

THE EVOLUTION OF ARABIC VOCABULARY AND LEXICAL STUDIES: BIBLIOMETRIC TRENDS FROM 1970 TO 2024

Nur Aishah Hanis Johan

Fakulti Pengajian Islam, Universiti Kebangsaan Malaysia,
43600 UKM Bangi, Selangor, Malaysia

Lily Hanefarezan Asbullah*

Fakulti Pengajian Islam, Universiti Kebangsaan Malaysia,
43600 UKM Bangi, Selangor, Malaysia

**Corresponding Author's Email: lilyhane@ukm.edu.my*

Article History:

Received : 6 February 2026

Accepted : 5 March 2026

Published : 9 April 2026

© Penerbit Universiti Islam Melaka

To cite this article:

Johan, N. A. H. & Asbullah, L. H. (2026). The Evolution Of Arabic Vocabulary And Lexical Studies: Bibliometric Trends From 1970 To 2024. *Jurnal 'Ulwan*, 11(1), 103-120.

ABSTRACT

This bibliometric analysis investigates the research landscape, growth trajectory, and thematic evolution of scholarly work on Arabic vocabulary from 1970 to 2024, utilizing data retrieved from the Scopus database. A total of 1,592 peer-reviewed articles were systematically selected after excluding review papers and non-research content, ensuring the inclusion of original studies focused specifically on Arabic vocabulary acquisition, usage, pedagogy, and linguistic development. VOSviewer software was used to generate bibliometric maps that visualize patterns of co-authorship, institutional affiliations, geographical distribution, and keyword co-occurrence. The results reveal that Arabic vocabulary research has undergone a significant transformation, marked by a sharp increase in publications particularly after the year 2000, with exponential growth continuing into 2024. Saudi Arabia, Jordan, and the United States were identified as the top contributing countries, while King Saud University and the University of Haifa stood out as the most productive institutions. Author keyword analysis revealed key research trends such as adaptation, aphasia, and applied linguistics, alongside emerging digital and computational themes like deep learning, Bidirectional Encoder Representations from Transformers (BERT), and social media which indicate a growing intersection between Arabic vocabulary studies, neurolinguistics, and artificial intelligence. Most of the research outputs are situated within the social sciences and humanities reflecting the foundational linguistic and cultural roots of Arabic vocabulary

studies, though increasing contributions from STEM and health sciences suggest a widening interdisciplinary relevance. This study not only maps the academic growth of Arabic vocabulary as a research field but also underscores its expanding global significance, its integration with modern technologies and its relevance to diverse educational and sociocultural contexts. The findings serve as a valuable resource for scholars, educators, and policymakers aiming to further explore and develop this crucial domain of Arabic language research.

Keywords: *Arabic vocabulary; bibliometric analysis; Scopus database; VOSviewer; research trends*

1.0 INTRODUCTION

The mastery of Arabic vocabulary is a critical element across various sectors, fundamentally impacting education, communication, and cultural expression. The significance of effective vocabulary teaching strategies in Arabic stems from its status as one of the most widely spoken languages globally, with an estimated 310 million native speakers and an additional 50 million non-native speakers, leading to its importance in multicultural societies where Arabic serves as a bridge between diverse communities. Empirical studies underscore that a limited vocabulary can significantly impede learners' reading comprehension and overall linguistic development, highlighting the urgent need for pedagogical approaches that directly address these gaps (Permatasari & Mauludiyah, 2022). Consequently, enhancing Arabic vocabulary acquisition emerges as not only a linguistic concern but also a pressing educational and societal necessity.

This urgency for targeted research into effective Arabic vocabulary teaching strategies is fueled by longstanding educational challenges and a plethora of instructional approaches currently available. Systematic analysis is critical to navigate this complexity. For instance, storytelling has been identified as an innovative teaching method that engages learners and enhances vocabulary acquisition (Arruum et al., 2022; , Akla, 2022). Evidence suggests that storytelling aids in vocabulary retention by associating words with narrative contexts, fostering active listening and comprehension skills (Vaahtoranta et al., 2019). Furthermore, contemporary educational practices have explored integrating digital tools and game-based learning to bolster vocabulary instruction (Mutholib et al., 2020). Comparative studies on various teaching aids reveal a pronounced need for integrative strategies that effectively combine traditional methods with modern technology (Mohamad et al., 2021). Therefore, this study seeks to explore instructional practices that merge these approaches into multimodal learning environments tailored to diverse learner needs.

To support this investigation, data were collected from authoritative academic databases, primarily Scopus. Utilizing Scopus provides a robust platform for tracking citations and analyzing research trends, while the Web of Science (WoS) enriches the review with a broader spectrum of interdisciplinary studies. A systematic review methodology was employed, applying stringent filters for document type (peer-reviewed journal articles) and publication year (2018–2023) to

guarantee the inclusion of contemporary and relevant sources (Permatasari & Mauludiyah, 2022; , Hà & Bellot, 2020).

In conclusion, enhancing Arabic vocabulary acquisition is integral not only for individual linguistic proficiency but also for broader educational and societal advancements. The systematic exploration of teaching strategies that blend storytelling, digital tools, and traditional methods can effectively address the diverse needs of Arabic language learners, thereby improving their vocabulary mastery and overall communication skills.

2.0 METHODS

2.1 Data Source And Search Strategy

The data collection for this bibliometric analysis on Arabic vocabulary was carried out in March 2025 using the Scopus database. The central focus of this study was centered around research concerning Arabic vocabulary, particularly about lexical development and usage. To initiate the process, text mining was performed by applying the search terms “Arabic” in conjunction with either vocab or lexical in the title and abstract. The query string was carefully refined to exclude non-research outputs by filtering journal articles in the final publication stage, published from 1970 to 2024 excluding records from 2025. This refinement initially retrieved 2,790 documents which were then narrowed down to 1,605 records by excluding other document types. To further ensure accuracy, potential review articles were filtered out based on keywords commonly used in reviews such as review, progress, update, and overview found in the title and abstract, as well as by examining the full text. After excluding 16 such review articles, a total of 1,592 valid records remained for analysis. The finalized search string explicitly excluded those identified review articles using their EID numbers.

This comprehensive search strategy reflects a notable narrowing of focus from general vocabulary studies to more targeted and original research on Arabic vocabulary. The significant reduction from 2,790 to 1,592 records underscores the dominance of non-article and review content in the field potentially indicating a scholarly environment that values theoretical and overview-based contributions. This trend might be interpreted as a sign that Arabic vocabulary, as a research niche, is evolving from broad conceptual discussions toward more empirical, data-driven studies. Such a filtering process also suggests an increasing academic rigor and specialization in Arabic lexical studies as only a small portion of publications met the strict criteria for inclusion. Fig. 1 illustrates the systematic steps undertaken in refining the central theme and sub-themes. The complete set of search strings, including potential reviews and final articles, is presented in the supplementary material.

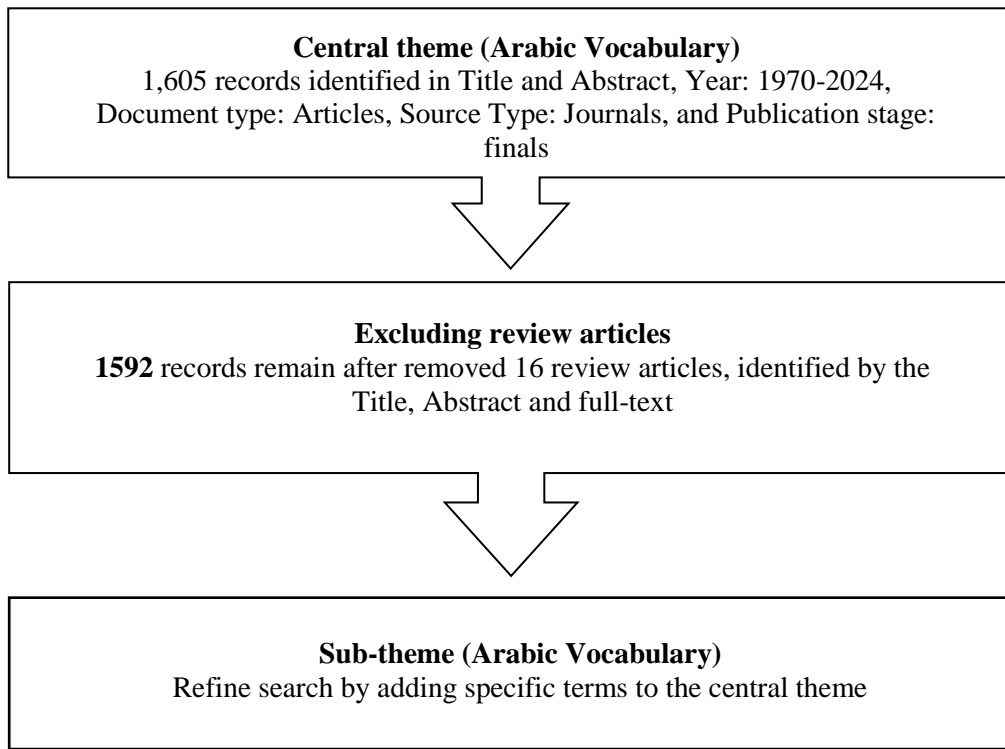


Figure 1: Research Framework On The Data Collection Process For Primary And Sub-Themes

2.2 Bibliometric Maps

The bibliometric maps were created from 1605 records obtained from the Scopus database using the VOSviewer software (version 1.6.17). The software is selected as the key tool of the study because of its free to use policy and can support a wide range of databases. The design of the maps used items as the representation of the topic of interest, which can be either countries or author keywords. The strength of the relationship between one item and another is measured by the software and displayed as a numerical number that can be easily interpreted. For example, when the connection between two items is stronger, the reading of the link will exhibit a higher numerical value. In terms of the co-authorship evaluation, the link strength connecting two items represents the shared amount of publication such as journal articles and conference proceedings between two countries. The publication was selected according to the affiliation of co-authors. So when an article has authors of different countries and affiliations, this will be identified as a joined publication for both countries. In addition, the total link strength acts as an indicator of the total co-authorship strength of a given country with the others. Likewise, the link strength between two author keywords for the co-occurrence bibliometric maps represents the number of articles where the keywords have occurred together. The construction of data representation for both tables and figures were made using a full counting

system. Further explanation of the available functions of the software can be obtained using the instruction manual (2021).

a) Evaluation of Co-authorship

The study of co-authorship covered an extensive range of datasets, including all 81 countries affiliated to 2808 authors. The countries were clustered based on their respective continents: Asia, Europe, America, Africa, and Oceania and the maps were presented in network visualization mode using a full counting system.

b) Evaluation of Co-occurrence

Assessment of the co-occurrence based on author keywords includes 4739 keywords from the extracted records. Before creating co-occurrence maps, the keywords selected by the VOSviewer were filtered to reduce the redundancy and ensure that each keyword would have a distinct meaning. For example, keywords such as acid medium, acid solution, acid media and acid mediums were counted as one and re-labelled as the acid solution. The process results in 4682 keywords that were used for the maps. The minimum occurrences of author keywords were set at default which is 4 with 177 keywords being generated. The constructed maps were presented in the overlay visualization mode using a full counting method. In this mode, the item's colour and link strength indicate the average publication year of documents containing the selected keywords.

3.0 RESULTS AND DISCUSSION

3.1 Publication Productivity And Trends Of Research Interest

Across a span of 55 years from 1970 to 2024, a total of 1590 publications were recorded with a primary focus on Arabic vocabulary (Fig. 2). Prior to the year 1990, the publication trend was relatively slow and sporadic, with fewer than 10 articles appearing per decade. A slight increase began in the early 1990s, where modest annual outputs of 3 –7 articles were recorded. However, this momentum lacked sustainability until a noticeable shift occurred after the year 2000. Starting from 2001, there was a consistent climb in annual publications, reaching 12 articles in 2003 and doubling to 24 by 2008. The turning point appeared between 2010 and 2015, when the publication productivity surged significantly, with yearly output climbing from 16 in 2010 to a remarkable 72 in 2015. This trajectory strongly suggests a growing scholarly recognition of Arabic vocabulary as an essential domain, particularly in alignment with digital language processing, second language acquisition, and Islamic educational needs.

From 2015 onwards, a rapid and sustained escalation in publication output was evident. The annual growth rate intensified further, reaching 104 publications in 2019 and then almost doubling within a span of four years, recording 179 articles in 2024 alone. This tremendous growth of over 1100% compared to the year 2000 signals a robust academic momentum and possibly reflects global trends in Arabic language learning, the rising demand for Arabic in non-native contexts, and the

integration of vocabulary research with computational linguistics. Notably, the period from 2020 to 2024 saw consistent triple-digit publication figures each year, indicating an era of maturation for the field. The post-2020 boom may also be attributed to increased access to digital publication platforms, broader interdisciplinary collaboration, and national policies encouraging research in native language preservation and enhancement. With such exponential growth, Arabic vocabulary research stands as a flourishing field poised for even deeper global impact in years to come.

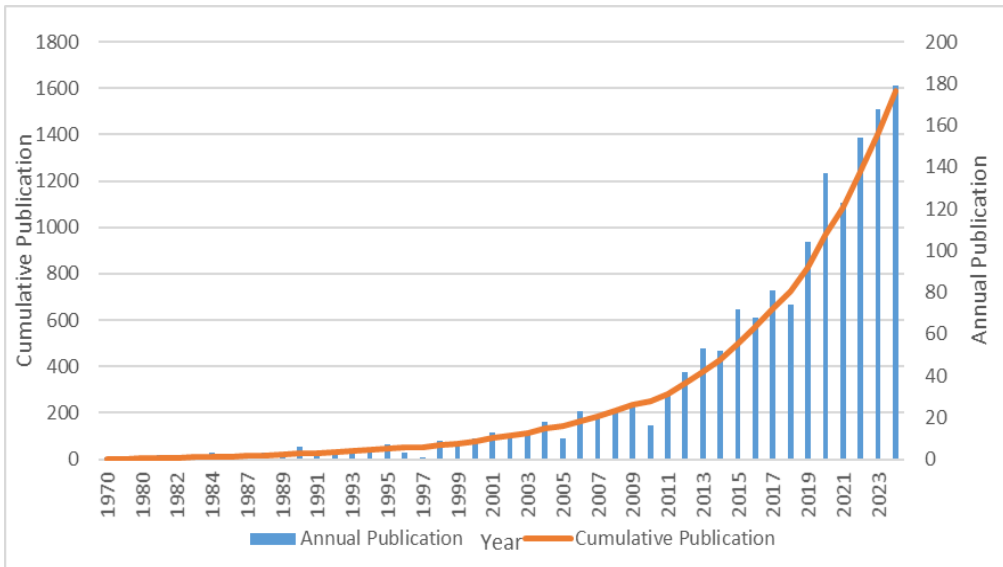


Figure 2: The number of research articles and conference papers on SEL indexed in Scopus from 1978 to 2024.

3.2 Preferred Journals

Based on the compiled data of the top 10 most productive journals in Arabic vocabulary research, it was observed that two publishers, Springer Nature and Academy Publication, each had two journals featured in the list (Table 1). The journals from Springer Nature were *Journal of King Saud University - Computer and Information Sciences* and *International Journal of Speech Technology*, while Academy Publication produced *Theory and Practice in Language Studies* and *Journal of Language Teaching and Research*. The remaining journals are scattered across a variety of publishers, including Oxford University Press, MDPI, and Babel-Verlag, reflecting the global academic interest in the field. The most productive journal was *Dirasat: Human and Social Sciences* from the University of Jordan, which published 37 articles, contributing 2.32% to the overall corpus. Interestingly, although it holds the top rank by volume, its most cited article received only eight citations, with a FWCi of 0.75, suggesting that quantity does not necessarily translate to impact in this journal's case.

By evaluating the CiteScore, it was discovered that *Journal of King Saud University - Computer and Information Sciences* held the highest metric among the top ten journals, with a CiteScore of 10.5 in 2023. This journal also produced the most cited article in the dataset “Building an Arabic Sentiment Lexicon Using Semi-Supervised Learning” which received 77 citations and a strong FWCI of 3.93. On the other hand, *RomanoArabica* recorded a CiteScore of 0.0 and showed limited citation activity, despite being among the top publishers in terms of volume. It is noteworthy that *Theory and Practice in Language Studies* produced a highly impactful article with 28 citations and a remarkable FWCI of 7.14, outperforming even higher-CiteScore journals. This variation illustrates that while CiteScore is a useful indicator of general journal quality, it does not always align with the citation performance of individual articles, especially in niche fields like Arabic vocabulary studies where specific topics may suddenly gain academic traction.

Apart from CiteScore, the Field-Weighted Citation Impact (FWCI) reveals deeper insight into research influence. For instance, *Theory and Practice in Language Studies* achieved a FWCI of 7.14, indicating that its article was cited over 600% more than the global average in similar subject areas. This high FWCI underscores the relevance and timeliness of its topic translating culture-bound expressions in today’s multilingual and politically sensitive world. Conversely, journals such as *Language Related Research* and *International Journal of Speech Technology*, despite being technically sound and well-established, recorded lower FWCI scores (0.29 and 0.64, respectively), pointing to limited global engagement with the specific subtopics they published. These disparities emphasize the importance for authors not only to target journals with reputable metrics but also to ensure alignment between their article themes and the current scholarly appetite. Ultimately, while CiteScore and FWCI are crucial quantitative indicators, selecting the right journal should also involve qualitative analysis of thematic fit and audience relevance.

3.3 Leading Countries And Institutions

Fig. 3 illustrates the list of the top 10 most productive countries along with their leading academic institutions in the field of Arabic vocabulary research. Collectively, the top three countries Saudi Arabia, Jordan, and the United States account for 556 publications, covering over 41% of the total scholarly output in this domain. Saudi Arabia emerged as the leading contributor with 208 publications (TPc), reflecting a strong national commitment to Arabic language research. Notably, King Saud University contributed 38 papers, matching the highest institutional output among all listed universities. Jordan and the United States followed closely with 174 publications each, but with differing collaboration dynamics and institutional footprints. The University of Jordan matched King Saud University with 38 articles, while Indiana University Bloomington, the U.S.’s top institution, contributed only 7 publications indicating a more dispersed institutional landscape in American scholarship.

Table 1: The topmost productive journals on LWST research with their most cited article from 1979 to 2024.

Rank	Journal	TP (%)	CiteScore (year)	The most cited article reference)	Times cited	FWCI	Publisher
1.	Dirasat Human and Social Sciences	37 (2.32)	0.1 (2023)	Translating Extra-linguistic Elements for Dubbing into Arabic: The Case of the Simpsons	8	0.75	The University of Jordan
2.	Theory And Practice in Language Studies	22 (1.38)	0.9 (2023)	A Corpus-Assisted Translation Study of Strategies Used in Rendering Culture-Bound Expressions in the Speeches of King Abdullah II	28	7.14	Academy Publication
3.	Babel	20 (1.25)	1.0 (2023)	The interaction between text difficulty and translation accuracy	25	-	Babel-Verlag
4.	Journal Of Semitic Studies	20 (1.25)	0.6 (2023)	New data on the delateralization of dcombining dot belowā and its merger with zcombining dot belowā' in classical Arabic: Contributions from old south Arabic and the earliest Islamic texts on dcombining dot below / zcombining dot below minimal pairs	13	0.98	Oxford University Press

5.	Journal Of King Saud University Computer and Information Sciences	19 (1.19)	10.5 (2023)	Building an Arabic Sentiment Lexicon Using Semi-Supervised Learning	77	3.93	Springer Nature
6.	Romanoarabica	18 (1.13)	0.0 (2023)	The grammaticalization of the modal particles in south Iraqi Arabic	7		Bucharest University Press
7.	International Journal of Speech Technology	17 (1.06)	5.0 (2023)	Arabic broadcast news transcription system	38	0.64	Springer Nature
8.	Journal Of Language Teaching and Research	16 (1.00)	1.0 (2023)	The impact of linguistic and cultural competence on translation quality: Pedagogical insights into translation problems	14	1.26	Academy Publication
9.	Language Related Research	15 (0.94)	0.9 (2023)	Phonological processes in Qur'an: The case study of Ibdāl	2	0.29	Tarbiat Modares University
10.	Languages	15 (0.94)	1.4 (2023)	Age, Experience and Language and Literacy Skills in English-Arabic Speaking Syrian Refugees	3	1.35	Multidisciplinary Digital Publishing Institute (MDPI)

TP: Total Publication;FWCI: Field-Weighted Citation Impact

An examination of the Single Country Publication (SCP) rate reveals insightful collaboration patterns. Israel demonstrated the highest intra-national collaboration with 84.62% of its research being authored exclusively by Israeli institutions. Jordan (79.31%) and the United States (73.56%) also exhibited strong domestic collaboration. These high SCP values suggest a well-established and internally driven research ecosystem, particularly in the Middle Eastern countries where Arabic vocabulary research is closely tied to national identity and educational priorities. Conversely, the United Kingdom recorded the lowest SCP among the top 10 at just 40%, indicating a high rate of international collaboration. While lower SCP may at first seem less cohesive, it is a testament to global integration, as UK-based scholars appear to actively engage in cross-border research networks.

Furthermore, the institutional output sheds light on how research efforts are centralized or decentralized within each country. The University of Haifa in Israel produced an impressive 53 publications the highest among all institutions listed highlighting the university's specialization and depth in Arabic linguistic studies. This may be reflective of the country's sociolinguistic landscape, where Arabic plays a significant role despite being a minority language. In contrast, countries like the United States and the United Kingdom, while ranking high in overall output, show more distributed institutional contributions, with no single university dominating the national output. Malaysia, represented by Universiti Kebangsaan Malaysia with 20 articles, exemplifies a growing regional interest, likely fueled by multilingual policies and the importance of Arabic in Islamic education. These findings suggest that Arabic vocabulary research is both a regional academic priority and an emerging global interest, with key institutions playing pivotal yet varied roles in driving scholarship forward.

Figure 4 illustrates the distribution of countries based on their respective continents. For example, the items with the red color represent the countries from Asia, green for Africa, blue for Europe, yellow for America and purple for Oceania. The distance and thickness of the lines between a country indicate the strength of their relatedness, which can correspond to the degree of collaborative works in the Arabic vocabulary research area. Based on the co-authorship analysis, Saudi Arabia appeared as the most productive country with 208 total publications, followed by Jordan and the United States, both with 174 publications. Countries such as Israel (117), the United Kingdom (95), Malaysia (83), Egypt (77), France (65), Spain (63), and the United Arab Emirates (63) also showed strong contributions. The leading academic institutions include King Saud University and The University of Jordan with 38 publications each, followed closely by the University of Haifa (53), Universiti Kebangsaan Malaysia (20), and CNRS Centre National de la Recherche Scientifique (18).

The high strength of international collaboration observed in countries like Saudi Arabia, Jordan, and the United States can be attributed to several factors, including strategic language planning, international academic networking, and the presence of strong institutional research capacities. For instance, the high Single Country Publications (SCP) percentage for Jordan (79.31%) and Israel (84.62%) reflects intensive national research efforts, possibly driven by local academic interest and linguistic heritage. Meanwhile, countries such as Malaysia and Egypt exhibit more balanced SCP ratios, suggesting active international collaboration. The

central positioning of countries like the United States and France in the co-authorship network may also reflect the involvement of diaspora scholars, research funding programs, and participation in global Arabic linguistic studies. These findings underline the importance of global academic mobility and regional leadership in fostering impactful research in Arabic vocabulary.



Rank	Country	TPc	S.C.P (%)	The most productive academic institution	TPi
1	Saudi Arabia	208	60.09%	King Saud University	38
2	Jordan	174	79.31%	The University of Jordan	38
3	United States	174	70.68%	Indiana University Bloomington	7
4	Israel	117	84.61%	University of Haifa	53
5	United Kingdom	95	40%	Swansea University	5
6	Malaysia	83	68.67%	Universiti Kebangsaan Malaysia	20
7	Egypt	77	48.05%	Cairo University	13
8	France	65	47.69%	CNRS Centre National de la Recherche Scientifique	18
9	Spain	63	69.84%	Universidad de Granada	7
10	United Arab Emirates	63	47.61%	University of Sharjah	13

Figure 3: The Top 10 Leading Countries And Academic Institutions In FLT Research Field

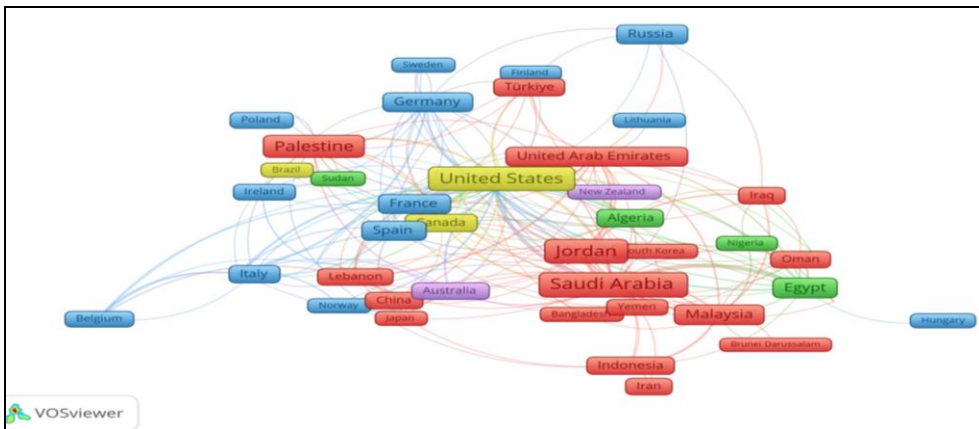


Figure 4: A Bibliometric Map Fabricated According To The Co-Authorship In Network Visualization Mode.

3.4 Leading Authors

Based on the data extracted from the Scopus database, Table 2 tabulates the ten most prominent authors in the field of Arabic vocabulary research. These authors are affiliated with institutions across eight countries: Israel (3 authors), Jordan (2 authors), Saudi Arabia (1 author), Australia (1 author), Tunisia (1 author), the United Arab Emirates (1 author), and Spain (1 author). The timeline of their first publications spans from as early as 1990 (Asaid Khateb) to the most recent in 2017 (Ibrahim Abdalla Asadi), illustrating the continuity and emerging interest in this niche. Interestingly, two authors Asaid Khateb and Raphiq Ibrahim have long-standing publication records (since 1990 and 1996, respectively) and are affiliated with the University of Haifa, suggesting a strong institutional focus on Arabic psycholinguistics and vocabulary acquisition. Notably 9 out of the 10 listed authors were the first authors in at least one of their key publications, highlighting their leading roles in driving Arabic vocabulary research.

Manuel Perea from Spain stands out with the highest total number of publications (292), a remarkable h-index of 53, and a citation count of 4809, placing him at the forefront in terms of academic impact. Following closely is Asaid Khateb, with 105 publications, a high h-index of 36 and an impressive citation count of 4095. Both scholars demonstrate sustained influence, possibly due to interdisciplinary contributions involving cognitive science and neurolinguistics. Moreover, younger researchers such as Ibrahim Abdalla Asadi (first publication in 2017) and Marwan A. Jarrah (first publication in 2015) have rapidly built a significant presence in the field an indicator of growing global interest in Arabic vocabulary, particularly in relation to second language acquisition and applied linguistics. The clustering of top authors within certain institutions like the University of Haifa and The University of Jordan may reflect the existence of localized research hubs, driven by collaborative networks and region-specific language challenges.

Table 2: Top Ten Prolific Authors

Rank	Author	Scopus Author ID	Year of 1st publication	Total publication	h-index	Total citation	Current affiliation	Country
1	Ibrahim, Raphiq	7102593444	1996b	56	18	1247	University of Haifa	Palestine
2	Asadi, Ibrahim Abdalla	57192683379	2017b	32	9	268	University of Haifa	Palestine
3	Jarrah, Marwan A.	57193088521	2015a	82	12	190	The University of Jordan The institution will open in a new tab	Jordan
4	Masrai, Ahmed M.	56674700900	2015a	23	10	251	Prince Sattam Bin Abdulaziz University	Saudi Arabia
5	Omar, Abdulfattah A.	57190662485	2016a	65	10	250	The Australian National University The institution will open in a new tab	Australia
6	Farghal, Mohammed	56487819800	1991b	43	11	348	Applied Science Private University	Jordan
7	Khateb, Asaid	7003972821	1990a	105	36	4095	University of Haifa	Palestine
8	Zrigui, Mounir	8951841700	2005b	170	19	609	Université de Monastir	Tunisia
9	Boudelaa, Sami	6603233995	1993b	26	13	466	United Arab Emirates University	United Arab Emirates
10	Perea, Manuel	7005089971	1996a	292	53	4809	Universidad Antonio de Nebrija	Spain

*Role in co-authorship, superscripts.

a. First author.

b. Co-author.

interdisciplinary approach in Arabic vocabulary research where vocabulary is studied not only from a linguistic but also from a cognitive and pathological perspective.

In terms of research area distribution, the Social Sciences and Humanities domain dominated the dataset, contributing a total of 2,286 publications, far surpassing other categories. This was followed by the STEM field with 601 publications, while Health and Life Sciences accounted for 161. Meanwhile, Education and Multidisciplinary studies contributed 28 publications, and Engineering & Applied Sciences trailed behind with only 8. The overwhelming representation of the Social Sciences and Humanities suggests that Arabic vocabulary research is deeply rooted in linguistic, cultural, and educational paradigms rather than technical or medical applications. The modest contribution from Health and Life Sciences, however, is coherent with the emergence of terms like *aphasia*, which signals growing interest in clinical linguistics and language disorders within Arabic-speaking contexts. The limited presence in Engineering and STEM may indicate a lack of integration of Arabic vocabulary studies in computational linguistics or language technology, an area that presents vast potential for future exploration. This distribution underscores a critical need for more cross-disciplinary research that bridges traditional linguistic inquiry with modern technological and clinical applications.

b) Topics of Interest

Recent fluctuations in scholarly interest and emergent themes within Arabic vocabulary research have notably illuminated the integration of advanced computational methodologies, particularly deep learning, into this domain. The prominence of "deep learning" in recent bibliometric analyses signifies a transformative shift in how Arabic vocabulary studies are approached. This significant interest aligns with the advent of techniques such as automatic feature extraction and sentiment analysis, which are crucial for enhancing natural language processing (NLP) applications specific to Arabic, including tasks like text classification and machine translation, as highlighted by Antoun et al. (Antoun et al., 2020).

Moreover, the adaptation of BERT-based models for Arabic, epitomized by AraBERT, has played a pivotal role in advancing performance metrics across vocabulary comprehension, named entity recognition, and semantic similarity assessments (Antoun et al., 2020). AraBERT's architecture allows for effective tokenization and leverages a substantial pre-training dataset, vital for navigating the morphological complexities of the Arabic language (Bensoltane & Zaki, 2021). The intersection of these advanced models with deep learning frameworks illustrates a significant advancement in Arabic NLP, fostering improved understanding and utilization of Arabic vocabulary.

The emergence of sociolinguistic themes, encapsulated by keywords such as "gender" and "social media," further signifies a multidimensional approach within the field. This intersection of language processing and social variables is essential for comprehensively understanding how digital communication influences lexical practices among Arabic speakers (Boulouard et al., 2022). Research has

demonstrated that socio-cultural factors significantly affect language production and perception, especially in multilingual and diglossic contexts prevalent in Arab societies (Sari, 2023). Mutholib et al. underscore the necessity of mastering vocabulary as a foundational element for language acquisition, which resonates with the current emphasis on cultural and societal nuances in Arabic vocabulary research (Mutholib et al., 2020).

The focus on vocabulary emphasizes its enduring relevance in enhancing language skills. This is critical not only for learners but also for educators seeking effective strategies and materials to aid vocabulary acquisition in Arabic (Mutholib et al., 2020). The significant publication activity around this theme suggests a heightened awareness of the role vocabulary plays in academic success and fluency in Arabic, further reiterating the integration of both AI-driven methodologies and socio-linguistic perspectives.

In summary, the current landscape of Arabic vocabulary research is marked by an increased focus on integrating sophisticated computational techniques alongside sociolinguistic insights. This dual approach illuminates how emerging technologies can enhance traditional vocabulary studies, making them more relevant in both academic and practical applications within Arabic language learning and usage.

3.6 Study Limitations

By limiting the search to terms “Arabic” in conjunction with “vocab*” or “lexical*” within titles and abstracts, the results presented by Scopus may not capture the entirety of Arabic vocabulary related research. This exclusion may occur as some researchers may adopt alternative expressions such as “terminology,” “linguistic items,” or “semantic fields” which do not directly align with the specified search string, thereby escaping retrieval. Furthermore, the refinement process, which excluded 16 review articles and restricted inclusion to final-stage journal articles, while improving specificity, may have inadvertently filtered out conceptual or emerging works crucial to foundational insights in the field.

In addition, the co-occurrence analysis conducted via VOSviewer might have overlooked low frequency but semantically rich keywords due to its minimum occurrence threshold, potentially obscuring niche but rising trends. This narrow funneling of data despite yielding 1,592 focused articles may reflect a deeper structural limitation: the fragmented terminological landscape in Arabic lexical studies. This fragmentation could stem from varied regional terminologies, diglossia, or transliteration inconsistencies across academic communities. Therefore, to enhance comprehensiveness and cross-regional inclusivity, future bibliometric studies are encouraged to triangulate Scopus data with other databases such as Web of Science, Dimensions, or Arabic-centric repositories. This would allow a more holistic capture of the field’s evolution and better illuminate underrepresented trajectories in Arabic vocabulary scholarship.

4.0 CONCLUSION

This study was conducted to present a comprehensive overview of the research trajectory on Arabic vocabulary from the year 1970 to 2024 using the Scopus database, incorporating 1,592 selected journal articles following a meticulous filtering process that excluded review papers and non-final publications. Initial publication activities before the 1990s were minimal, with fewer than 10 articles per decade, but a gradual and then exponential rise was observed from the early 2000s onward. The publication count reached a remarkable 179 articles in 2024 alone, marking an over 1100% increase since the year 2000. This explosive growth reflects the shifting academic focus towards Arabic lexical studies, particularly considering technological advancements, increasing global interest in Arabic as a second language and interdisciplinary intersections with computational linguistics. Bibliometric mapping through VOSviewer also highlighted those terms such as adaptation, aphasia, and deep learning have gained substantial traction, suggesting a merging of traditional linguistic paradigms with cognitive sciences and AI technologies. Countries such as Saudi Arabia, Jordan, and the United States emerged as leading contributors, with institutions like King Saud University and the University of Haifa anchoring much of the scholarly output. Interestingly, while countries like Israel demonstrated strong national research ecosystems, the United Kingdom and Malaysia showed higher tendencies toward international collaboration, offering fertile ground for future partnerships. In sum, this study not only charts the prolific rise of Arabic vocabulary research but also captures the evolving interdisciplinary landscape and its implications for language learning, cultural preservation, and digital innovation.

Author Contributions

Johan, N. A. H. and Asbullah, L. H.: Introduction, discussion on the organization of ideas, language style review, and final draft editing; the overall preparation of the article was carried out collectively.

Conflict Of Interest

This manuscript has not been published elsewhere, and all authors have agreed to its submission and declare no conflict of interest regarding the manuscript.

REFERENCES

- Akla, A. (2022). Storytelling method to improve vocabulary for non-native Arabic speaker children. *Arabiyatuna Jurnal Bahasa Arab*, 6(1), 171-192.
- Antoun, W., Baly, F. & Hajj, H. (2020). Arabert: transformer-based model for Arabic language understanding. *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pp. 9-15.

- Arruum, B., Wardah, W. & Yuniarti, Y. (2022). Designing storytelling instruction for senior high school speaking. *Pena Kreatif Jurnal Pendidikan*, 11(2), 197-202.
- Bensoltane, R. & Zaki, T. (2021). Comparing word embedding models for Arabic aspect category detection using a deep learning-based approach. *E3s Web of Conferences*, 297(6), 01072.
- Boulouard, Z., Ouaiassa, M., Ouaiassa, M., Krichen, M., Almutiq, M. & Gasmi, K. (2022). Detecting hateful and offensive speech in Arabic social media using transfer learning. *Applied Sciences*, 12(24), 12823.
- Hà, T. & Bellot, A. (2020). Assessing storytelling as a tool for improving reading comprehension in the efl primary classroom. *English Teaching Practice & Critique*, 19(2), 169-196.
- Mohamad, N., Baharun, M., Rahman, A., & Ramli, Z. (2021). Receptive and productive arabic academic vocabulary mastery: a study of students in malaysian public universities. *International Journal of Asian Social Science*, 11(3), 154-167.
- Mutholib, A., Munjiah, M. & Muflichah, S. (2020). The use of android-based dictionary games for Arabic vocabulary learning in elementary school. *Proceedings of the 2nd International Conference on Quran and Hadith Studies Information Technology and Media & the 1st International Conference on Islam, Science and Technology*, <https://doi.org/10.4108/eai.2-10-2018.2295457>
- Permatasari, F. & Mauludiyah, L. (2022). Improving students' Arabic vocabulary by memorizing with cognitive retroactive transfer (crt). *Alsinatuna*, 8(1), 33-50.
- Sari, L. (2023). Content analysis of animation video instagram account @marhaban.academy as a medium for learning arabic vocabulary. *Journal of Arabic Language Learning and Teaching*, 1(1), 25-34.