in focus. Its Weighted Inclusion Index is 0.36, indicative of a moderate connection, which suggests that "language" is related to "learners," but is not entirely subsumed by it. Its Inclusion Index is 0.33, signaling a partial inclusion, which means that "language" is a substantial but not central theme of the evolving thematic landscape. A Stability Index of 0.09 (5 occurrences) points to a stable thematic trajectory. The transition from "learners" to "size," in the context of "English" and "text," points to thematic diversification.
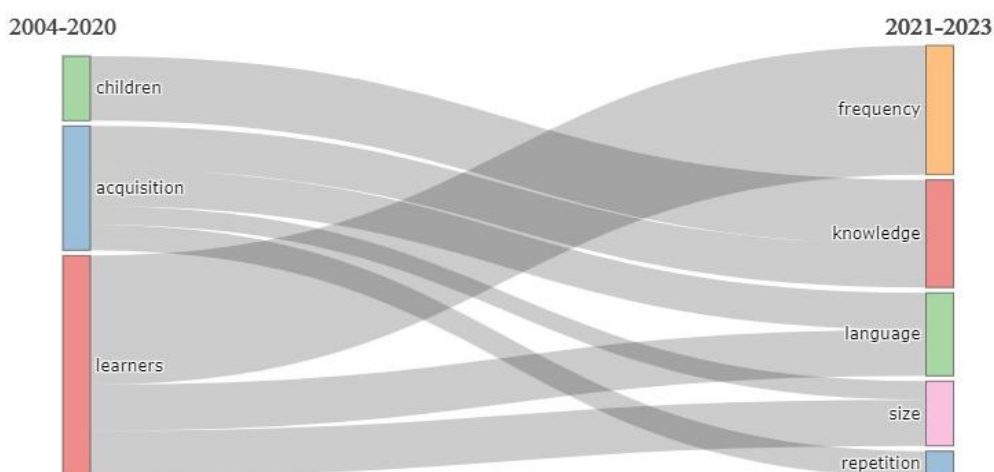

Fig. 3 Thematic Evaluation

Furthermore, Figure 4, Thematic Map, depicts the visual representation of thematic clusters along with associated metrics, providing an overall view both of the prominence and interconnectedness of different themes within the analyzed dataset. The "learners" cluster is immediately apparent, with a high Callon Centrality of 9.07, indicative of a central and influential position in the thematic network. This cluster exhibits a high Callon Density of 314.39, emphasizing a dense web of connections, indicating that "learners" is closely linked to other themes. The Rank Centrality of 6 indicates a relatively lower central ranking, while the Rank Density of 2 signifies a moderate density ranking. Therefore, the red cluster color distinguishes "learners" thematically (Fergnani, 2019). The "acquisition" cluster demonstrates notable prominence with a Callon Centrality of 6.39, indicating a central position in the thematic landscape. The high Callon

Density of 436.59 underscores dense interconnections, highlighting the significance of "acquisition" in the broader context. A Rank Centrality of 4 suggests a relatively central ranking, while the Rank Density of 3 indicates a moderate density ranking. The orange-red cluster color represents the "acquisition" theme. "Children" emerge as a distinct cluster with a Callon Centrality of 2.98, signifying a substantial but less central influence. The exceptionally high Callon Density of 857.5 indicates a tightly interconnected thematic space around "children." A Rank Centrality of 2 suggests a relatively higher central ranking, while the Rank Density of 6 signifies a high-density ranking. The green cluster color signifies the "children" theme. The "model" cluster boasts a Callon Centrality of 7.64, indicating a central thematic presence. The Callon Density of 576.19 suggests a dense network around the "model" theme. A Rank Centrality of 5 indicates a moderate central ranking, while the Rank Density of 4 suggests a moderate density ranking. The purple cluster color signifies the "model" theme. "Adjustment" stands as a cluster with a Callon Centrality of 4, representing a moderate level of centrality. The Callon Density of 604.17 indicates a dense thematic network associated with "adjustment." A Rank Centrality of 3 suggests a moderate central ranking, while the Rank Density of 5 signifies a moderate density ranking. This shows that brown cluster color is attributed to the "adjustment" theme, as also evident in Ding et al. (2023).
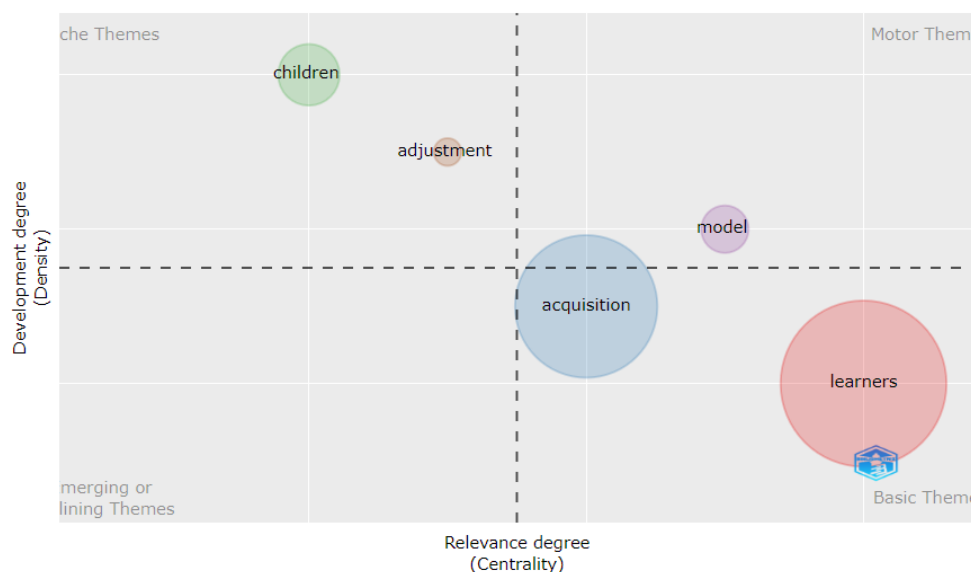

Fig. 4 Thematic Map for Time Slide One

Similarly, the thematic clusters and their associated metrics are depicted shedding light on the prominent themes of the period and their occurrences (Casadei et al., 2023). The "learners" cluster utterly dominates Period 1, with substantial occurrences on a number of associated words (i.e., "language", "english", "vocabulary", etc.). As evident from eight occurrences, the prominence of this cluster reveals its thematic importance. This may suggest a focus of the "learners" on the process of language acquisition. The array of terms in this cluster indicates processes of Period 1 language acquisition from several angles. Moreover, the "children" cluster occurs four times, reflecting a preoccupation with learning language in a context filled with children. This may reflect the exploration of particular ways of forming or acquiring language, or it may point to relevant constraints and affordances that would be encountered by new speakers." The "model" cluster, occurring three times, points to a thematic emphasis on the models or theories of language within Period 1. It is followed by the term "adjustment" (occurs twice) which may mark adjustments to these language models or strategies. The "adjustment" cluster, with two occurrences, may signify a thematic exploration of adjustments in language learning or instructional approaches. The limited occurrences suggest a less pervasive but still notable thematic presence during Period 1 (Droogan & Peattie, 2017).

Insights are provided for the Figure 5 Thematic Map for Time Slide Two. Table rows in the associated table correspond to words, their frequency of occurrence, and the thematic cluster they belong to, along with designated cluster label. The "knowledge" cluster shows four occurrences. This suggests a thematic emphasis on knowledge during Time Slide Two. Meaningful discussions or research might be related to the acquisition or application of knowledge within a specific context. The inclusion of "children" in the "knowledge" cluster would suggest a focus on knowledge acquisition among younger learners. The "size" cluster, with five occurrences, suggests a thematic focus on the size or magnitude of linguistic elements. The associated words like "english," "coverage," "text," and "words" indicate a comprehensive examination of size-related aspects within the linguistic domain.
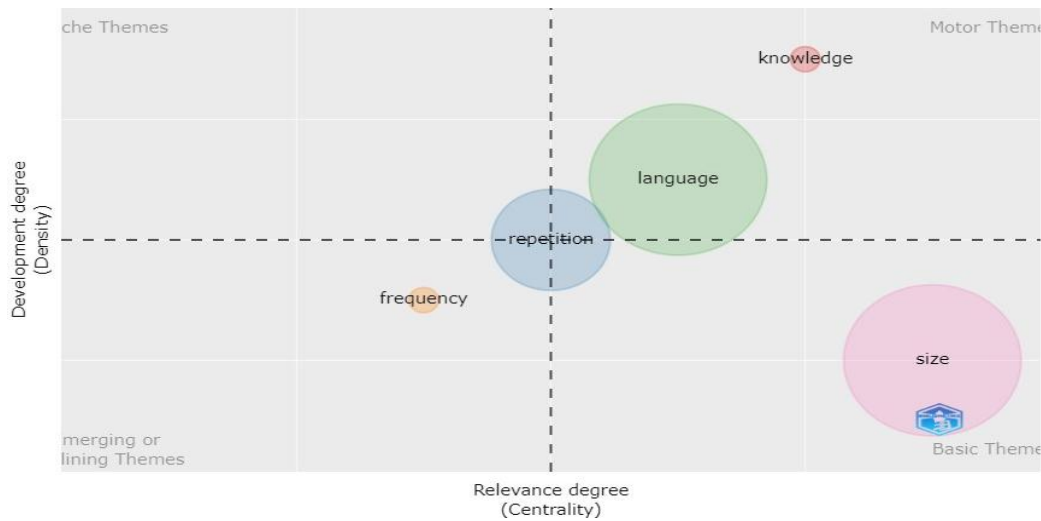
Fig. 5 Thematic Map for Time Slide Two

Figure 6, Thematic Map for Period Two, provides a quantitative evaluation of thematic clusters during a specific timeframe. The Callon Centrality, Callon Density, Rank Centrality, and Rank Density metrics offer insights into the prominence and interconnections of different thematic clusters. The "size" cluster emerges as the most influential cluster with a Callon Centrality of 12.25, underlining its central role in a thematic network at this period. This suggests that discussions or research related to the size of linguistic elements are not just prominent, but key components of the overall thematic landscape. The high Callon Density of 321.67 adds another dimension in the shape of the dense interconnections within the "size" cluster. This underscores a comprehensive coverage of size-related aspects from throughout the overall linguistic domain. The "knowledge" cluster (nine) demonstrates a noteworthy Callon Centrality (9.26), indicating considerable influence and prominence during the period of investigation. A Callon Density of 605.77 demonstrates a highly dense network of connections, suggesting that knowledge-related themes were intricately probed during this period. The Rank Centrality and Rank Density values of six and seven, respectively, indicate moderate central ranking but lower density ranking within the thematic network. During Period Two, the "language" cluster (10) demonstrated substantial influence and prominence in the form of a Callon Centrality of 8.55. The Callon Density of 477.84 indicates a dense thematic network associated with language, suggesting that linguistic aspects were extensively explored. A Rank Centrality of 5 and Rank Density of 5 indicates a moderate central and data set ranking within the thematic landscape. The "repetition" and frequency clusters. Each with a Callon Centrality of 5.40 and 2.44 exhibit their thematic ubiquity. Callon Densities of 477.08 and 447.47 suggest dense networks. However, as Rank Centrality and Rank Density metrics indicate their moderate central and density dataset rankings within the thematic network in Period Two.
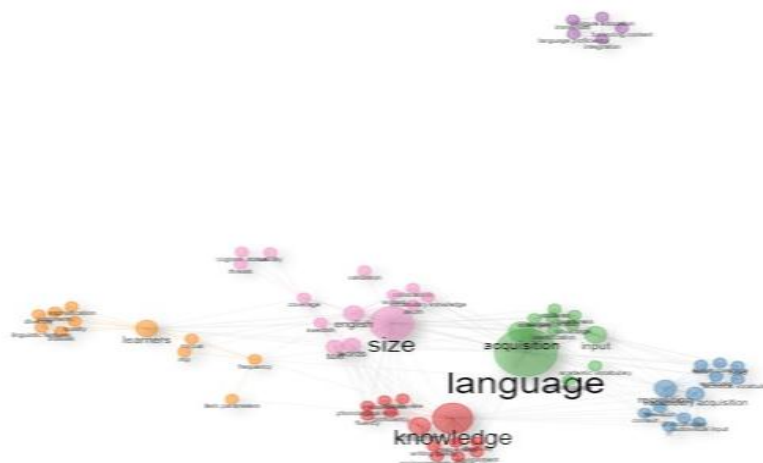

Fig. 6 Thematic Map Period Two

By conducting a multidimensional scaling analysis of the topics represented by the 32 papers associated with cluster 4 of red, a single cluster emerged that is displayed as the red cluster in Figure 6. The keywords that comprise this cluster, "testing," "teachers," "English proficiency," "English medium instruction" (EMI), "EFL," and notably, "academic vocabulary," are positioned away from the center of the coordinate plot, underscoring the fact that the nature of the relationship between these keywords is one of not closeness.

In other words, "academic vocabulary," we can say from the diagram, shares a cluster with other keywords such as "testing" and "teachers," but it is extremely not the same thing. Examining the spatial representation of common nouns in abstracts provides an initial sense of the thematic divergence within the red cluster that underscores "academic vocabulary" as a unique research topic within vocabulary studies. This representation might offer language education researchers and educators some useful insights (Tseng, 2018). For example, the existence of one prominent cluster and the isolated positioning of "academic vocabulary" within it can serve as a reminder of the intricacies of the various thematic flows within vocabulary research, which may be a useful guide for future research directions. In relation to the current study, it could be seen as highlighting, for example, that future research may benefit from more closely attending to the very specific and detailed discussion of "academic vocabulary" and what vocabulary researchers can learn from it within the context of either language teaching or testing.

3.5 Factorial Analysis

We also performed a bibliometric analysis on the field of research in the "Vocabulary Research". In order to carry it out, we drew the co-occurrence cluster of the keywords in the 1,779 papers obtained from the 6,856 retrieved from the query the day the retrieval took place. Configuration cluster of the co-occurrence anchors (the articles) of the cluster of keywords was obtained from one factorial technique, the multiple correspondence analysis, which works with categorical variables trying to reduce the number of factors necessary to reveal the relationships among a set of variables. Three methods of factor analysis of bibliometric were suitable for carrying out that future of factorial implementation: the analysis of correspondence, the analysis of multiple correspondence and the factor analysis of the proximities on the variables. These approaches were utilized to ascertain the proximity between keywords and the overarching subject. The anticipation was that this method would unveil a close association. The conceptual structure maps, depicted in Figures 4 to 6, offer a more intricate depiction of the proximity and divergence within the research domain.

The output of the correspondence analysis in Figure 7 reveals the presence of two distinct clusters. The red cluster encompasses keywords such as "vocabulary depth," "updated vocabulary levels test," "high-frequency words," "lexical coverage," "teacher cognition," "corpus linguistics," "knowledge," "L2 learners' vocabulary knowledge," "secondary school," "textbooks," "English language culture teaching," and "spin." In contrast, the blue cluster includes the keyword "controlled productive knowledge," "word frequency counts," "English as a foreign language (EFL)," "collocation size," and "teachability collocations." The relationship among the keywords within the "Vocabulary Research" topic is identified as not particularly strong. However, it is evident that certain keywords like "vocabulary depth," "updated vocabulary levels test," "high-frequency words," "lexical coverage," "teacher cognition," "corpus linguistics," "knowledge," and "L2 learners' vocabulary knowledge" exhibit a close association with each other. Conversely, the keyword cluster consisting of "secondary school," "textbooks," "English language culture teaching," and "spin" appears notably distant from the other keywords within the same cluster, indicating a weaker relationship with the rest. This correspondence analysis sheds light on the thematic organization within the field of vocabulary research.

On the other hand, the observed distance of certain keywords suggests thematic divergence or less direct relationships within the same cluster. The clustering of keywords provides valuable insights for researchers, educators, and practitioners interested in vocabulary research. It allows for the identification of cohesive sub-themes and areas of concentration within the broader context of vocabulary studies. The strength and weakness of relationships between keywords guide the understanding of the field's structure, aiding in the development of targeted research agendas and effective teaching strategies in language education.
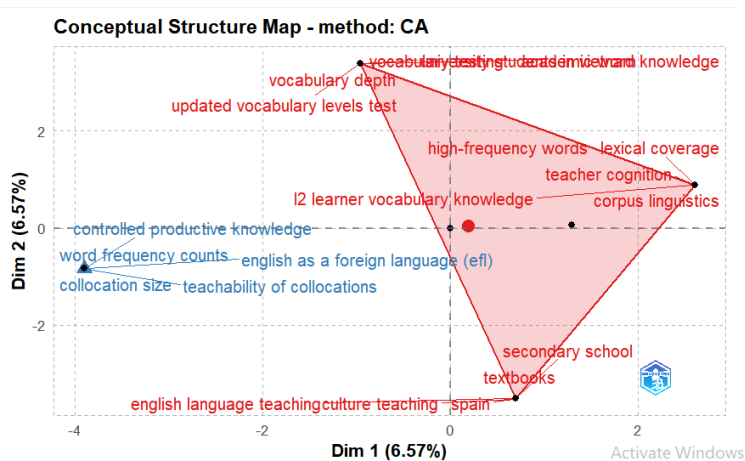

Fig. 7 Correspondence Analysis (CA)

The outcomes of the multiple correspondence analysis are visually represented in Figure 7. The spatial arrangement of points and their distribution along dimensions offers valuable insights into the findings. The analysis reveals the formation of a single cluster, prominently depicted as a large red cluster. This cluster encompasses keywords like "testing," "teachers," "English proficiency," "English medium instruction," and "EFL." Significantly, the cluster appears equidistant from the center of the coordinates, suggesting a balanced distribution in relation to the analyzed dimensions. When we visualize the keyword hierarchy, we can see that it is the largest cluster in the entire dataset. Multiple correspondences analysis (MCA) is a useful technique for studying relationships and patterns among categorical variable, which in this case, are keywords. By looking at the distribution of points in the cluster in space, one can see how the keywords are connected or what they are associated with. If the points in the cluster from a complete geometric pattern about the centroid in any direction, then one might safely assume that the keywords are all equally distant from the centroid, suggesting homogeneity: a clear indication that the keywords are underpinned by a thematic or a contextual vocabulary/lexis research. The size and centrality of the read cluster have serious implications for both researchers and practitioners involved in language education. We see the words testing, teachers, and English proficiency. A clear cluster like this might be taken as evidence of strong thematic coherence, with research and scholarly attention concentrated at the intersect of language testing, teaching practices and English language proficiency — a likely candidate for three-term descriptor of critical areas of concern within the broader landscape of language education.

4. Research Implication

The results have far-reaching implications for academic institutions, educators, and policymakers within the field of applied linguistics. Firstly, the insight into identified vocabulary research trends offers a window onto the changing terrain of language education. It allows institutions to align their curricula and their research output to the direction of the field, ensuring the continuing relevance and currency of their academic program. More to the point, in offering us a detailed understanding of the thematic clusters in the field (Tan et al., 2023), this research gives educators the tools to similarly adjust their teaching methodologies such that they too align with the most salient research area, birthing a culture of innovation in language instruction. And finally, policymakers may also benefit from the findings presented here. Through the allocation of resources to the key areas identified through a bibliometric and thematic analyses of the type conducted here, they may in turn be better positioned to form and inform language education policies. Furthermore, recognizing patterns of international collaboration can prompt institutions to forge partnerships with prolific contributors to enhance the global reach of vocabulary research (Afjal, 2023).

In practice, educators and language professionals can employ the identified clusters and themes as they sharpen their own instructional methods. For instance, more knowledge about the salience of certain topics or ideas - say, "learners" - can inform the construction of lesson plans that dovetail with the broad contours of contemporary research. Educators can adapt their approaches in light of the heightened importance of such themes as "size" or "knowledge." At a more granular level, the thematic analysis can help to reveal interdisciplinary angles that make it possible for practitioners to weave insights from related fields into the fabric of language instruction. For language professionals developing materials or assessments, understanding the frequency of specific themes and their interconnectedness can provide the basis for creating relevant and impactful resources (Shernoff et al., 2017). The social implications of this research also extend to language learners, researchers, and the broader community. Language learners have the potential to gain from instructional approaches that are designed based on the thematic clusters identified in this study through potentially improved language acquisition. Researchers can also make use of these findings to identify gaps or areas for further exploration in the language, thus fostering collaboration and the exchange of knowledge amongst the academic community. The broader community stands to gain from the advances in language education that will stem from improved instructional methods as they contribute to improved communication, cultural understanding and social integration.

Theoretical implications emanate on two levels from these factorial analyses. In terms of methodology, this study has employed rarely used factorial analysis techniques in applied linguistics research, such as correspondence analysis, multiple correspondence analysis, and multidimensional scaling. The methods introduced here allow scholars to uncover complex relationships among thematic elements that methods typically used in vocabulary research would not reveal (Gibson, 2017). However, it is not just the ability to carefully scrutinize the data that these techniques offer scholars. When clusters have been identified, they can be used in much the same way that items are in factor analysis to identify the centrality of the elements of the cluster, and thus begin the process of utilizing them in formulating theoretical constructs. Teachers and researchers working in language acquisition need such theoretical frameworks in their explorations of new

teaching strategies and materials that seek to individuate vocabulary instruction to match students' level of cognitive development. Finally, these thematic maps and transitions depict the evolving theoretical landscape facing these scholars, with the hope that just as they can work to improve the cluster structure of an individual map, they can also, as they did in so many cases in the progression from map to map, work to improve the theoretical structure of the evolving landscape.

5. Conclusion

Through a symbiosis of bibliometric and thematic analyses, research trends on vocabulary in applied linguistics unraveled important insights on how the landscape of language education has been reshaped over the past two decades. By marrying both quantitative and qualitative analyses, a mixed-methods approach provided a comprehensive depiction of scholarly discourse. A set of 118 articles published in the top 10 most productive countries was used to run a bibliometric analysis, from which a careful examination of publication trends, growth, citation patterns, international collaboration in the field, and other relevant phenomena were scrutinized. By doing so, the present study was able to quantitatively discern how vocabulary research in applied linguistics has evolved thematically in the last 20 years. Subsequent to factorial analyses with a wide array of statistical techniques (e.g., correspondence analysis, multiple correspondence analysis, and multidimensional scaling (MDS) analysis), intricate relationships among the thematic elements were uncovered. The analyses of these two methodologies collectively illuminated the thematic clusters, affording a detailed conceptualization of the interrelations within the corpus of L1vocabulary research. The thematic clusters' visualisation and concomitant commentary allowed for the parsing of cohesive subthemes and loci of drift, which informed lines of inquiry and instructional design.

Alongside the factorial outputs, the thematic maps and shifts laid bare a thoroughgoing accounting of the course of the thematic transition. In summary, the preponderance throughout the research corpus of "learners", "acquisition", "children", "model", "adjustment" and their correlates are manifest. These themes offer focal points for educators, researchers, and policymakers, highlighting opportunities for targeted interventions and concentrated research efforts. Further, the detection of the clusters "size" and "knowledge" serves to underscore the multifaceted nature of vocabulary research, pointing to directions for future research efforts. The evolving nature of the themes across time, particularly the movement from "acquisition" to "knowledge" and "learners" to "frequency," suggests shifts in emphasis for research on vocabulary across time. The stability indices for these transitions allow for some conclusions about the continuity and coherence of these thematic paths, yielding information about whether they represent consistent research emphases on which scholars, educators, and policymakers might choose to spend resources.

Funding Information
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Conflict
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References
1. Afjal, M. (2023). Bridging the financial divide: a bibliometric analysis on the role of digital financial services within FinTech in enhancing financial inclusion and economic development. *Humanities and Social Sciences Communications*, *10*(1). https://doi.org/10.1057/s41599-023-02086-y
2. Aria, M., & Cuccurullo, C. (2017). bibliometrix : An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007
3. Barrera, D., Kayacik, H. G., Van Oorschot, P. C., & Somayaji, A. (2010). A methodology for empirical analysis of permission-based security models and its application to android. In *Proceedings of the 17th ACM conference on Computer and communications security* (pp. 73-84). https://doi.org/10.1145/1866307.1866317
4. Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.
5. Benzie, H. J. (2010). Graduating as a 'native speaker': international students and English language proficiency in higher education. *Higher Education Research & Development*, *29*(4), 447–459. https://doi.org/10.1080/07294361003598824

6. Berahmand, K., Bouyer, A., & Samadi, N. (2018). A new centrality measure based on the negative and positive effects of clustering coefficient for identifying influential spreaders in complex networks. *Chaos Solitons & Fractals*, *110*, 41–54. https://doi.org/10.1016/j.chaos.2018.03.014

7. Braun, V., & Clarke, V. (2020). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, *18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

8. Brunfaut, T., & McCray, G. (2015). Looking into test-takers' cognitive processes whilst completing reading tasks: A mixed-method eye-tracking and stimulated recall study. ARAGs Research Reports Online, AR/2015/001.

9. Casadei, P., Bloom, M., Camerani, R., Masucci, M., Siepel, J., & Ospina, J. V. (2023). Mapping the state of the art of creative cluster research: a bibliometric and thematic analysis. *European Planning Studies*, *31*(12), 2531–2551. https://doi.org/10.1080/09654313.2022.2158722

10. Clarivate Analytics. (2024). Web of Science Fact Sheet. https://clarivate.com/webofsciencegroup/solutions/web-of-science-would-you-like-to-know-more/

11. Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*, 66–76. https://doi.org/10.1016/j.esp.2013.08.001

12. Ding, M., Zhang, J., Shen, G., Zheng, Q., & Yuan, H. (2023). From photographic images to hierarchical networks—Color associations of a traditional Chinese garden. *Color Research & Application/Color Research and Application*, *48*(6), 735–747. https://doi.org/10.1002/col.22886

13. Droogan, J., & Peattie, S. (2017). Mapping the thematic landscape of Dabiq magazine. *Australian Journal of International Affairs*, *71*(6), 591–620. https://doi.org/10.1080/10357718.2017.1303443

14. Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*, *105*(3), 1809–1831. https://doi.org/10.1007/s11192-015-1645-z

15. Fàbregues, S., & Fetters, M. D. (2019). Fundamentals of case study research in family medicine and community health. Family Medicine and Community Health, 7(2). 10.1136/fmch-2018-000074

16. Farsani, M. A., & Jamali, H. R. (2023). Collaboration network of applied linguistics research articles with different methodological orientations. *Studies in Second Language Learning and Teaching*, *13*(4), 727-754.

17. Fergnani, A. (2019). Mapping futures studies scholarship from 1968 to present: A bibliometric review of thematic clusters, research trends, and research gaps. *Futures*, *105*, 104–123. https://doi.org/10.1016/j.futures.2018.09.007

18. Foroudi, P., Akarsu, T. N., Marvi, R., & Balakrishnan, J. (2021). Intellectual evolution of social innovation: A bibliometric analysis and avenues for future research trends. *Industrial Marketing Management*, *93*, 446–465. https://doi.org/10.1016/j.indmarman.2020.03.026

19. Gibson, C. B. (2016). Elaboration, generalization, triangulation, and interpretation. *Organizational Research Methods*, *20*(2), 193–223. https://doi.org/10.1177/1094428116639133

20. Greenacre, M. (2018). *Compositional data analysis in practice*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429455537

21. Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis.* Sage Publications.

22. Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice and using software*. Sage Publications.

23. Lestari, M., & Wahyudin, A. Y. (2020). Language learning strategies of undergraduate efl students. *Journal of English Language Teaching and Learning*, *1*(1), 25–30. https://doi.org/10.33365/jeltl.v1i1.242

24. Liu, Z. (2022). Mapping the research trends of third language acquisition: A bibliometric analysis based on Scopus. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1021517

25. Martínez-López, F. J., Merigó, J. M., Valenzuela-Fernández, L., & Nicolás, C. (2018). Fifty years of the European Journal of Marketing: a bibliometric analysis. *European Journal of Marketing*, *52*(1/2), 439-468.

26. McLean, S., & Kramer, B. (2015). The creation of a new vocabulary levels test. *Shiken*, *19*(2), 1-11.

27. Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. *C. Bardel, C. Lindqvist, & B. Laufer (Eds.) L*, *2*, 57-78.

28. Nagy, W., & Townsend, D. (2012b). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, *47*(1), 91–108. https://doi.org/10.1002/rrq.011

29. Nakata, T., & Webb, S. (2015). Does studying vocabulary in smaller sets increase learning? *Studies in Second Language Acquisition*, *38*(3), 523–552. https://doi.org/10.1017/s0272263115000236

30. Nenadic, O. ., & Greenacre, M. (2007). Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package. *Journal of Statistical Software*, *20*(3), 1–13. https://doi.org/10.18637/jss.v020.i03

31. O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W., Parker, T. H., Gurevitch, J., Page, M. J., Stewart, G., Moher, D., & Nakagawa, S. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biological Reviews/Biological Reviews of the Cambridge Philosophical Society*, *96*(5), 1695–1722. https://doi.org/10.1111/brv.12721

32. Pellicer-Sánchez, A. (2015). Learning L2 collocations incidentally from reading. *Language Teaching Research*, *21*(3), 381–402. https://doi.org/10.1177/1362168815618428

33. Peters, E. (2013). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research*, *18*(1), 75–94. https://doi.org/10.1177/1362168813505384

34. Rosas, S. R., & Kane, M. (2012). Quality and rigor of the concept mapping methodology: A pooled study analysis. *Evaluation and Program Planning*, *35*(2), 236–245. https://doi.org/10.1016/j.evalprogplan.2011.10.003

35. Rymes, B. (2015). *Classroom discourse analysis: A tool for critical reflection*. Routledge.

36. Shernoff, D. J., Sinha, S., Bressler, D. M., & Ginsburg, L. (2017). Assessing teacher education and professional development needs for the implementation of integrated approaches to STEM education. *International Journal of STEM Education*, *4*(1). https://doi.org/10.1186/s40594-017-0068-1

37. Tan S., Harji M. B., & Hu X. (2023). A Bibliometric Analysis of English for Specific Purposes from 2011 to 2023 Using Citespace: Visualizing Status, Themes, Evolution, and Emerging Trends. *Journal of Language and Education, 9*(3), 159-175. https://doi.org/10.17323/jle.2023.17632

38. TENG, M. F., HUANG, Y., & MIZUMOTO, A. (2023). Incidental Vocabulary Learning Through Word-Focused Exercises: The Association with Vocabulary Learning Strategies. *Asian Journal of English Language Teaching*, *32*.

39. Tseng, J. J. (2018). Exploring TPACK-SLA interface: Insights from the computer-enhanced classroom. *Computer Assisted Language Learning*, *31*(4), 390-412.

40. Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*(1), 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

41. Watts-Taffe, S., Fisher, P., & Blachowicz, C. (2017). Vocabulary instruction: Research and practice. *Handbook of research on teaching the English language arts*, 130-161.

42. Webb, S., & Chang, A. C.-S. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? Studies in Second Language Acquisition, 37(4), 651-675. https://doi.org/10.1017/S0272263114000606

43. Williams, J., & Rebuschat, P. (2016). *Implicit learning and second language acquisition*. Evanston, IL, USA: Routledge.

44. Xodabande, I., Torabzadeh, S., Ghafouri, M., & Emadi, A. (2022). Academic vocabulary in applied linguistics research articles: A corpus-based study. *Journal of Language and Education, 8*(2), 154-164. https://doi.org/10.17323/jle.2022.13420

45. Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. *Journal of the American Society for Information Science and Technology*, *63*(7), 1313–1326. https://doi.org/10.1002/asi.22680

46. Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in Quantitative Methods for Psychology, 9(2), 79-94.