

Mapping the Evolution of Arabic Language Research: A Bibliometric Approach

رسم خريطة لتطور البحث في اللغة العربية: منهج بيبليومتري

Abdulroya Benseng,^{1, *} Sakirin Tam^{2,}¹ Academy of Islamic and Arabic Studies Princess of Naradhiwas University, Narathiwat, Thailand² Faculty of Information Communication Technology, University of Puthisastra, Cambodiaعبد الرويا بنسينغ^{١, *}، ساكيرين تام^٢¹ أكاديمية الدراسات الإسلامية والعربية، جامعة الأميرة نارادياس، ناراثيوات، تايلاند.² كلية تكنولوجيا المعلومات والاتصالات، جامعة بوئيساسترا، كمبوديا.

ABSTRACT

This bibliometric analysis examines the progression of research on the Arabic language using a thorough dataset obtained from Scopus and shared on GitHub. This study uses quantitative measurements to analyze the field's progression, pinpointing significant trends, notable contributors, and thematic emphasis areas. The analysis shows a strong connection between linguistic studies and technology improvements through the incorporation of computer approaches like natural language processing and deep learning. Geographical contributions demonstrate a worldwide interest in Arabic language studies, with Saudi Arabia, Malaysia, Jordan, and Tunisia being the top countries in terms of scholarly output and citation effect. The study delves into a detailed examination of sociolinguistic aspects, specifically focusing on gender, in Arabic language research. The discipline is expanding dynamically and showing interdisciplinary characteristics through a consistent growth in research output and a diverse range of themes indicated by keyword analysis. The co-occurrence network analysis demonstrates the interconnection of research issues, with "Arabic languages" serving as a core node connecting several sub-disciplines. This bibliometric analysis illustrates the expansion and thematic development of Arabic language research, highlighting its growing complexity and worldwide influence. This study enhances comprehension of the present state and future trends of Arabic language studies by emphasizing the use of computational technology and the importance of international cooperation.

الخلاصة

يفحص هذا التحليل البيبليومتري تقدم البحث في اللغة العربية باستخدام مجموعة بيانات شاملة تم الحصول عليها من Scopus ومشاركتها على GitHub. تستخدم هذه الدراسة قياسات كمية لتحليل تقدم المجال، وتحديد الاتجاهات الهامة، والمساهمين البارزين، ومجالات التركيز المواضيعية. يُظهر التحليل وجود علاقة قوية بين الدراسات اللغوية والتحسينات التكنولوجية من خلال دمج مناهج الكمبيوتر مثل معالجة اللغة الطبيعية والتعلم العميق. تُظهر المساهمات الجغرافية اهتمامًا عالميًا بدراسات اللغة العربية، حيث تعد المملكة العربية السعودية وماليزيا والأردن وتونس الدول الأولى من حيث الإنتاج العلمي وتأثير الاستشهاد. وتتناول الدراسة دراسة تفصيلية للجوانب الاجتماعية اللغوية، مع التركيز بشكل خاص على النوع الاجتماعي، في أبحاث اللغة العربية. يتوسع التخصص بشكل ديناميكي ويظهر خصائص متعددة التخصصات من خلال النمو المستمر في مخرجات البحث ومجموعة متنوعة من المواضيع التي يشير إليها تحليل الكلمات الرئيسية. يوضح تحليل الشبكة المتزامنة الترابط بين قضايا البحث، حيث تعمل "اللغات العربية" كعقدة أساسية تربط بين العديد من التخصصات الفرعية. يوضح هذا التحليل البيبليومتري التوسع والتطور المواضيعي لأبحاث اللغة العربية، مع تسليط الضوء على تعقيدها المتزايد وتأثيرها العالمي. تعزز هذه الدراسة فهم الوضع الحالي والاتجاهات المستقبلية لدراسات اللغة العربية من خلال التأكيد على استخدام التكنولوجيا الحاسوبية وأهمية التعاون الدولي.

Keywords

الكلمات المفتاحية

Arabic Language , Bibliometric Approach, Arabic Language Research
اللغة العربية، المنهج البيبليومتري، بحوث اللغة العربية.

Received

استلام البحث

5/10/2023

Accepted

قبول النشر

19/12/2023

Published online

النشر الإلكتروني

10/1/2024

1. INTRODUCTION

Bibliometric analysis is a quantitative approach used to assess the impact and trends within a specific field of academic research. By examining various metrics such as publication counts, citation analysis, and co-occurrence of keywords, this method provides insights into the development, structure, and dynamics of scholarly disciplines. It enables researchers to identify influential authors, documents, and themes, as well as to track the evolution of knowledge over time[1].

This paper intends to use bibliometric analysis to trace the evolution of research trends dedicated to Arabic language studies. Our journey starts with a literature review, where we shall be pulling the thread of existing studies to lay a foundation for further understanding of the field's background and current state of affairs. Methodology section, which covers the process and the criteria for data capture and analysis, stresses Scopus as a primary data source and the R language for data manipulation and visualization.

The findings are discussed in the result section expressing the production levels based on an annual scale, the most relevant sources & authors, and geographical contributions & thematic focuses of the Arabic language research. This section aims to answer several pertinent questions: What has the rise in the volume of research on the Arabic language over the years? Which are the nations and organizations in the frontline of this academic undertaking? What kind of themes and approach are being the most often used in this field?

The discussion integrates these findings, exploring the implications of the observed trends and the interconnections between different facets of the research landscape. It examines the role of computational methods in Arabic language studies, the global distribution of research contributions, and the thematic diversity within the field. Finally, the conclusion synthesizes the main outcomes of the bibliometric analysis, reflecting on the growth, challenges, and future directions of Arabic language research. By providing a holistic view of the field, this paper contributes to a deeper understanding of its development and identifies areas ripe for further investigation.

2. METHODOLOGY

The purpose of this bibliometric study is to chart the development of Arabic language research by reviewing all relevant literature. Due to its large coverage of peer-reviewed literature and its wide acceptability in academic research for trustworthiness and comprehensiveness, Scopus is justified as the major database for our technique, which incorporates an organized approach to data gathering, analysis, and interpretation. What follows is an account of the search methodology, data extraction steps, analytic software, and constraints that were faced.

2.1. Database and Search Strategy

Scopus database is chosen for this work because it profoundly indexes journals that contribute a sizeable number of the articles related to the Arabic language studies. This choice incorporates Scopus, having wide and trustworthy coverage on scientific and research publications, thus enabling a complete bibliometric study. The research was organized by a special keyword plan to avoid the situation of the unrelated documents. The keywords used for searching the database were confined to the article titles and keywords only, utilizing the search string: TITLE ("arabic language") AND KEY ("arabic language"). Use of this analytical technique helped us to be effective in our core research area by selectively extracting those studies that have no relations and thus, use the proper dataset for analysis.

2.2. Data Collection

There were 431 papers that were found to be pertinent to the subject of Arabic language research. Our research was based on these materials, which included a wide range of academic outputs such as articles, reviews, and conference papers. This analysis relied heavily on accurate and comprehensive bibliographic metadata, which included the following fields: authors, document type, language, publication year, title, total citations, abstract, affiliation, keywords, DOI, and cited references.

2.3. Data Analysis Tools

For the data analysis, we employed R Studio, benefiting from the R programming language. The biblioshiny package was a good choice since it could manipulate the bibliometric data in a flexible way and also fit well for generating informative figures and tables. This package was the enabler of the isolation, processing and illustration of data and it enabled advanced analysis of the history of Arabic linguistic researches.

2.4. Limitations and Data Quality

When it came to bibliographic metadata, the investigation ran across problems with its completeness. Some information elements, including DOI and keywords plus, were partially or entirely missing from the dataset; others, like cited references and science categories, were entirely missing as well. Notwithstanding these caveats, the research did not depend entirely on the missing metadata to draw conclusions. Title, abstract, affiliation, keywords, DOI, and corresponding author data were less comprehensive than areas like author, document type, journal, language, and publication year, among others. The

quality of the information varied. Table 1 of the study explains how a cautious approach was needed in data interpretation and analysis due to the difference in metadata quality. This study's technique takes advantage of Scopus's extensive database, as well as RStudio and the biblioshiny package, to overcome the obstacles given by missing information. The academic community will benefit greatly from this method since it guarantees a thorough and concentrated examination of the development of research on the Arabic language.

TABLE I. COMPLETENESS OF BIBLIOGRAPHIC METADATA

Status	Missing %	Missing Counts	Description	Metadata
Excellent	0	0	Author	AU
Excellent	0	0	Document Type	DT
Excellent	0	0	Journal	SO
Excellent	0	0	Language	LA
Excellent	0	0	Publication Year	PY
Excellent	0	0	Title	TI
Excellent	0	0	Total Citation	TC
Good	0.93	4	Abstract	AB
Good	0.93	4	Affiliation	C1
Acceptable	13.69	59	Keywords	DE
Poor	23.9	103	DOI	DI
Poor	40.6	175	Keywords Plus	ID
Poor	45.48	196	Corresponding Author	RP
Completely missing	100	431	Cited References	CR
Completely missing	100	431	Science Categories	WC

3. BIBLIOMETRIC ANALYSIS DELIVERABLES

This section presents the results of a detailed bibliometric analysis of the evolution of Arabic language research, by drawing on a meticulously curated dataset from Scopus. This in-depth examination is further divided into different themes, each of which reflects different facets of the academic ecosystem within this domain. We start by giving an overview of the main core dataset statistics collected, namely publication count by type, source-wise publication, and annual growth, in order to provide a macroscopic view of research output over time. We continue by analyzing yearly scientific production, as we uncover trends in publication volume that is commensurate with expanding interest and scholarly activity in this field.

We move on to the analysis of average citations per year, as this provides an interesting view into the impact and influence of the research being produced within this domain. We look into the most relevant sources and authors, as we uncover the key journals, conferences and contributors who are driving the discourse in Arabic language studies. We look into the geographical distribution of corresponding authors and citation patterns of different countries, as we uncover the global nature and collaborative aspect of research in this domain. We next move to deeper thematic analysis, as we identify the most cited countries and the prevalence of topics under research, as revealed by keyword frequency. Finally, we present a co-occurrence network analysis of these topics, exposing the interrelatedness of topics and the complex web of themes that underpin Arabic language research. Each sub-section of this results section builds to a comprehensive insight into the dynamics of the field and to the key outcomes and insights from the bibliometric analysis.

3.1. Main information

The most important point that can be gathered from Scopus as the data source for the aspired bibliometric analysis allows a comprehensive view of a field progress through the 1946-2024 period. This period covers a number of 431 documents that were published on 302 different sources, and sources are less than the amount of journals, books, and so on. The presentation of this example testifies to the stable and sustainable rate of 1.42% per year, which is the sign of continuous research process over the period. As regards the average age of document 6.8 years indicates that the majority of the research is relatively new and reveals ongoing interests and research progress in this field. Additionally, there is an average citations rate of 4.856 documents which is an evidence of the value and importance of these researches explored in this field. This is definitely the case, as indeed, sources were all lacking in the data set hinting at a critical shortcoming in the study.

The abundance of keywords in these documents—1,865 Author's Keywords (DE) and 1,866 Keywords Plus (ID)—shows how varied and specific the subjects covered by Arabic language studies are. With such a wide range of terms, it's easy to see how the area of Arabic language research is complex and multi-faceted.

According to the numbers, 1,175 people have written for this collection of work, with 70 of those people penning individual papers. There is a clear indication of a strong sense of community collaboration here, even though there are still 71 documents with a single author who contributed significantly to the research. Moreover, with an average of 3.12 co-authors

per document and an impressive 22.51% of documents including international co-authorships, the global reach and cooperative nature of Arabic language studies are further demonstrated by the collaboration.

The information on the type of documents that break down gives further views on the research output, with articles at 246 becoming the most frequent document type, in comparison to the 168 conference papers. This gives a clear view on the role of just recording research findings in both peer-reviewed journal articles and conference proceedings. Alongside the more common documents like book chapters and reviews, we have a plethora of research works which lend richness and depth to the scholastic terrain.

Figure 1 brings to light the conditions of Arabic language research, which is a lively and prospering field filled with diversified topics, co-operative research initiation, as well as a combination of traditional and modern approaches of information dissemination. The continuous yearly growth pace along with the number documentary pieces and authors is a clear representation of the dynamic and changing manner of Arabic literature studies while data which is incomplete solicits the need for better and improved bibliographic methods.

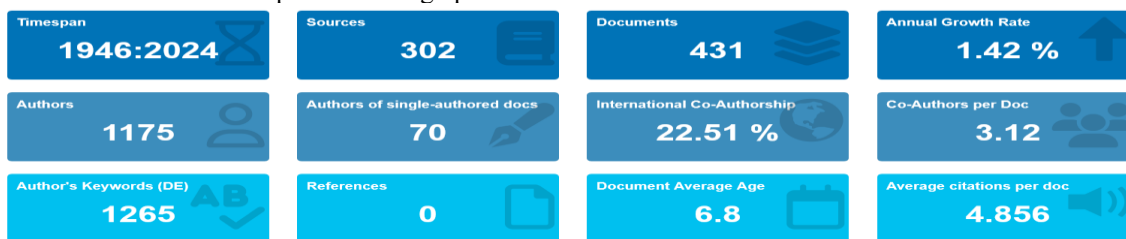


Fig.1. Main information

3.2. Annual scientific production

Figure 2 shows the yearly scientific output in Arabic language research from 1946 to 2024. It tells an interesting story of how the field has grown over nearly eight decades. From its 1946 debut in a single article to its long hiatus until its 1987 modest return, the timeline traces its events in chronological order. Up until a considerable increase in the early 21st century, this pattern of irregular publishing persisted, with a few articles emerging here and there.

When we look at the numbers more closely, we can see that there is a major turning point about 2001. After that, the number of publications starts to rise, and the yearly scientific output becomes more consistent. The increase is most noticeable beginning in 2008, when the yearly article count goes from the low single digits to the high double digits, with the highest points being in 2021 and 2023, with 49, 50, and 54 articles, respectively. During this time, there was a surge in both public and academic interest in Arabic language studies around the world, which led to the field's most rapid and consistent expansion to that point.

It is too early to establish the annual trend, but the three articles covering the year 2024 show that the field is still productive. From 1946 to 2024, the discipline of Arabic language research seems to have matured and attracted more interest, as shown by the increasing volume of submissions.

Language and cultural studies have advanced, research networks have grown, and digital platforms for publication and collaboration have become more accessible, all of which contribute to the yearly increase in scientific production. One possible explanation for this expansion is the increasing number of articles with co-authors from outside of the Arab world; this trend suggests that a more inclusive and varied group of scholars is working together to deepen our grasp of Arabic.

The information in the form of the Figure 3 indicates that there were times of stagnant and short-lived research in Arabic as well as periods of doing not much and just a growing research in the field of Arabic studies quite on a gradual basis, after which came the time for the research in Arabic to prosper. This development is one of the indications of how alive and dynamic scholarly research methods are and the fact that it contacts the developments in academic history and inquiry.



Fig.2. Annual scientific production

3.3. Average Citations

Figure 3 depicts the Average Citations Per Year section, which offers a comprehensive evaluation of the influence and acknowledgment of Arabic language research across the years, as evaluated by citations. Using data collected from 1946 to 2024, this study provides a historical perspective on the impact and citation of research in this field.

At first glance, the mean total citations per article (MeanTCperArt) is quite low; for example, the 1946 piece only received 4 citations on average, which shows that the field was interested in it from the beginning. Articles published in the following decades, such as 1987 and 1991, have a small increase in citations per year (MeanTCperYear), suggesting a slow but consistent pace of interest in the research done during these years.

Citations skyrocketed in 2002, with one item receiving an astounding 82 citations (or 3.57 citations year on average). It is possible that the groundbreaking nature or immediate applicability of this anomaly caused it to attract a great deal of interest.

As we move into the 21st century, specifically from 2005 onwards, there's a notable increase in both the quantity of research (as indicated by N, the number of articles) and the average citations per year. This period marks a phase of enhanced visibility and influence of Arabic language research, with 2005 standing out as articles from this year received an impressive mean of 12.75 total citations per article, translating to an average of 0.64 citations per year.

The trend of citations delineates some ups and downs in subsequent years with the appearance of two plunges; they are in 2007, 2008, and especially in 2016 when the research accomplished the highest citation rate which is 7.63 MeanTCperArt, representing a peak of scholarly interest and notice. While this is contrasted by years that citation impact per year is high, such as in 2014 or 2016, that fluctuation illustrate the divergent character of research impact in the long-term perspective. The 2020-2023 data, the most recent holding period (3 years), demonstrates a combined citation pattern that varies from year to year. More particularly, it shows that the articles of 2020 and 2021 having the author best-cited is a good sign of continued interests and relevance in the field's research outcomes. Nevertheless, this decline happens in the last figure, while 2023 cites per year mean is 0.32, this number has a meaning that the new articles haven't had enough time for a while to attain the citations.

Concluding the analysis in 2024 will be the observation that MeanTCperYear is now 0, and the decrease due to the students having written the articles lately explains the fact that they do not yet have any citations. Therefore, a distinction is made between current outcry and the resulting influence caused by the delay in the accumulation of citation and the need for time to assess the impact of newly published works.

By summarizing the "Citations, Items Per Year" part, as presented in figure 4, we can see the fluctuations of recognition or impact that Arabic language research has been through decades. It shows the changing nature of the field, which is glaringly displayed by the popular periods of scholarly research and academic influence contrasting with the time-dependency issues, where the latest research articles have higher citation rates than older ones.

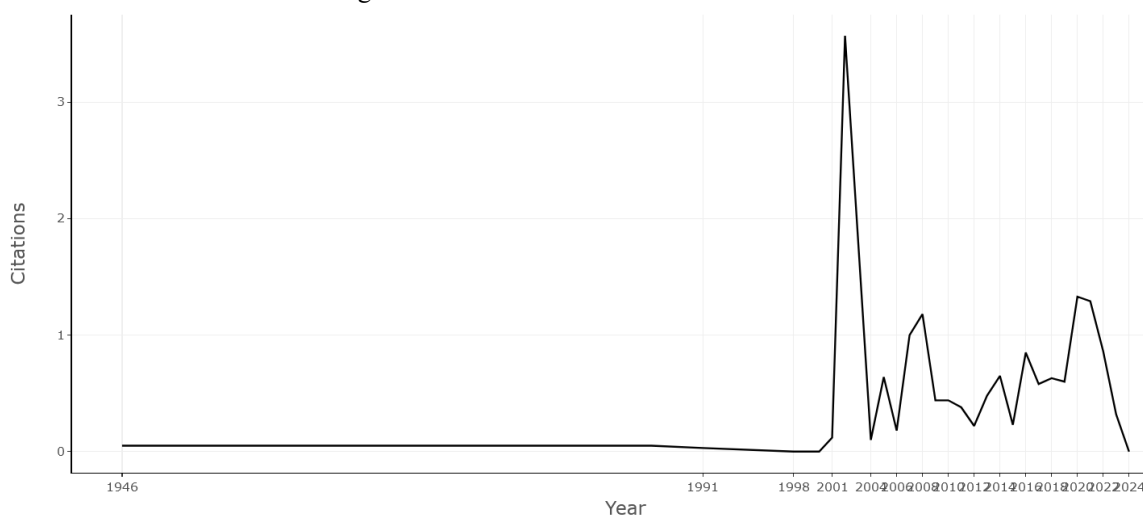


Fig.3. Average citations per year

3.4. Relevant sources

Figure 4 shows the distribution of articles across journals and conference proceedings, which, according to the bibliometric analysis on the evolution of Arabic language research, reveals a number of noteworthy sources that have made substantial contributions to the field. These sources are ranked according to the number of articles published.

Among the sources that disseminate research pertaining to Arabic language studies within the human and social sciences sector, "DIRASAT: Human and Social Sciences" stands out as the leading source with 11 articles. Because Arabic language

research encompasses more than just linguistics but also cultural, social, and educational studies, this journal is highly regarded in the area.

The next highest placing journal, “Language Related Research”, presents 8 articles, demonstrating that it is a well-used platform for the publication of linguistic studies on Arabic. As the academic community pursues studies of the intricate and nuanced Arabic language, it is clear that more and more disciplines are taking an interest in the linguistic details of the Arabic language. “Lecture Notes in Computer Science, especially the subseries on Artificial Intelligence and Bioinformatics”, comes next with 7 articles, reflecting the growing intersection of Arabic language research with computer science – with natural language processing (NLP) and AI in particular. As these technologies are brought to bear on Arabic language data, the inherent challenges, as well as the technological prospects, are being revealed.

Equally strong contributors, the “International Journal of Speech Technology” and “WANLP 2021 - 6th Arabic Natural Language Processing Workshop, Proceedings of the Workshop,” added 6 articles each that bolster the importance of speech technology and NLP as they relate to Arabic language studies by representing current leading approaches to research in this area. Their focus is on cutting-edge research in the field, such as speech recognition, synthesis, and processing technologies that focus on the Arabic language.

Other publishers provided equally important contributions. “Advances in Intelligent Systems and Computing,” “Colloquium in Information Science and Technology, CIST,” “Communications in Computer and Information Science,” “Information Sciences Letters,” and the “International Journal of Advanced Computer Science and Applications” each contributed 5 articles that represent the technologically and computationally diverse approaches being used in Arabic language studies. Their contributions run the gamut from intelligent systems and computing to information science and advanced computer applications.

Figure 4 reveals that Arabic language research is a multi-faceted domain that is inherently interdisciplinary and collaborative, with linguistics being a central bordering discipline. Moreover, the prominence of conferences in contrast to journals in this bibliometric analysis attests to a field in which there is a vibrant exchange and circulation of ideas that is fast-paced and decentralized. This interpenetration of traditional scholarly journals and recent technological fora is emblematic of a rapidly evolving domain that is indeed undergirded by innovation and collaboration, and is one in which the boundaries of investigation are ever expanding.

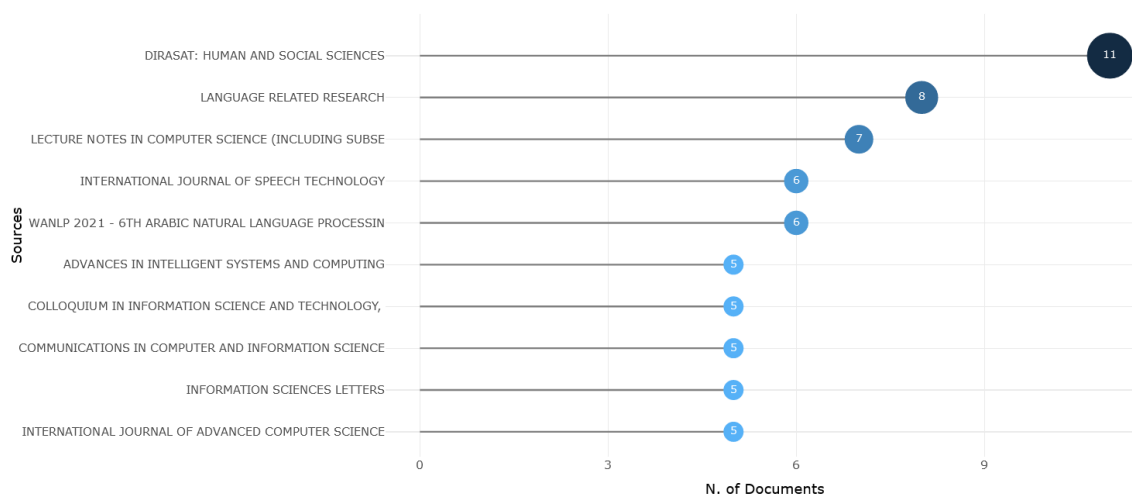


Fig.4. Relevant sources

3.5. Relevant authors

Figure 5 presents the most relevant authors in the domain of Arabic language research analyzed in this In-Depth Analysis. The chart not only lists the number of articles each author has published but also offers a fractionalized view of their contributions. This fractionalized perspective provides an extra layer of insight into the collaborative nature of their work, further fleshing out their impact on the academic landscape.

ZRIGUI M tops the table as the most prolific author, publishing a total of 10 papers and a fractionalized contribution of 3.67. Indicating not only prolific research production, but significant participation in collaborative research efforts within the Arabic language research community [2-11]. ALOULOU C and ISMAIL Z follow with 5 articles each. Their fractionalized contributions of 1.58 illustrate a commensurate level of participation in collaborative research. GHANI MTA also has 5

articles, which produce a fractionalized contribution of 1.50 that reflects a different pattern of collaboration with these authors [12-14].

DAUD WAAW, FEKIH-ROMDHANE F, HAJJAR M, HALLIT S, LAATAR R, and MALAEB D each contributed 4 articles, as well. Their fractionalized contributions differ, which reflects their differing engagements in collaborative research. DAUD WAAW and LAATAR R have a fractionalized contribution of 1.25 and a fractionalized contribution of 0.67, indicating heavier engagement in collaboration for the former group of authors as opposed to the latter group. The fractional contribution of FEKIH-ROMDHANE F, HALLIT S, and MALAEB D, is substantially lower (0.63) indicating their work may include more solo or lead authorship [12, 14-17].

This analysis illuminates the teamwork that is intrinsic to doing research in Arabic language. The variety of fractionalized contributions indicates the many fluid collaborations that can exist in academic research. Some researchers may opt for working in larger groups and diluting their fractional contribution per paper, while others elect to form smaller units working more often in just pairs or alone, which would result in their having a higher fractionalized contribution. The appearance of these authors across the numerous studies and papers points to their substantial contribution to furthering research of all kinds in Arabic language. Their work has likely spanned a variety of linguistic and computational approaches to language processing, in addition to cultural studies and beyond. The varying collaborative patterns seen here in their fractionalized contributions point to the emergence of teamwork and partnerships of all kinds as being the infrastructures that push the boundaries of knowledge and discovery in Arabic language studies. The blend of individual and combined efforts serves to amplify the academic discourse, enlivening the research environment with its dynamism.

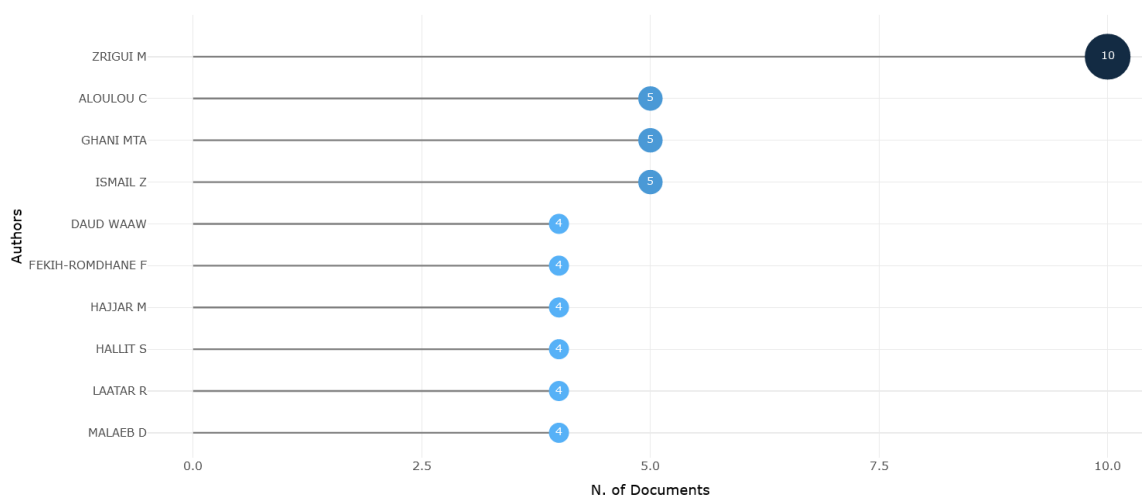


Fig.5. Relevant authors

3.6. Relevant affiliations

The bibliometric analysis of the evolution of Arabic language research also reveals the most productive affiliations in this field, as illustrated in Figure 6. This overview of affiliations serves to present a larger view of the universities dedicated to the study of the Arabic language, and research institutions that have critically shaped Arabic language research. According to this representation, King Saud University leads the chart with an impressive 21 articles. Its prominence demonstrates the university's dedication to the study of the Arabic language, and its position as one of the principal university centers for scholarly contributions to this field. The overriding presence of King Saud University underscores the importance of Arabic studies to the region, and its commitment to promoting academic research and inquiry into the language, culture, and linguistics of the Arabic-speaking world.

In second place, the University of Sfax, with 15 articles, reveals a significant research effort by Tunisia in the study of the Arabic language. The prominence of this university demonstrates its commitment to disseminating Arab and Islamic values and to promoting research in the humanities and arts for the well-being of the Arab-Muslim world, and its unique standing among the world's universities. With 14 articles, Universiti Kebangsaan Malaysia shows the strength and breadth of international interest in the scholarly study of the Arabic language. This demonstrates the university's commitment to research involving Arabic language and literature, Islamic history and civilization, Islamic banking and finance, and Arabic class for adult at international level; again indicating that research and learning in the areas of Arabic language and culture are thriving outside of the Arab world.

Yarmouk University, and The University of Jordan, are next, with 12 and 10 articles respectively. Jordan's places further underscore the leadership role Jordanian universities have taken in this field, and the extensive research that is currently

taking place there on all aspects of the Arabic language. In this environment, it is clear that Jordanian scholars have uniquely well positioned to contribute to the development of this field, and to play a decisive role in our understanding of the Arabic language. Other noteworthy affiliations include the University of Monastir, with 9 articles, Universiti Pendidikan Sultan Idris, with 7, University of Genoa, with 8, The University of Jordan (with 7), and Tarbiat Modares University, with 7 of its own. This wide-ranging list not only underscores the broad geographical distribution of universities engaged in leading this study, but the diversity of interdisciplinary research that is currently being inspired and carried out across these universities. The distribution of articles among these affiliations, further described in Figure 6, illustrates the collaborative and widely distributed nature of research on the Arabic language. It unspools a rich network of critical scholarship from universities and institutions around the world, deeply engaging analytical projects and pedagogical applications of the Arabic language. This rigorous, widespread approach to these aspects alone of the Arabic language provides overarching insights into the research and contribution of this language itself across the globe. It further emphasizes its modern and strictly framework and articulation, and its adaptability to rigorous and large-scale examination.

In the first fifteen years of the twenty-first century alone, then, Arabic language studies has completed a rigorous, sustained and interdisciplinary journey from a mostly if not exclusively institutionally-based set of analyses to critical and imaginative interfaces of knowledge, culture, technology and modern languages. This journey will undoubtedly continue to unfold in unpredictable and imaginative ways as Arabic language studies evolves and reorganizes, and it pushes and is pushed by other languages and discourses in critical, reflective and inventive directions.

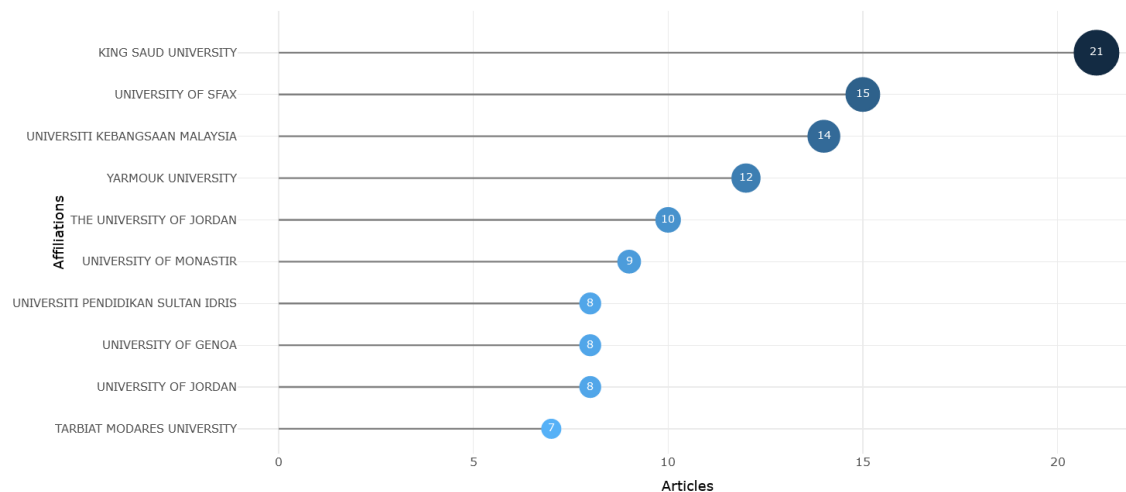


Fig.6. Relevant affiliations

3.7. Corresponding Author's Countries

The analysis of Corresponding Author's Countries (Figure 7) offers a more comprehensive view into the regional distribution of the top contributors to Arabic language research, as it takes into account the location of the corresponding authors. Two crucial parameters are taken into consideration; Single Country Publications (SCP) and Multiple Country Publications (MCP) along with their corresponding frequencies and MCP ratios to convey the extend and nature of the research taking place in these countries.

SCP, or research articles published within one country, indicate the internal research efforts of that country. MCP, on the other hand, present the articles that arise from international collaborations, demonstrating the degree to which researchers in a given country work with their colleagues from across the globe. These deliver valuable insights into the global research landscape of Arabic language studies, showcasing both local research capabilities and international research networks.

The first thing to note is the sheer volume of articles being produced by the countries with a strong interest or cultural investment in the Arabic language. The country with the highest number of articles, denoted here without a specific name, leads with 206 articles, of which 167 are SCP and 39 are MCP, showing a relatively balanced production of both national and international research.

Following closely behind are Saudi Arabia with 28 articles (24 SCP and 4 MCP) and Malaysia with 27 articles (26 SCP and 1 MCP), also exhibiting both healthy domestic and active international research outputs. Next in line are Jordan and Tunisia, with Jordan leaning more towards everyone's favorite type of research; 19 articles (13 SCP and 6 MCP) and hence a higher MCP ratio, indicating a wide network of international research partners in the Jordanian academic community.

Crawling alongside are Algeria who chart out 14 articles and the UAE who contribute 13, both with a more active position in the global research community - a higher MCP ratio - indicating important academic partnerships with foreign scholars in pushing the boundaries of research in Arabic. Rounding out said region's contributions to the field are Egypt, Indonesia,

and Iran with Egypt and Indonesia tipping the scales more towards SCP and hence a higher propensity to collaborate, whereas all 9 articles from Iran are SCP, meaning they may be focusing more on work within their country's borders. Again, the analysis of Corresponding Author's Countries and their SCP and MCP outputs (Figure 7) not only reveal the global distribution and collaborative nature of Arabic language research, but also showcases the myriad different degrees to which different countries are engaged in this type of research. This variance in research collaboration patterns hint at the universal appeal of Arabic language studies and the equal importance of local and international collaborations in moving this field forward.

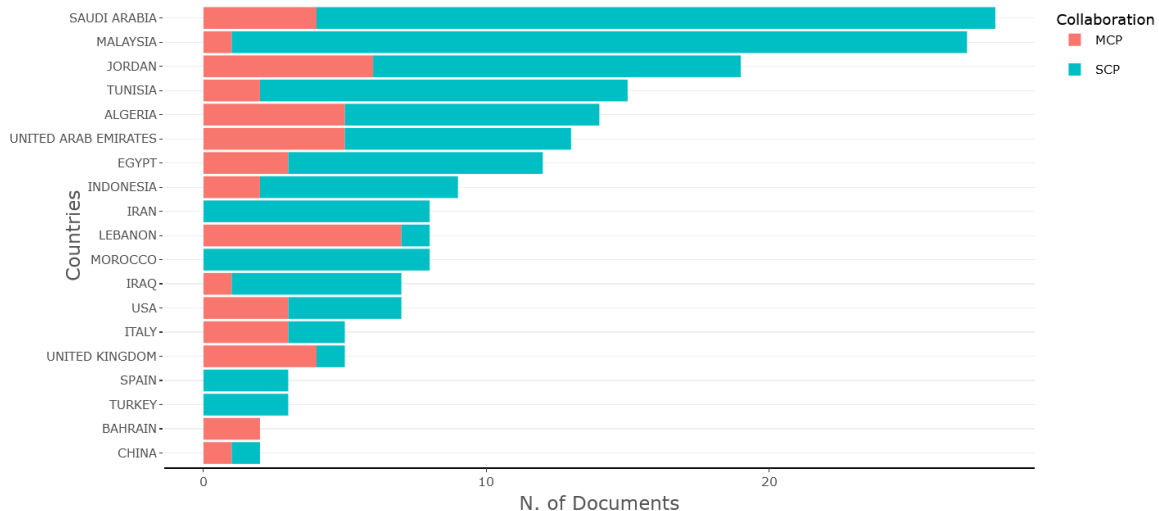


Fig.7. Corresponding Author's Countries

3.8. Most cited countries

The bibliometric analysis of the most cited countries with regards to Arabic language research is presented in Figure 8. This not only highlights the countries whose research outputs have accrued the most citations, but also does so by comparing these metrics via average article citations, thus providing a more meaningful way of ascertaining the qualitative impact of research outputs. Saudi Arabia is ranked top with a total citation count (TC) of 196, which translates to an average of 7.00 citations per article. This metric demonstrates Saudi Arabia's marked influence and recognition of its outputs within the field in question, which illustrates Saudi Arabia's position in terms of advancing scholarly knowledge and being commensurately recognized by scholars at large.

Singapore accrues a remarkable 122 TC, but perhaps more strikingly still, delivers an exceptional average of 122.00 citations per article, which alludes to a singularly impactful output, the like of which, has resonated on an uncommon scale within the field. This extraordinary average would suggest that research originating from Singapore is perhaps less voluminous, however it is of considerable relevance and significance, not only to the discipline, but moreover, the wider global academic discourse surrounding Arabic language studies. The United Kingdom also wields considerable influence, with 91 TC and an impressive average of 18.20 citations per article. This is indicative of strong academic outputs from the UK, which are collectively being internationally recognised, which in turn, furnishes high quality and influential research output to the field of Arabic language studies and thus enhances the existing canon.

The list of most cited countries, below, reveals an interesting global spread to Arabic language studies. Egypt tops the list with a total citation count of 647 and an average per article of 4.50, underlining that it continues to engage internationally with and make significant contributions to the field. Saudi Arabia is second on the list, with a total citations count of 223 and an average per article of 5.20, confirming its sustained influence. While fewer in number, research from these countries tends to be cited at a higher rate, signifying their ongoing impact on the field. Jordan and Tunisia follow with total citations of 80 and 78 and averages per article of 4.20 and 5.20, respectively, demonstrating consistent engagement with and contribution to Arabic language studies on behalf of their research outputs. Spain and the USA each present interesting cases, with high average citations per article of 23.30 and 9.70 respectively, indicating that while they may output fewer research papers, their contributions are cited higher and valued by the scholarly community at a greater rate. This level of citation suggests that their research introduces significant advances, or critical insights, into Arabic language studies.

Malaysia, Lebanon, and Morocco follow with total citations of 65, 50, and 45, and average citations per article on the lower end of the spectrum compared to the leaders but nonetheless signaling meaningful impact and contribution. In the case of Malaysia's researched, with an average of 2.40 citations per article, for example, while it's fewer research papers might not bring as high a citation count, the fact it's research is included in this list would suggest that it has a broad and active

engagement in the field, even compared to other countries. This form of international participation on behalf of these most cited countries for Arabic language research is a testament to not just the quantity of research output from these countries, but the quality and influence of it as explained through the raw and average citation counts below. This diverse international participation creates a well-rounded and engaged scholarly conversation on Arabic language studies, which, in turn, is propelled forward by these interlinking contributions.

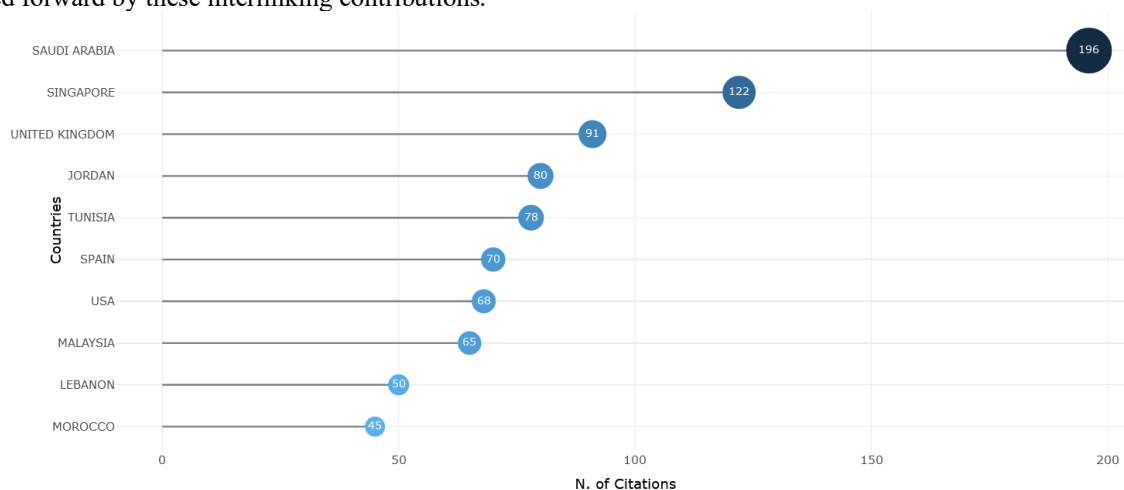


Fig.8. Most cited countries

3.9. Word Cloud

The word cloud as a depiction of the word frequency analysis of the Arabic language research (shown in Figure 9) gives an overview of the main themes and foci that are prevalent in the field. These data show the most centrally occurring terms and correspondingly, where the main research directions, methodologies, and topics which dominate the Arabic language studies. At the fore is the term "arabic languages" which appears 204 times, indicating the main focus of the research corpus and highlighting the general interest in the Arabic language, including the different dialects and languages as well as the study of various linguistic properties of the language. Secondly, "natural language processing systems" with 66 mentions, "computational linguistics" (45 times) and "language" also with 45 occurrences — shows the high emphasis on language oriented research towards technological perspectives, using the full range of available computational tools and theories, and by virtue of alphabetically-ordered names, the commonplace intersection between linguistics and computer science.

The roughly equal appearance of "female" terms and "male" terms, each 37 times with "human" at 37, surprisingly indicates that the authors of Arabic NLP research are either very evenly split across genders or else are interested in the phenomenon in their language of study across genders. Similarly, "arabic (language)" with 34, "article" and "sentiment analysis" also with 34, demonstrates a coherent focus on output within the corpus and is concerned with the study of Arabic-related sentiment, showing a high concern for the concomitant corpus and the clear interest in use of NLP technologies to address opinion mining and emotion detection problems within Arabic texts.

The appearances of "deep learning" for 26 and "natural language processing" with 25, with "sentiment analysis" show the high degree to which computational techniques are being brought to bear upon the language and accordingly, the advance of sophisticated AI and machine learning methods for the examination of Arabic linguistic data. Closer inspection reveals a trend in research methods and demographics, with terms like "adult," "questionnaire," "adolescent.," "young adult" indicating the varied age groups studied, and methodologies like "classification (of information), social media," "speech recognition" revealing the varied application of Arabic language research, from understanding social media content, to developing speech recognition systems.

The incidence of words related to reliability and validity, such as "reliability," "reproducibility" and "internal consistency," along with "cronbach alpha coefficient," suggests a concerted effort by the academic community to generate reliable and replicable research findings. Lastly, the term "saudi arabia" 10 times suggests either a geographical focus, or significant contribution to this field from the region, along with terms like "e-learning," "learning algorithms," pointing to an interest in applying linguistic research to e-learning technologies / methodologies. Figure 9, then, not only provides evidence of multiple themes and modes of inquiry in Arabic language research, but also of its inter-disciplinary nature, combining linguistics, computational and technological approaches, a high level of methodological rigor, and the eventual application of findings to diverse domains, including, education, technology, and the analysis of social media content.



Fig.9. Word cloud

3.10. Co-occurrence network analysis

Co-occurrence network analysis (see Figure 10) provides an intuitive representation of the interconnectedness of key terms within Arabic language research. The visualization highlights thematic clusters and the relative importance of each term according to various metrics such as Betweenness, Closeness, and PageRank, and illustrates its affiliation with other concepts. This “map” of research focus areas shows how different concepts and techniques relate to one another and shows which themes are focal points within the network.

The “overall” network analysis of the top 100 total linkages between terms that appeared in one-another’s 20 most similar term lists (via Cosine sim). This division it into two main clusters, which correspond to two main thematic areas within Arabic language research. Cluster 1 consists of terms that focus on computational and technological aspects of Arabic language studies (e.g. “arabic languages”, “natural language processing systems”, “computational linguistics”, “sentiment analysis”). Of these, the term “arabic languages” has the highest Betweenness and PageRank scores, indicating that it acts as an important connector throughout the network, and thus is a major focal point in terms of research. This centrality highlights foundational interest in the language in that respects both its diverse linguistic aspects and its significant role in connecting various themes of research.

Cluster 2 is made up of terms that seem to focus much more on the human (participatory) and methodological aspects of research (e.g. “language”, “female”, “human”, “male”, “article”). The term “human” has a high Betweenness score, suggesting that it acts as an important way-term that bridges between many themes in that cluster, highlighting above all the interdisciplinary nature of Arabic language research, in which humanistic, social, and computational science themes are inseparably important. Of these, the terms “arabic (language)” and “article” have medium-Betweenness and higher PageRank scores, indicating that they act as important ways in which research findings are disseminated, and that many disparate strains of research within this branch of the field follow these themes.

More interestingly, the presence of terms such as “female,” “male,” and “researcher,” with notable PageRank scores, particularly in Cluster 2, may point to a gender-focused dimension within Arabic language studies. This could reflect, for instance, research into gendered language use or the representation of gender in Arabic text corpora. In the same cluster, the term “questionnaire” emerges, underscoring the methodological approaches employed in the field; these approaches emphasize the role of surveys in gathering linguistic behaviors and attitudes.

The metrics used in this analysis—Betweenness, Closeness, and PageRank—provide a further window into the relative importance and connectivity of each term within the network: Betweenness flags terms that serve as bridges between different research areas, Closeness measures how close a term is to other terms in the network, and therefore how accessible and relevant a term is overall, to use the terminology of Easley and Kleinberg (2010), and PageRank gauges the likelihood of ending up at a term by following random jumps in the network, which tells us about the overall significance of a term.

This co-occurrence network analysis provides a richly detailed account of the thematic and methodological landscape of research on Arabic language. The identification of key terms and their interrelations supply concrete data on the important axes along which scholarly inquiry is directed, the often interdisciplinary nature of this work, and the crucial nodes that connect various areas of study, helped to map the current state of Arabic language studies and to imagine some future directions for the field, including the growing importance of methods that are computational in nature, the gender dimensions of research, and the humanistic dimensions of linguistic scholarship.

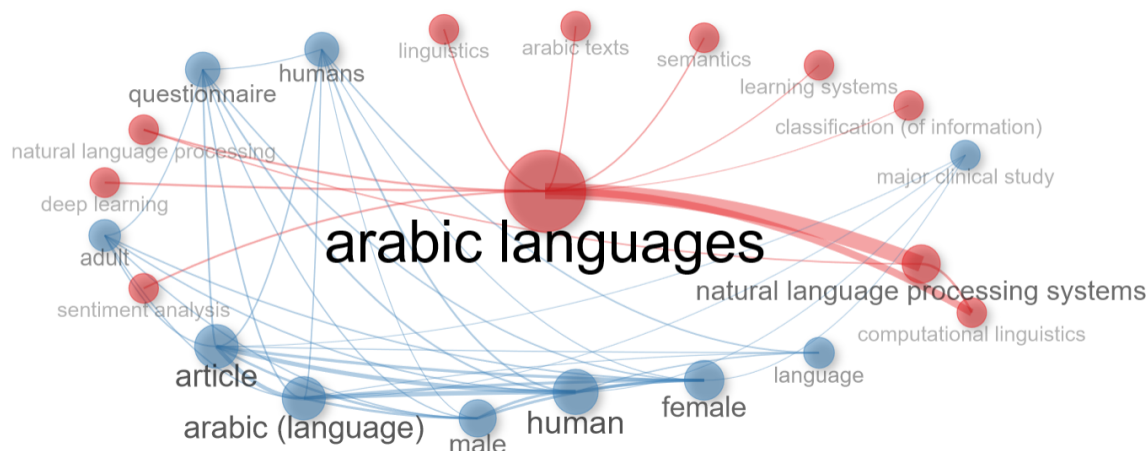


Fig.10. Co-occurrence network

4. DISCUSSION

As expected, the bibliometric analysis of Arabic language research uncovers a multidimensional landscape shaped by the convergence of thematic pursuits, geographical contributions, and methodological evolutions. Building on the previous findings, this reviewer triangulates these results to illuminate the dynamic interplay that characterizes the field's historical trajectory and contemporary landscape overtime.

The analysis has drawn attention to the central role of computational mechanisms in the production of Arabic language research. This trend is clearly manifest in the prevalence of "natural language processing systems," "computational linguistics," and "deep learning" among the top keywords. This emphasis is consistent with the significant proportion of articles that are attentive to the technological deployment of their findings, as shown in the most relevant sources section. A range of journals and conferences in computational science and technology - "Lecture Notes in Computer Science," "International Journal of Speech Technology," etc. - clearly occupy a vital role as a pathway for demarcating this field of inquiry. This intersection of linguistic discourse and computational production signals, at a minimum, a significant area of growth and potential topical impact for the field.

The contribution of geographies reveals the broader contextual bestiary that underscores the field's thematic development. When considering the corresponding author's countries, most cited countries, and total citizenship, the global multiplicity invested in the study of Arabic is made clear. Saudi Arabia, Malaysia, Jordan, and Tunisia emerge as significant contributors, suggesting that the thematic development of Arabic language research requires a geographical context. Saudi Arabia's position in both the quantity of articles produced and citations received also underscores its pivotal role as a global site for the advancement of Arabic linguistic knowledge. Finally, the reach across continents in terms of co-authorship demonstrates the rich diversity of the field's international contributors, enriching through a variety of partnerships that transcend geopolitical limitations locally. This distribution of contributions, and the respective citations they receive, is indicative of the universal relevance of Arabic language studies, consistently propagating an academic discourse that is able to transcend local boundaries.

The balanced occurrence of "female" and "male" terms, their central placement within the co-occurrence network, as well as the mention of emotion lexicon and questionnaire, suggest a deeper exploration of the constellation of sociolinguistic dimensions. These dimensions involve various processes that linguists have grappled with, such as gender representation and gender dynamics in language use, and are increasingly coming within the domain of engineering-oriented methodologies. The annual scientific production has apparently grown steadily over time, particularly during the last few years, leading to an expansion of research interests and methodologies that becomes evident in the word count analysis. This is paralleled by both the frequent mention of "deep learning" and "natural language processing," which captures the trend towards an increased application of advanced computational methods that has been noticed in the literature, and the above-noted role as pioneers that researchers on the "arabic languages" theme occupy. As a result, the field similarly follows disciplinary trends towards the digital turn.

Finally, the analysis of co-occurrence networks identifies "arabic languages" as a crucial node. It not only bridges several sub-disciplines, from sociolinguistic inquiries to computational linguistics, but illustratess the commonality of themes that promote this kind of interconnection that is sustained by the central divide among the various clusters along technological and (research) development dimensions, akin to the field's distinctive "dual" character. The bibliometric analysis of Arabic

language research paints a complex portrait of a field in ascendancy – one that is being transformed by computation, networked across the globe, and is fundamentally interdisciplinary. The synthesis of findings across dimensions of analysis reveals a field that is widening in its thematic and methodological range, and is deepening in its impact and relevance. The confluence of technological and humanistic inquiry, evidenced in the present study, suggests that Arabic language studies are a vibrant site for the co-evolution of inquiry around the linguistic structures of Arabic and the cultural phenomena that they contain.

5. CONCLUSION

The bibliometric analysis of the Arabic language research revealed several findings, pointing at the dynamic evolution and wide scope of the field. The emergence of computational methods, such as natural language processing and deep learning as evidenced by the rise of SVM classifiers among others is a clear marker of technological adaptation and adoption of these methods in the field. A geographical analysis showed that Arabic studies are a matter of global interest and among the leading countries in the field we find Saudi Arabia, Malaysia, Jordan and Tunisia, pointing to a worldwide relevance of these studies and signaling a high level of collaborative research. Sociolinguistic dimensions, including gender, are revealing of the field's continued concern with nuanced explorations of linguistic phenomena in their cultural context and the increase in scientific production over the years mirrors a general growth of academic interest in the field where the more traditional linguistic inquiries coexist and legend side-by-side with modern computational techniques. The co-occurrence network analysis further confirms the interdisciplinary of the research and in doing so it reveals how the foundational role of Arabic linguistic studies still serves as a bridge field to a variety of sub-disciplines. The analysis reveals a field that is undergoing a vibrant oscillation between a host of technologies and cultural studies, situated for a major contribution to understanding the multitudes of the Arabic language in the digital era.

Data source

The data supporting this bibliometric analysis were rigorously extracted from Scopus because of its thorough coverage and well-documented reliability for academic literature, particularly those pertinent to Arabic language research. The extracted dataset is available here on GitHub in a BibTeX file format to serve as a resource for transparency and further research by the scholarly community. Not only does this action reflect our strict adherence to open science principles, it also serves to allow others to reproduce and verify the integrity of our work. As our dataset is available for access via this link , which is a clear and accessible foundation for both verification and further investigation among other academics.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper

Acknowledgments

The authors would like to extend their heartfelt thanks to institution, for the moral support provided during the course of this research. The encouragement and guidance provided by the institution have helped tremendously in completing this research.

References

- [1] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *Journal of Business Research*, Article vol. 133, pp. 285-296, 2021. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- [2] A. Mahmoud and M. Zrigui, "Similar meaning analysis for original documents identification in arabic language," (in English), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11683 LNAI, pp. 193-206, 2019. https://doi.org/10.1007/978-3-030-28377-3_16
- [3] N. Farhani, N. Terbeh, and M. Zrigui, "Recognition of objects for the description of images in Arabic language with convolutional neural network," 2019, pp. 8947-8954: International Business Information Management Association, [IBIMA](https://doi.org/10.1007/978-3-030-28377-3_16).
- [4] A. Mahmoud and M. Zrigui, "Sentence Embedding and Convolutional Neural Network for Semantic Textual Similarity Detection in Arabic Language," (in English), *Arabian Journal for Science and Engineering*, vol. 44, no. 11, pp. 9263-9274, 2019. <https://doi.org/10.1007/s13369-019-04039-7>

- [5] N. Terbeh and M. Zrigui, "Identification of Pronunciation Defects in Spoken Arabic Language," (in English), *Communications in Computer and Information Science*, vol. 781, pp. 355-365, 2018. https://doi.org/10.1007/978-981-10-8438-6_28
- [6] M. A. Ben Mohamed, S. Zrigui, A. Zouaghi, and M. Zrigui, "N-scheme model: An approach towards reducing Arabic language sparseness," 2016: Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICTA.2015.7426895>
- [7] M. A. Ben Mohamed, S. Mallat, M. A. Nahdi, and M. Zrigui, "Exploring the potential of schemes in building NLP tools for Arabic language," (in English), *International Arab Journal of Information Technology*, vol. 12, no. 6, pp. 566-573, 2015. [Research Gate](https://www.researchgate.net/publication/318111111)
- [8] A. Boudhief, M. Maraoui, and M. Zrigui, "Elaboration of a model for an indexed base for teaching Arabic language to disabled people," 2014, pp. 110-116: IEEE Computer Society. <https://doi.org/10.1109/CSIT.2014.6805987>
- [9] A. Zouaghi, L. Merhbene, and M. Zrigui, "Word sense disambiguation for arabic language using the variants of the lesk algorithm," 2011, vol. 2, pp. 561-567. [Research Gate](https://www.researchgate.net/publication/318111111)
- [10] M. Mars, G. Antoniadis, and M. Zrigui, "Statistical Part Of Speech tagger for Arabic language," 2010, vol. 2, pp. 894-899. [Research Gate](https://www.researchgate.net/publication/318111111)
- [11] M. Mars, G. Antoniadis, and M. Zrigui, "A morphological analyser for Arabic language," 2007, pp. 297-298. [Research Gate](https://www.researchgate.net/publication/318111111)
- [12] R. Laatar, C. Aloulou, and L. Hadrach Belguith, "Towards a historical dictionary for Arabic language," (in English), *International Journal of Speech Technology*, vol. 25, no. 1, pp. 29-41, 2022. <https://doi.org/10.1007/s10772-020-09704-z>
- [13] Z. Ismail, N. Rasit, and T. Supriyatno, "Relationship Between Oral Language Anxiety and Students' Arabic Language Learning Outcomes in Malaysian Secondary Schools," (in English), *International Journal of Language Education*, vol. 7, no. 1, pp. 143-161, 2023. <https://doi.org/10.26858/ijole.v1i1.37368>
- [14] M. T. A. Ghani and W. A. A. W. Daud, "The Impact of Digital Game-Based Learning Towards Arabic Language Communication," (in English), *Jurnal Komunikasi: Malaysian Journal of Communication*, vol. 39, no. 1, pp. 407-424, 2023. <https://doi.org/10.17576/JKMJC-2023-3901-23>
- [15] F. Fekih-Romdhane et al., "Translation and validation of the mindful eating behaviour scale in the Arabic language," (in English), *BMC Psychiatry*, vol. 23, no. 1, 2023. <https://doi.org/10.1186/s12888-023-04614-1>
- [16] M. Hajjar, A. E. S. Al Hajjar, K. Zreik, and P. Gallinari, "An improved structured and progressive electronic dictionary for the Arabic language: iSPEDAL," 2010, pp. 489-495. <https://doi.org/10.1109/ICIW.2010.80>
- [17] D. Malaeb et al., "Validation of the pathological narcissistic inventory (PNI) and its brief form (B-PNI) in the Arabic language," (in English), *BMC Psychiatry*, vol. 23, no. 1, 2023. <https://doi.org/10.1186/s12888-023-04644-9>