

Three Decades of Malay Linguistics Research: A Scientometric Analysis of Trends, Impact, and Emerging Directions

Lalu Nurul Yaqin ^a

lahu.yaqin@ubd.edu.bn

Malay Language, Linguistics, and Literature Program
Faculty of Arts and Social Sciences
Universiti Brunei Darussalam, Brunei Darussalam

Yabit Alas

yabit.alas@ubd.edu.bn

Malay Language, Linguistics, and Literature Program
Faculty of Arts and Social Sciences
Universiti Brunei Darussalam, Brunei Darussalam

Salinah Ja'afar

b1salina@um.edu.my

Malay Linguistics Program, Academy of Malay Studies,
Universiti Malaya, Malaysia

ABSTRACT

The study presents a scientometric analysis of the development of Malay linguistics research over the past three decades. Due to rapid technological changes in linguistics and policy needs, this study examines publication trends, institution and country contributions, themes, and co-citation cluster networks to provide an overview of the field's intellectual structure. The study utilises CiteSpace version 6.3. R3 Advance to analyse records from the Scopus database (1994-2024). The findings indicate substantial growth in publication output, with Malaysia publishing the most papers, followed by Indonesia, Australia, Singapore, and Brunei Darussalam. Universiti Kebangsaan Malaysia, Universiti Sains Malaysia, Universiti Malaya, Universiti Putra Malaysia, Universiti Teknologi Mara, and International Islamic University of Malaysia are the top contributing institutions to research in Malay linguistics. The emerging themes indicate a shift in research focus from traditional linguistic studies toward computational linguistics, digital text processing, and AI-driven applications. The key research clusters were identified, spanning dialectology, historical linguistics, sociolinguistics, and computational linguistics. Among influential contributors, Asmah Haji Omar, with a citation count of 55 and burstness of 4.08, is the most influential author, significantly shaping the discipline. The GEMA Online Journal of Language Studies has the highest citation bursts (1.81), which are crucial in advancing research in computational linguistics, sociolinguistics, and language identity. This study provides critical insights into the evolution of Malay linguistics research, offering valuable evidence to guide researchers and education policymakers in formulating strategic policies for academic innovation and disciplinary advancement.

Keywords: CiteSpace; language digitalization; Malay Language; Malay linguistics; Scientometric Analysis

^a Main & corresponding author

INTRODUCTION

The study of Malay linguistics has undergone significant changes since its inception as a lexical description, with lexicography evolving into technology-enabled new language description and application. In the 19th century, van der Tuuk (1864) examined Malay dictionaries, laying the foundation for systematic Malay lexicography and language standardisation. This was followed by Schlegel's (1890) study of Chinese loanwords, which expanded the scope of inquiry to include external linguistic influences. Later, Hayashi's (1959) investigation into Javanese place names signalled a growing interest in the sociohistorical dynamics of the Malay world, while Teeuw (1959) focused more directly on tracing the historical development of the Malay language itself.

The latter half of the 20th century saw an increase in interdisciplinarity in the study of Malay linguistics. Landmark papers by Veiga et al. (1981) investigated cognitive factors using pictorial stimuli in Malay and Portuguese, and Tan (1981) was the first to apply information theory to linguistic entropy. At the same time, sociolinguistics insight started to take hold, as evidenced by Ota's (1984) examination of the use and attitudes toward Malay in Chinese-speaking communities. Furthermore, Adelaar (1989) and Collins (1989) have laid the fundamental foundations for understanding the Austronesian roots of Malay, emphasising its historical phonological trends and manuscript traditions. Technological integration was evident in the 1990s: Shariff et al. (1990) pioneered computer research on pronunciation recognition, demonstrating one of the early applications of artificial intelligence in Malay linguistics. This wide-ranging work reflects an evolving frontier at the intersection of philology, cognitive science, sociolinguistics, and applied technology.

As globalisation continues to grow and transform in digital spaces, researchers in linguistics are increasingly turning to technology, corpus linguistics, and writing in digital communication as ways to understand language and patterns of language use better. For instance, McLellan (2023) examines the impact of digital discourse on the current usage of Malay, where changes and patterns in code-switching attitudes also emerge. Pillai and Kamaruddin (2006) also investigated how social media and other digital tools affect language practices, particularly in an urban, multilingual context. These are some of the developments shaping modern Malay linguistics, which stands at the crossroads between conventional and new linguistic methods, and between improved technology.

This shift has led to novel frames for language, change, and conservation, enabling researchers to undertake new analyses of language within the curatorial scope of study. Malay linguistics has utility beyond the historical record of scholarship. Malay is the national language of Malaysia and Brunei Darussalam, and a key lingua franca for the region. It plays a central role in education, policy, and social cohesion (Deterding & Athirah, 2017). Studies of language change have increasingly been combined with a discussion of the effects of recent changes, such as digitisation, multilingualism, and artificial intelligence, on the structure of language use and language maintenance. The increase in the number of technological facilities (e.g., NLP [Natural Language Processing] and CL [Computational Linguistics]) in Malay research marked a shift in trend toward technologically oriented paradigms. However, we have limited insight into how research trends have evolved, which scholars and institutions publish the most, and how emerging themes are reflected in the Malay linguistic research.

Therefore, scientometric analysis plays a crucial role in bridging the knowledge gap regarding the evolution of research in the Malay linguistics. Thus, this study provides an organised and evidence-based review of the Scopus database from 1994 to 2024. The study employs

scientometric analysis to describe the intellectual trajectory and impact, including citation analysis, co-citation network mapping, and keyword clustering. While scientometric analysis has been extensively used in other linguistic studies to plot research trajectories and academic circles, less has been done to apply it in Malay linguistics. This study aims to address this gap by examining research patterns, impact, and thematic orientation in a structured manner. The study also aims to identify the most prolific authors, institutions, and countries contributing to the theme, focusing on emerging research themes based on keyword and citation analysis. It also highlights how international and institutional links have influenced the direction of research on Malay linguistics. A theme burst and citation burst analysis are also used to find the knowledge gap and suggest future research directions. The novelty of this study lies in its application of advanced scientometric analysis to Malay linguistics, a field where such methodologies have been relatively underutilised. Specifically, it seeks to answer the following research questions:

1. How have publication trends evolved, and what factors influence research output?
2. Which institutions and countries contribute the most, and how do regional collaborations impact knowledge production?
3. What are the dominant research themes and emerging topics revealed by keyword and cluster analysis?
4. How do co-citation networks identify the most influential authors, references, and journals and their role in research development?

By addressing these questions, this study sheds light on the intellectual contours of Malay linguistics and offers strategic insights to guide future research directions and inform academic policy development.

LITERATURE REVIEW

HISTORICAL DEVELOPMENT OF MALAY LINGUISTICS

Research into Malay linguistics has developed continuously from its early days of lexical, phonological, and syntactic description and is now characterised by the application of computer models. Pioneering works in Malay phonology and syntax have provided vantage points on its language system through the lens of its Austronesian origin (Kaland, 2019; Mahdi, 2024). The earlier work investigated phonetic features and sentence structures, which were significant in characterising the identifying properties of Malay among the Austronesian languages. This was further compounded by the breadth of dialectal differences and the effects of migration and trade on linguistic diversity (Adelaar, 2016).

External factors, especially inter-regional trade, colonialism, and cultural transmission, contributed to the formation of Malay linguistics. Trade networks facilitated language contact, resulting in the adoption of loans and grammatical elements from Arabic and Sanskrit, as well as enriching and modifying the syntactic structure of Malay (Joby, 2023). The colonial era introduced other linguistic systems, influenced by Western impact, which also modernised and standardised Malay (Omar, 1971). Reading classical Malay manuscripts has also demonstrated the historical variability of scripts and linguistic practices, as well as how early documents inform contemporary theorising about standardisation and language formation (Rahman et al., 2018). Dialect translation focuses on various regional forms with unique phonological and morphological features, as found

in the Tioman-Aur locality (Hamid et al., 2017). The transition from descriptive linguistics to technology-assisted techniques, including digital corpora, speech recognition, and artificial intelligence, has further shaped modern analyses of Malay linguistics (Aneja et al., 2020).

CONTEMPORARY RESEARCH AREAS OF MALAY LINGUISTICS

Phonology-oriented studies of Malay have focused on phonetic variation in dialects. Nasalisation substitution and other linguistic processes, such as those found in Malay dialects like Sarawak Malay, have been investigated, uncovering the complexity of the language's phonemic system (Alexander, 2018). These results indicate the complex interplay between diachronic phonetic change and ongoing sound change. The morphosyntactic studies are limited to complex verbs and the relationship between Malay and various other languages in a multilingual society. There have been studies on the agglutinative nature of Malay philology and the effects of language contact on grammatical patterns (Naziman & Jaafar, 2018). This study has contributed to a theoretical account of Malay syntax and, more broadly, to the phenomenon of language contact. Sociolinguists have examined code-switching, multilingualism, and language identity in Malay-dominant settings. These researchers have made significant contributions to understanding a speaker's negotiation of linguistic identity in various sociocultural settings (Gil, 2024; Ting et al., 2021). The intersection of politics, culture, and language in Malay society can also be a significant driver of Malay sociolinguistic developments. Recent research has demonstrated how policy implementations and the ever-changing patterns of human migration can influence language usage.

Corpus linguistics has made a significant contribution to the current state of Malay linguistics research, enabling large-scale linguistic investigations. Digital corpora have enabled pattern recognition and thematic exploration in spoken and written Malay (Don, 2010; Noah et al., 2015; Rahim et al., 2021). Incorporating NLP applications has also made significant contributions to computational research in Malay linguistics, while challenges remain in representing dialectal variety and agglutinative morphological structure (Sze et al., 2010). Diachronic corpus studies and linguistic text mining have also enabled deeper analyses of historical Malay vocabulary and syntax. Historical points of trade, colonisation, and amalgamation with one another have shaped the modern-day expression of Malay (Suryadi, 2015). Investigation of historical sources remains a primary means of tracing the evolutionary roots of the language.

SCIENTOMETRIC STUDIES IN MALAY LINGUISTICS

Recently, scientometric research has been applied to analyse similar problems in linguistics. These perspectives provide valuable insights into the evolution and advancement of linguistic research. The basic techniques in scientometrics include citation analysis, co-citation analysis, and keyword cluster analysis, which help us understand the intellectual structure of language research (Yaqin et al., 2025). Analysis of the citation network is necessary for identifying important researchers and seminal papers, whose high citation frequency and context can be recognised. Such an approach highlights the leading contributors and demonstrates scholarly impact in Malay linguistics (Yaqin et al., 2024). According to Chen (2020), co-citation analysis generates alternative ways of understanding by presenting thematic overlap and intellectual connections and relationships between more than one linguistic study, specifically by detecting changes in research topics over time.

Keyword clustering techniques have been utilised to identify research trends by clustering relevant keywords and concepts. This method enables researchers to track the evolution of research preferences in linguistics and to recognise new areas of investigation that reflect new techniques and emerging science in linguistics (Yaqin et al., 2024). However, applying scientometrics to studies of non-English languages introduces challenges due to limited database coverage and underrepresentation of high-quality publications, which can lead to assessment bias.

Based on scientometric results, this study is significant in identifying the impact of research in Malay Linguistics. The scientometric analysis is advantageous for systematically monitoring publication trends, identifying key researchers in a particular field, and visualising the interdisciplinary collaboration (Chen & Song, 2019). Linguistic trends have been identified by incorporating citation network analysis into machine learning, offering ground-breaking opportunities for more accurate impact measures and precision values. Analysing the created citation networks and machine learning techniques deployed for scientometric studies opens new possibilities for refining the citation impact estimates and detecting emerging topics.

RESEARCH METHODS AND TOOLS

SCIENTOMETRIC ANALYSIS AND CITESPACE

The scientometric analysis quantitatively analyses collaboration networks, academic influence, and research development. This method utilises bibliometric indicators, such as topic clusters, co-authorship relationships, and citation structures, to provide a comprehensive picture of the evolution of knowledge in a research field (Chen, 2012; Yaqin et al., 2025). The intellectual structure of a study field is mapped using CiteSpace, a widely used bibliometric visualisation software (Chen, 2006). An increasingly common bibliometric visualisation tool is the map of science, which represents the intellectual structure of a scientific field.

CiteSpace helps identify top articles and research trends using burst word finding, co-citation network, and structural variation analysis (Chen, 2017). Moreover, CiteSpace enables theme clustering, dynamic monitoring of the research track, and the detection of knowledge gaps, which help us better understand the evolution of research (Chen & Song, 2019). Because of its ability to reveal the growth pattern of scientific knowledge, CiteSpace has been widely used in systematic scientometric studies (Chen, 2017).

DATA RETRIEVAL AND QUERY FORMULATION

The data is retrieved from Scopus, a vast bibliographic database of academic journals. Scopus is a trusted database for bibliometric studies due to its comprehensive and well-indexed nature. Table 1 presents the data source and parameters of Malay linguistics research over the three decades discussed in this article.

TABLE 1. Data Source and Parameters

Database	: Scopus
Timeline	: January 1994 to December 2024
Search Query	: TITLE-ABS-KEY ("Malay language" OR "Bahasa Melayu" OR "Malay linguistic*" OR "Malay dialect*" OR "Malay phonology" OR "Malay morphology" OR "Malay syntax" OR "Malay semantics" OR "Malay pragmatics" OR "Malay sociolinguistics" OR "Malay corpus linguistics" OR "Malay historical linguistics" OR "Malay computational linguistics") AND PUBYEAR > 1993 AND PUBYEAR < 2025 AND (LIMIT-TO (SUBJAREA , "SOC") OR LIMIT-TO (SUBJAREA , "ARTS"))
Subject Area	: Social Sciences and Arts and Humanities
Document Type	: Article (545), Conference paper (72), Book chapter (55), Review (29), Book (6), Conference review (4), Erratum (3), Editorial (2), Note (1)
Sample records	: 710 Articles
Authors	: 1413
Journals	: 310
Institutions	: 877
Countries/Regions	: 53
Duplicated Records	: No One
Invalid Records	: No One

DATA ANALYSIS

CiteSpace version 6.3. R3 Advance was used to probe the Scopus database, identifying research clusters, citation bursts, and co-citation networks. Various bibliometric indicators were employed in order to enhance the interpretability and robustness of the analysis.

Co-citation analysis identifies foundational research in Malay linguistics by examining frequently co-cited references and uncovering significant intellectual milestones. Citation burst detection tracks sudden increases in citation activity, highlighting emerging research areas and shifts in scholarly priorities. Cluster analysis groups publications into thematic clusters, assessing changes in citation networks, and identifying dominant research themes. Temporal trends and evolutionary pathway analysis visualise topic development over time, tracing paradigm shifts and key academic contributions. Collaboration network mapping evaluates interdisciplinary and international cooperation by analysing co-authorship networks and institutional affiliations, revealing the influence of research partnerships in shaping Malay linguistics.

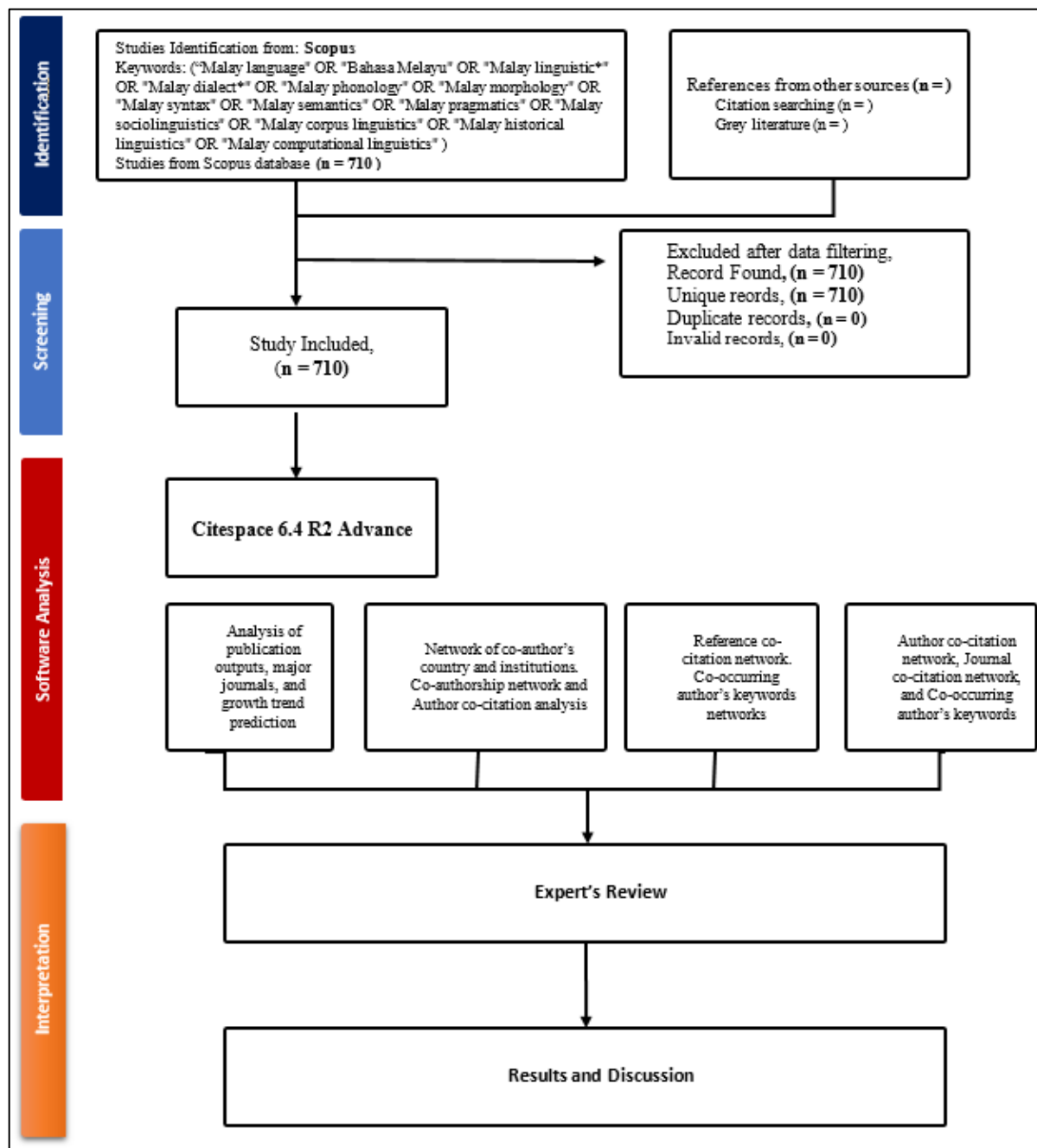


FIGURE 1. Scientometric Analysis Workflow Using CiteSpace

FINDINGS AND DISCUSSION

This section presents the main scientometric findings on publication trends, institutional and regional contributions, emerging patterns and themes, co-citation structures, cited references, and the journal co-citation network in Malay linguistics.

PUBLICATIONS TEMPORAL DISTRIBUTION

Figure 2 presents a bar chart of the analysis of the publishing pattern in Malay linguistics research from 1994 to 2024, which shows an upward trend in research papers over the last three decades. The number of new publications has ranged from one to six studies per year from 1994 to 2008, indicating its slow recognition worldwide. Between 2009 and 2010, there was a significant increase in the number of publications, reaching 23 per year, reflecting the growing interest among scholars.

The number of published items gradually increased between 2011 and 2016. It peaked at 37 items in 2017, primarily due to institutional backing, international cooperation, and advancements in digital linguistic tools, including improvements in computational processing and corpus-based studies. The following years witnessed a steady increase, with 50, 57, and 69 publications published in 2018, 2019, and 2020, respectively. While there was a minor drop in 2021 and 2023, with 51 and 47 publications, years of research remained strong, finishing with 62 publications that year. This continuous trend reflects the growing convergence of AI, NLP, and language-growing efforts, likely influencing future research. This trend signifies the broadening of Malaysian linguistics into new transdisciplinary directions, supported by technological advances and international collaboration.

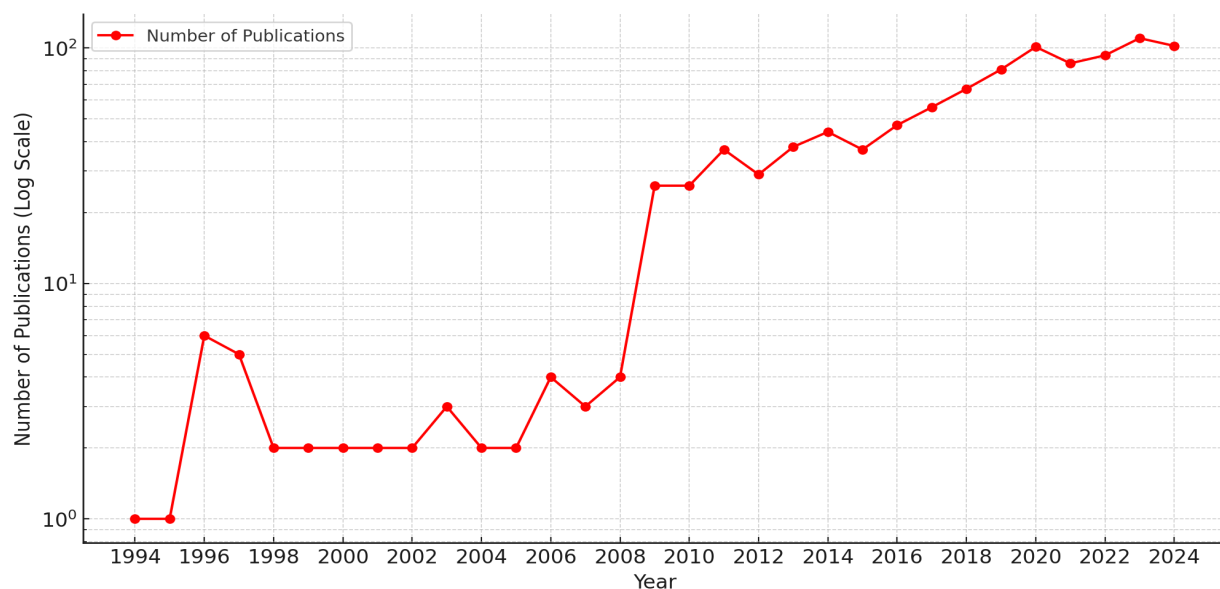


FIGURE 2. Annual publication trends in Malay linguistics (1994-2024).

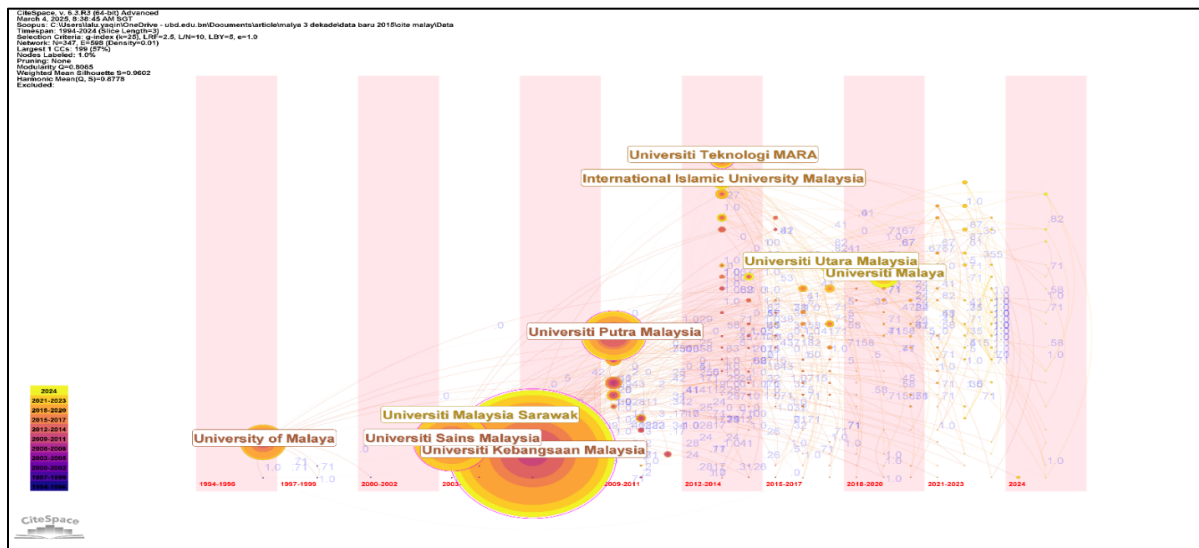


FIGURE 3. Timezone cluster of institutions productive in Malay linguistics (1994-2024).

INSTITUTIONAL AND REGIONAL CONTRIBUTIONS TO KNOWLEDGE PRODUCTION

Figure 3 illustrates the institution's productivity in Malay linguistics between 1994 and 2024. The network indicates Malaysia as the central hub, confirming the country's dominance in the sector. The largest of the three central universities and institutions is Universiti Kebangsaan Malaysia (UKM), which boasts 129 citations, clearly demonstrating its centrality in language studies. Some of the important works in this area include the work of Jalaluddin et al. (2008). This is followed by Universiti Sains Malaysia (USM), which has made significant contributions to computational linguistics and dialect translation, with 59 citations. Moreover, Universiti Putra Malaysia (UPM) has also contributed to linguistic approaches, dialect vitality, and language teaching, utilising 50 references. UM also makes significant contributions to lexicography and language standardisation projects. At the same time, Universiti Teknologi MARA (UiTM) and the International Islamic University Malaysia (IIUM) have been conducting research in corpus linguistics, collocation analysis, and bilingual education.

The co-citation analysis illustrates the research themes underpinning the paradigm of Malay linguistics. The Students' Achievement cluster (36 members) is the most prominent, focusing on educational linguistics, multilingualism, and proficiency. The Sustaining Coastal Terengganu Malay Dialect is the second-largest research theme, with 22 members. This topic addresses the conservation of dialects and language viability, which directly influences research works, for instance, Mansor (2018). Another topic that has been researched is the translation of Malay dialects, which examines interlingual effects and translation methods. Some influential works include Seong (2014). Linguistic analysis of Malay studies examines the exchanges of teacher training and didactic research languages, with special reference to curriculum innovation and teacher training methods. Lastly, lexicography and the standardisation of language influence dictionary expansion, computational lexicography, and etymology research. Arabic and Malay, focusing on loanwords, sentence structures, and digital evolution. From a research impact perspective, UKM, USM, and UPM have strong interdisciplinary connectivity between linguistics, education, and computation. UiTM and IIUM have significantly influenced lexicography, corpus linguistics, and bilingual studies.

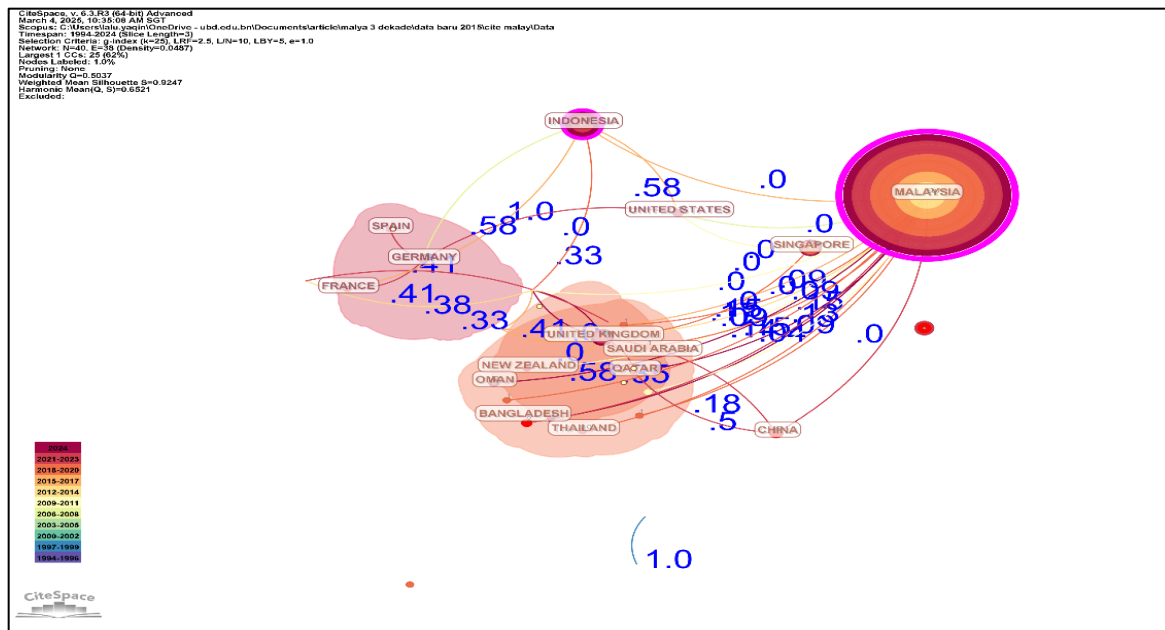


FIGURE 4: International Collaboration Network in Malay Linguistics Research (1994-2024)

Figure 4 illustrates global research collaboration. Network analysis reveals the subject's global contributions, highlighting key research areas, institutional affiliations, and international connections. Malaysia's rise as a research-led destination has brought to the fore its supreme importance in Malay linguistics. It further illustrates the intellectual history of Malay linguistics studies, highlighting the impact of citation and collaboration on various topics.

Four main research groups can be identified as describing the global network of collaboration. The largest cluster is that of Papuan Malay (11 members; silhouette value = 1.000). This cluster, chaired by Malaysia, with 507 citations, focuses on sentiment analysis, language structure, and multilingual corpora. Handayani et al. (2018) investigated sentiment analysis in Malay, and Sulaiman et al. (2019) focused on the role of Malay in detecting epidemic signals. The second cluster (cluster #1) on Papuan Malay has a size of 5 and a silhouette of 0.94. It examines the phonology and grammar of Malay and Indonesian. This field is closely related to Indonesia (41 citations), the USA (17 citations), and Germany (10 citations). The articles are Kaland & Gordon (2022) on Papuan Malay and American English phonetics, and Cole & Hermon (2008b) on Malay/Indonesian voice quality contrasts. Cluster #2 (silhouette: 0.935) is the influence of society on the history of Malay. The reuse of cultural and linguistic systems applicable to multilingual cultures has been studied in Singapore (25 citations), the United Kingdom (20 citations), and China (5 citations). Morve et al. (2023) is an important book that examines the impact of English on the sociocultural context of the Malay language and its implications for language policy. Cluster #3: Language Society (0.686) examines the impact of language on education policies and the sense of identity. The top citations are from Australia and Brunei Darussalam: 35 and 21 citations, respectively.

Malaysia is the principal hub of Malay language study, as indicated in Table 2, with 507 citations and the highest centrality score (0.58) in the research landscape (Table 2). Indonesia follows with 41 citations and a centrality score of 0.12, making a substantial contribution to lexical studies, syntax, and phonetics. Singapore (25 citations, 0.00 centrality), Brunei Darussalam (21 citations, 0.06 centrality), and Australia (35 citations, 0.13 centrality) demonstrate robust regional

involvement in sociolinguistic research. The United Kingdom (20 citations, 0.06 centrality) and the United States (17 citations, 0.05 centrality) are other significant contributors, with a focus on phonological research and multilingual corpora. The additional contributions from Germany, Thailand, and Japan underscore the increased worldwide participation in Malay linguistics studies.

TABLE 2. Top 10 Productive Countries in Malay Linguistics Research (1994-2024)

No	Country	Years	Frequency	Centrality
1	Malaysia	1995	507	0.58
2	Indonesia	1997	41	0.12
3	Australia	1997	35	0.13
4	Singapore	1996	25	0.00
5	Brunei Darussalam	1996	21	0.06
6	United Kingdom	1996	20	0.06
7	United States	2006	17	0.05
8	Germany	2008	10	0.06
9	Thailand	2011	10	0.00
10	Japan	1998	7	0.00

The partnership network highlights strong research ties between Malaysia and its neighbouring countries, particularly Indonesia, Singapore, and Brunei, reflecting shared linguistic heritage and policy-driven collaboration. Beyond the region, Australia and the United Kingdom also serve as important partners, contributing through interdisciplinary work that links linguistics with cultural and educational research.

EMERGING THEMES AND TOPICS IN MALAY LINGUISTICS RESEARCH

Analysing frequently occurring keywords in Malay linguistics (1994-2024) reveals evolving research priorities. Cluster analysis yields several distinct clusters, as illustrated in Figure 4. The study cluster Malay Version and Linguistic Adaptation (Cluster #0) is the largest cluster, with 42 papers and a high silhouette value of 0.9, indicating strong thematic coherence. This cluster focuses primarily on adapting standardised linguistic tools for studying Malay and conducting psychometric analyses and linguistic validation of the tools. Moreover, the focus on linguistic standardisation and phonetic development in this domain has emphasised the relevance of formal language models.

The Large Vocabulary Continuous Speech Recognition (Cluster #1), the second largest with 30 papers (silhouette value: 0.975), explores works on automatic speech recognition (ASR) and natural language processing (NLP). Research in this area has had a significant impact on phonetics, phonology, and machine translation. Sze et al. (2010) are among the most relevant works for Malay speech modelling and transfer learning for Instant Malaysia English NLP models. MT, call, and text automation are important for LL studies in the field. Another central research theme, Jawi Revivalist Discourse and Historical Linguistics (Cluster #2), be represented by 27 studies (silhouette value = 0.965). This cluster examines the evolution of Jawi and its impact on the sociopolitical realm, as well as its revival in contemporary discourse. Motoki (2021) investigates the relationship between Islamization and Jawi script revivalism, contributing to the revival of interest in old writing systems, specifically the Lingua Franca and Multilingualism in Southeast Asia (cluster #3). The last section, comprising 23 documents (silhouette value: 0.845), focuses on the function of Malay as a lingua franca. The studies in this bundle focus on the role

of general-purpose digital multilingualism, digital multilingual communication, multilingual identity construction, and language choice.

Malay NLP and Text Applications (Cluster #4) focuses on the computational processing of text, including subjective, ironic, and emblematic language, cyberbullying, and misinformation detection, with the help of twenty studies (silhouette value: 0.893)—the improvement in lemmatisation and morphological analysis in related research in the area of remarkable computational linguistics frameworks. The Automated Teller Machine (ATM) Natural Interaction (Cluster #5) comprises 14 studies (silhouette value: 1.0) and explores the field of human-computer linguistic ergonomics. The Malaysian Experience in Linguistic Education (Cluster #6) emphasises student proficiency in Malay and language pedagogy, as evidenced by the 12 studies (silhouette value: 0.88). Research on teacher training, gamification, and ICT integration suggests a modernised approach to Malay language learning. The standardisation and Phonetic Research (Cluster #7) examines language standardisation, phonetic analysis, and psychometric validation with 10 experiments (silhouette value: 0.978). The endeavour to build structured linguistic metrics is emphasised by research, such as Mukari et al. (2006), on auditory testing in Malay, focusing on phonetic consistency in translations. These findings demonstrate progress, although further study is required in historical linguistics and classical Malay texts.

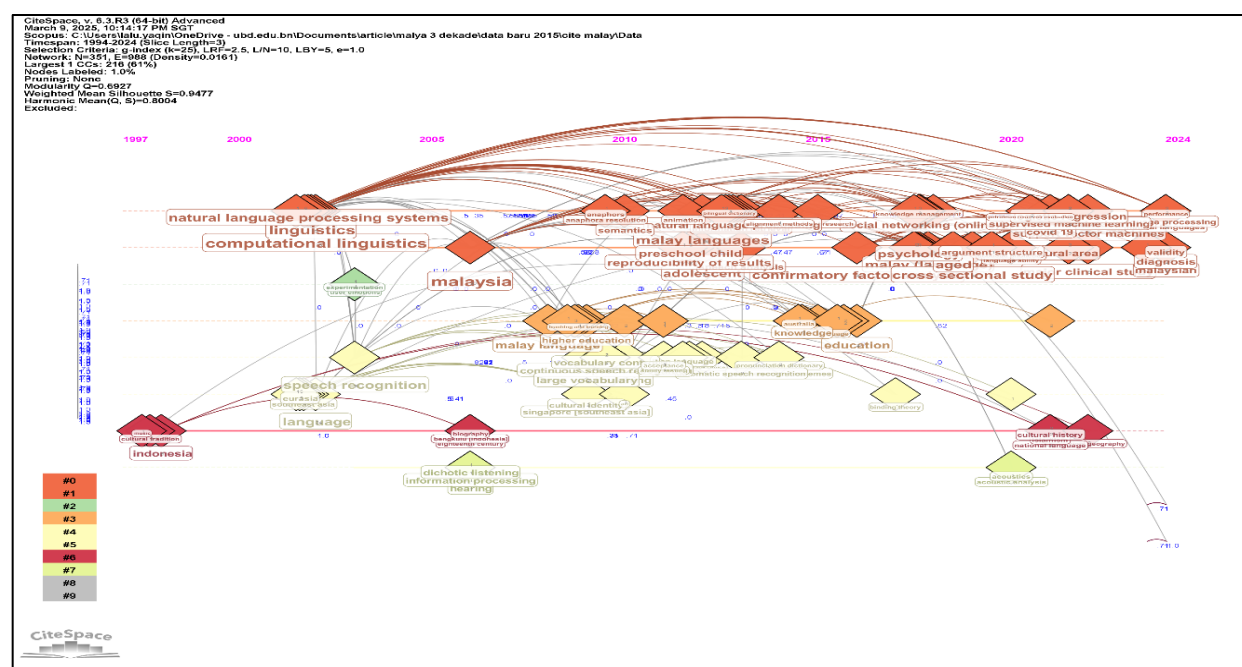


FIGURE 5. A timeline map of the Keyword Analysis and Emerging Research Themes in Malay Linguistics (1994-2024)

CO-CITATION ANALYSIS

Figure 6 visualises timeline maps of the evolving intellectual landscape of Malay linguistics (1994-2024) through co-citation network analysis. The research clusters highlight key themes, including dialectology, historical linguistics, sociolinguistics, and computational linguistics. These clusters underscore the contributions of prominent scholars in shaping knowledge within the field's diverse disciplines.

Theoretic Account (Cluster #0), the most prominent cluster, comprises 51 members with a silhouette value of 0.863. This cluster concentrates on optimality theory, Malay dialect phonology, and constraint-based language analysis. Omar AH (55 citations), Karim NS (44 citations), Collins JT (39 citations), and Hassan A are among the most highly cited researchers in this field (36 citations). Their research has substantially impacted Malay linguistic theory, particularly in Kelantan Malay dialects and existential sentence patterns in Malay. The Brunei Darussalam cluster (Cluster #1) comprises 51 members (silhouette value: 0.721) and focuses on language policy and multilingual education in Brunei. Fishman JA (23 citations), Asmah HO (15 citations), and Nothofer B (14 citations) are among the distinguished researchers who have made significant contributions to the field of Malay language vitality, classroom code-switching, and English-Malay policy interactions.

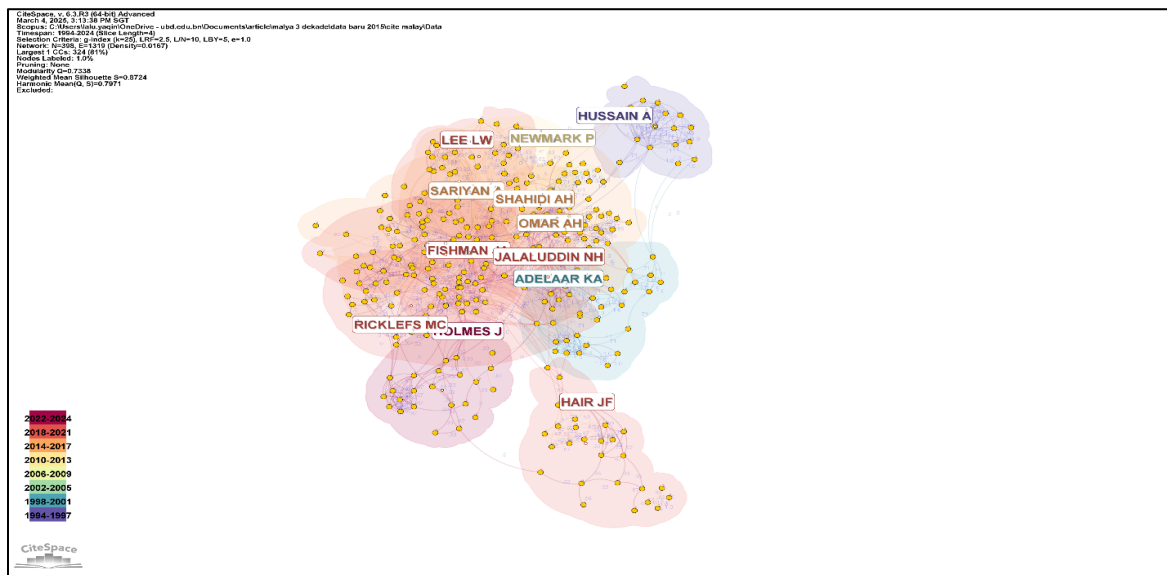


FIGURE 6. Co-Citation Network Analysis of Malay Linguistics Research (1994-2024)

The Malay World cluster (Cluster #2) comprises 31 individuals (silhouette value: 0.902) and prioritises Malay manuscript studies and historical linguistics. Ricklefs MC (14 citations), Anderson B (12 citations), and Chambert-Loir H are significant contributors (10 citations). This cluster investigates the socio-political aspects of language evolution, the influence of Islam on Malay linguistics, and historical Malay texts. Another key research group, Inventive Skill (Cluster #3), comprises 27 members (silhouette value: 0.978) and provides teacher training, educational evaluation, and psychometric analysis for Malay language instruction. Psychometric applications in linguistic study have been profoundly influenced by scholars such as Hair JF (15 citations), Fornell C (8 citations), and Mahamod Z (8 citations).

The Sociolinguistic Perspective cluster (Cluster #4), comprising 26 members (silhouette value: 0.97), investigates sociolinguistic shifts, language interaction phenomena, and code-switching in Malay-speaking groups. Adelaar KA (7 citations), Hussainmiya BA (6 citations), and Campbell L (5 citations) are among the scientists in this cluster who have been highly cited, underscoring the importance of sociolinguistic studies in multilingual cultures. Additional noteworthy clusters include Language Choice (Cluster #7, 24 members, silhouette value: 0.917), which investigates the impact of sociocultural factors and multilingualism on language use, and Inquisitive Semantics (Cluster #8, 22 members, silhouette value: 0.825), which explores lexical

CiteSpace v. 5.3.R3 (64-bit) Advanced
March 5, 2025, 9:22:25 AM SGT
Script: C:\Users\jaya\OneDrive - ubd.edu.bn\Documents\article\matya 3 dekade\data baru 2015\cite matya\data
Timespan: 1994-2024 (Slice Length=4)
Selection Criteria (g, q5, cov): 5, 5, 20; 4, 3, 20; LRF=2.5, UR=10, LB=5, e=1.0
Network: N=27, E=65 (Density=0.1852)
Largest CC: 26 (96%)
Nodes Labeled: 1.0%
Pruning: None
Modularity Q=0.9596
Weighted Mean Silhouette S=0.7349
Harmonic Mean(Q, S)=0.8252
Excluded:

2018-2021
2016-2017
2014-2017
2010-2013
2006-2009
2002-2005
1998-2001
1994-1997

CiteSpace

The 1994-2024 Malay linguistics co-citation network analysis indicates five key issue clusters in Figure 6, highlighting significant scholarly contributions and interactions. This cluster covers phonetics, dialectology, corpus linguistics, sociolinguistics, and language education. The network provides an overview of the most influential Malay linguists, their impact, and future research directions.

eISSN: 2550-2131
ISSN: 1675-8021

The second largest cluster, Theoretic Account of Malay Dialects (silhouette score: 0.745), encompasses theoretical aspects of dialectology, sociolinguistics, and linguistic structure. The work encompasses studies in corpus-based language analysis, phonological models, and language policy. Substantial insights by Kasdan et al. (2020) on partial reduplication in the Malay lexicon and elements of the Arabic loanwords in Malay dictionaries. Rahim (2014) investigated corpus observations of the Malay language, while Ahmad et al. (2011) examined vowel sequences and dialectal variation. The most cited studies in Malay dialectology, syntax, and linguistic structure are those by Omar AH (47 citations) and Jalaluddin (21 citations), respectively.

The third cluster, Language Events and Communication (silhouette score: 0.947), examines the sociolinguistic practices in digital/ media communication. Studies in this area also explore the usage of SMS and virtual language (Rahman et al., 2009). The latter includes Noah et al. (2018)'s study of news discourse in the Malay media and McKay and Rubdy's (2009) study of bilingual education. The fourth cluster is Unplanned Language Environment (silhouette score: 0.982), representing studies on bilingualism, code-switching, and language development. The third cluster, High-Level Thinking Skills and Mobile Learning (silhouette score: 0.909), focuses on the combination of multimedia, digital learning, literacy, and language learning.

CITED REFERENCES IN MALAY LINGUISTICS RESEARCH

Figure 8 visualises key references shaping Malay linguistics, highlighting 15 influential publications. The "Television Advertisement" cluster reflects a shift toward structural variation, lexical borrowing, and corpus-based analysis, signalling evolving research priorities in the field.

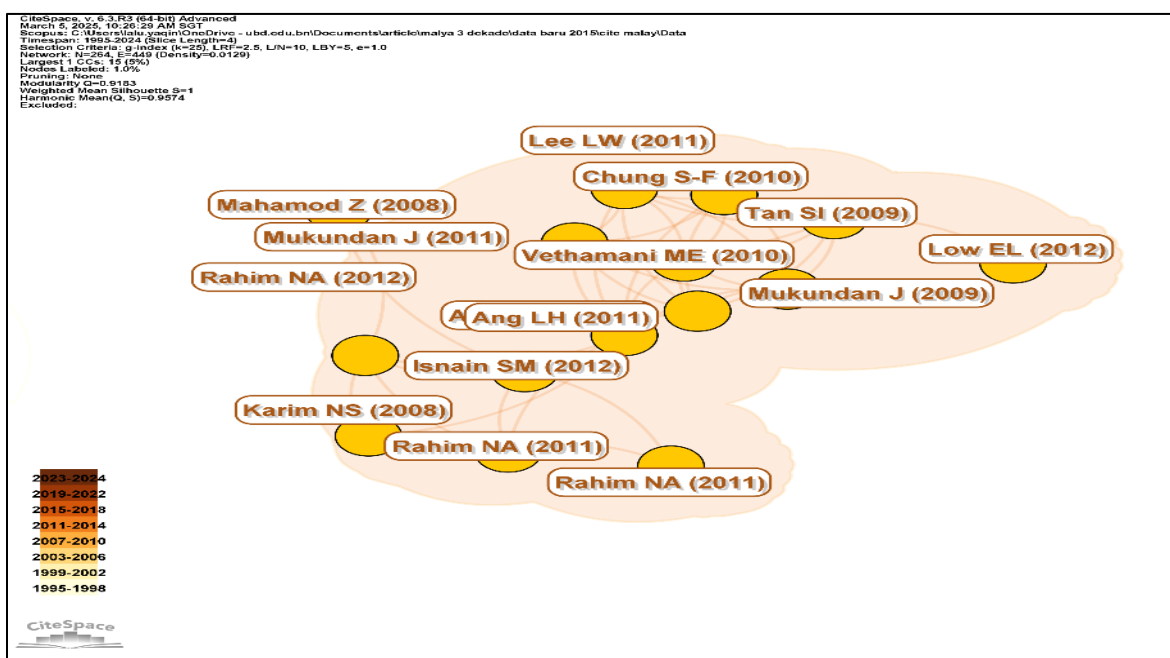


FIGURE 8. Cited references in Malay linguistics research (1994-2024)

The bibliometric research reveals substantial citation patterns, with Lee & Wheldall (2011) appearing as the most-cited reference, focusing on Malay word recognition skills among early readers. Similarly, the significance of cognitive linguistics and psycholinguistics in Malay language studies is further emphasised by the contributions of Masterson et al. (2010) to research on word-reading accuracy and vocabulary acquisition. The growing influence of digital discourse on the use of the Malay language is underscored by the junction of linguistics and media studies, as indicated by Abd Rahim & Harun (2011). This approach also emphasises transitioning from conventional grammatical research to computational and corpus-based methodologies. The emergence of NLP-driven methodologies underscores the increasing significance of digital humanities (Joharry & Rahim, 2014).

JOURNAL CO-CITATION NETWORK IN MALAY LINGUISTICS

Figure 9 visualises the journal co-citation network in Malay linguistics, highlighting key research themes, publication sources, and citation interactions. The analysis identifies ten major clusters with distinct interdisciplinary connections. The *GEMA Online Journal of Language Studies* is the most influential, exhibiting the highest citation burst (1.81) and centrality (0.16). It leads research in computational linguistics, sociolinguistics, and language identity. Similarly, the *Pertanika Journal of Social Sciences & Humanities* is essential for translation studies, sociolinguistics, and lexicography.

Meanwhile, the *Journal of Multilingual and Multicultural Development* plays a key role in linguistic diversity research and language policy discussions. The journal clusters demonstrate the multidisciplinary nature of Malay linguistics. Cluster #0 (*Melayu dalam Kalangan*) focuses on digital language campaigns and linguistic identity. Cluster #1 (*Bekenu Sarawak*) explores language preferences among ethnic minorities, particularly the Kedayan and Murut. Clusters #2 and #3 analyse Malay proverbs, Arabic loanwords, etymology, linguistic signage, and code-switching in Malaysia and Brunei.

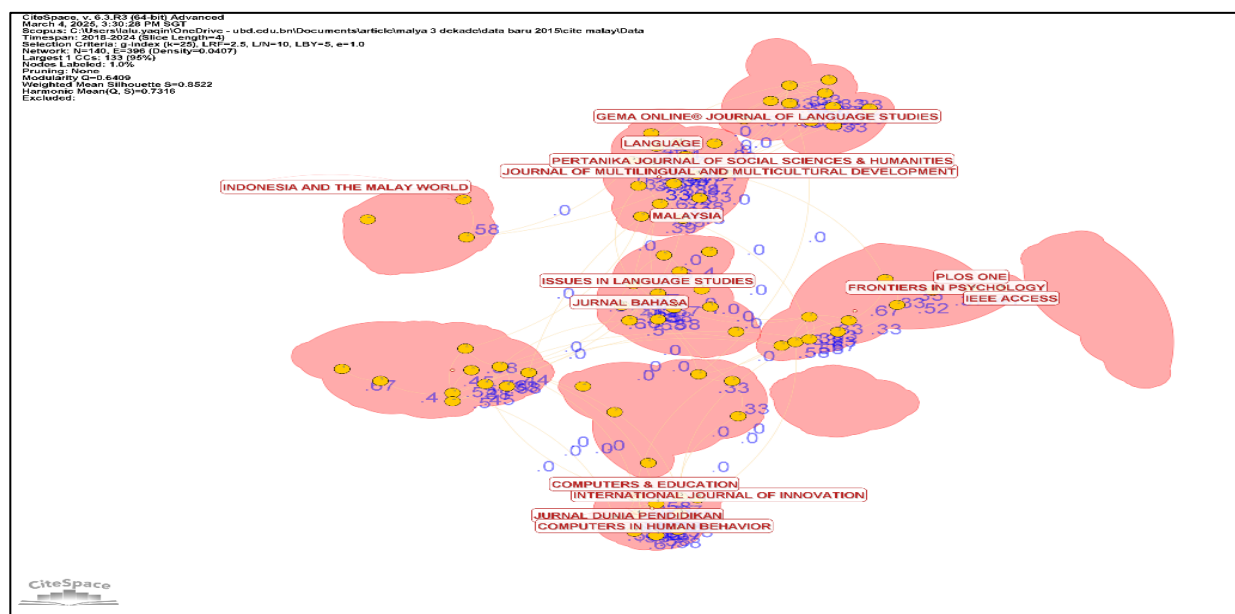


FIGURE 9. Journal co-citation network in Malay linguistics (1994-2024)

Frontiers in Psychology and PLOS ONE play a significant role in Cluster #5 (Language-Free Method), reflecting the influence of psycholinguistics. The growing impact of technology in linguistics is evident in Cluster #6 (Inventive Skill), where studies in the *International Journal of Innovation* and *Computer-Human Interaction* discuss AI-driven language learning and gamification tools. Historical linguistics and traditional Malay manuscripts remain active research areas, particularly in Cluster #7 (Malay Manuscript), with *Indonesia and the Malay World* as the most cited journals. Citation bursts indicate rising trends in computational linguistics, multilingualism, and language policy. *Indonesia and the Malay World* (1.50) emphasises historical linguistics, *The Journal of Multilingual and Multicultural Development* (1.28) explores linguistic variation, and *Pertanika Journal of Social Sciences & Humanities* (1.44) shapes identity research and cultural linguistics. The journal co-citation network reveals the intellectual structure of Malay linguistics, highlighting interdisciplinary integration across computational approaches, corpus linguistics, and psycholinguistics.

CONCLUSION, LIMITATION, AND FUTURE RESEARCH

The scientific study of Malay linguistics from 1994 to 2024 demonstrates significant growth and transformation. Over the past three decades, research has expanded beyond traditional phonological and syntactic studies to encompass the growing role of digital language processing and artificial intelligence. This study has limitations. Relying solely on Scopus may exclude relevant studies from other databases, thereby limiting the comprehensiveness of the dataset. While CiteSpace and scientometric analysis effectively track quantitative data, they lack qualitative depth. A mixed-methods approach integrating qualitative content analysis with bibliometric data could provide deeper insights into Malay linguistics research. Future research should focus on interdisciplinary collaboration to enhance the accuracy and robustness of computational tools. Integrating computational linguistics with historical studies could offer a more comprehensive understanding of the evolution of the Malay language. Strengthening partnerships between digital humanities, artificial intelligence, and linguistics is essential for further advancing the field. Expanding linguistic corpora and developing multilingual frameworks are critical in shaping the future of Malay linguistics research.

REFERENCES

- Ab Rahman, R., Musanif, A. M., Abd Rahim, N., & Hamzah, Z. A. Z. (2009). Language Events and Virtual Communication from the Aspect of Ethnographic Communication. *Ubiquitous Learning*, 1(4), 59.
- Abd Rahim, N., & Harun, N. I. (2011). Emotions among the primary school students: A Malay drama program (software). *UPENA Johor*, 10, 185–198.
- Adelaar, A. (2016). Austronesians in Madagascar: A Critical Assessment of the Works of Paul Ottino and Philippe Beaujard. *Palgrave Series in Indian Ocean World Studies*, 77 – 112.
- Adelaar, K. A. (1989). Malay influence on Malagasy: linguistic and culture-historical implications. *Oceanic Linguistics*, 28(1), 1–46.
- Alexander, O. (2018). Notes on Structural Distinctions in Malay Dialects. *Wacana Journal of the Humanities of Indonesia*, 19(2), 327. <https://doi.org/10.17510/wacana.v19i2.706>

- Aneja, S., Nur Afikah Bte Abdul Mazid, S., & Aneja, N. (2020). *Neural Machine Translation model for University Email Application*. 74–79. <https://doi.org/10.1145/3421515.3421522>
- Bakar, N. A., Zahid, I., & Wan Ali, W. Z. K. (2024). Arabic Loanwords in Kamus Dewan Perdana: Questions of Etymology, Word Class and Definition. *Jurnal Arbitrer*, 11(3), 314 – 333.
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377. <https://doi.org/10.1002/asi.20317>
- Chen, C. (2012). Predictive effects of structural variation on citation counts. *Journal of the American Society for Information Science and Technology*, 63(3), 431–449.
- Chen, C. (2017). Science Mapping: A Systematic Review of the Literature. *Journal of Data and Information Science*, 2(2), 1–40.
- Chen, C., & Song, M. (2019). Visualizing a field of research: A methodology of systematic scientometric reviews. *PLOS ONE*, 14(10), e0223994.
- Cole, P., & Hermon, G. (2008b). Voice in Malay/Indonesian. *Lingua*, 118(10), 1500–1553.
- Collins, J. T. (1989). Malay dialect research in Malaysia: The issue of perspective. *Bijdragen Tot de Taal-, Land-En Volkenkunde/Journal of the Humanities and Social Sciences of Southeast Asia*, 145(2), 235–264.
- Coluzzi, P. (2020). Jawi, an endangered orthography in the Malaysian linguistic landscape. *International Journal of Multilingualism*, 1–17.
- Deterding, D., & Athirah, I. (2017). Brunei Malay. *Journal of the International Phonetic Association*, 47(1), 99–108.
- Don, Z. M. (2010). Processing Natural Malay Texts: A Data-Driven Approach. *Trames Journal of the Humanities and Social Sciences*, 14(1), 90.
- Gil, D. (2024). Borrowing within Malayic; The role of exotericity. *Wacana*, 25(3).
- Hamid, N. A., Aman, R., & Shahidi, A. H. (2017). Tioman-aur-pemanggil Malay dialect variants: A comparative linguistics analysis. *GEMA Online Journal of Language Studies*, 17(2), 219–244. <https://doi.org/10.17576/gema-2017-1702-13>
- Handayani, D., Abu Bakar, N. S. A., Yaacob, H., & Abuzaraida, M. A. (2018). Sentiment Analysis for Malay Language: Systematic Literature Review. *Proceedings 2018 International Conference On Information And Communication Technology For The Muslim World (Ict4m)*, 305–310.
- Hayashi, H. (1959). Thoughts on Javanese place names. *Japanese Journal of Human Geography*, 11(2), 97–116.
- Jalaluddin, N. H., Awal, N. M., & Bakar, K. A. (2008). The mastery of English language among lower secondary school students in Malaysia: A linguistic analysis. *European Journal of Social Sciences*, 7(2), 106 – 119.
- Joby, C. (2023). Missionary linguistics in the East Indies in the seventeenth century. *Historiographia Linguistica*, 50(1), 1 – 34. <https://doi.org/10.1075/hl.00128.job>
- Joharry, S. A., & Rahim, H. A. (2014). Corpus research in Malaysia: A bibliographic analysis. *Kajian Malaysia*, 32, 17–43.
- Kaland, C. (2019). Acoustic correlates of word stress in Papuan Malay. *Journal of Phonetics*, 74, 55–74. <https://doi.org/10.1016/j.wocn.2019.02.003>
- Kaland, C., & Gordon, M. K. (2022). The role of f0 shape and phrasal position in Papuan Malay and American English word identification. *PHONETICA*, 79(3), 219–245.

- Kasdan, J., Baharuddin, R., & Shamsuri, A. S. (2020). Covid-19 in the Corpus Of Malay Terminology: A Socio-terminological Analysis. *Gema Online Journal Of Language Studies*, 20(3), 221–241. <https://doi.org/10.17576/gema-2020-2003-13>
- Lee, L. W., & Wheldall, K. (2011). Acquisition of Malay word recognition skills: lessons from low-progress early readers. *Dyslexia*, 17(1), 19–37.
- Mahdi, W. (2024). Austroasiatic loanwords in Austronesian languages. *Wacana*, 25(3). <https://doi.org/10.17510/wacana.v25i3.1780>
- Mansor, N. S. (2018). Addressing family members in the Malay culture: Practices of “young-adult” university students. *Lenguas Modernas*, 2018(51), 63 – 76.
- Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children’s printed word database: Continuities and changes over time in children’s early reading vocabulary. *British Journal of Psychology*, 101(2), 221–242.
- McKay, S. L., & Rubdy, R. (2009). The social and sociolinguistic contexts of language learning and teaching. *The Handbook of Language Teaching*, 7–25.
- McLellan, J. (2023). Digital Technologies, Social Media, Global and Local Languages in Southeast Asia. *Kemanusiaan*, 30, 9–24. <https://doi.org/10.21315/kajh2023.30.s1.2>
- Mohd Noah, S. A., Mohamad Ali, N., & Hasan, M. S. (2018). Determining features of news headline in Malay news document. *GEMA Online Journal of Language Studies*, 18(2), 154–167. <https://doi.org/10.17576/gema-2018-1802-11>
- Morve, R. K., Wen, X., & Mansour, N. (2023). The role of English and the sociocultural structure of Bahasa: a study of Brunei Darussalam. *Asian-Pacific Journal of Second and Foreign Language Education*, 8(1). <https://doi.org/10.1186/s40862-023-00186-5>
- Motoki, Y. (2021). Script of islam, script of malay: Jawi revivalist discourse in early independent indonesia and its relationship to Malaya. *Japanese Journal of Southeast Asian Studies*, 58(2), 141 – 163.
- Mukari, S. Z., Keith, R. W., Tharpe, A. M., & Johnson, C. D. (2006). Development and standardization of single and double dichotic digit tests in the Malay language. *International Journal Of Audiology*, 45(6), 344–352.
- Naziman, S. N. M., & Jaafar, S. R. S. (2018). The adaptation strategies of English loanwords among Kelantan Malay dialect speakers. *3L: Language, Linguistics, Literature*, 24(4), 128–142. <https://doi.org/10.17576/3L-2018-2404-10>
- Noah, S. A., Omar, N., & Amruddin, A. Y. (2015). Evaluation of Lexical-Based Approaches to the Semantic Similarity of Malay Sentences. *JOURNAL OF QUANTITATIVE LINGUISTICS*, 22(2), 135–156.
- Omar, A. H. (1971). Standard language and the standardization of Malay. *Anthropological Linguistics*, 75–89.
- Ota, I. (1985). Language environment for Chinese-speaking societies in Malaysia and Singapore. *Geographical Review of Japan*, A, 58, 318–339.
- Pillai, S., & Kamaruddin, F. (2006). The variety of Malaysian English used in radio advertisements. *Varieties of English in Southeast Asia and Beyond*, 39–54.
- Rahim, H. A. (2014). Corpora in language research in Malaysia. *Kajian Malaysia*, 32, 1–16.
- Rahim, H. A., Hasan, R. A., Hong, A. L., & Joharry, S. A. (2021). The Diachronic Malaysian English Corpus (DMEC): Design, Development and Challenges. *GEMA Online Journal of Language Studies*, 21(4), 88–109. <https://doi.org/10.17576/gema-2021-2104-05>

- Rahman, A. H. A., Abdullah, F. binti, Ali, A. M., Kadir, F. K. A., Adam, F., & Ismail, D. (2018). Jawi script and writings in the Malay world: Approaches and contributions to islamic discourse. *International Journal of Civil Engineering and Technology*, 9(4), 1269 – 1275.
- Ramli, I., Jamil, N., & Ardi, N. (2020). Formant characteristics of Malay vowels. *International Journal of Evaluation and Research in Education*, 9(1), 144 – 152.
- Schlegel, G. (1891). Chinese loanwords in the Malay language. *T'oung Pao*, 391-405.
- Seong, G. S. (2014). *Bahasa Cina-Bahasa Melayu: Kebolehterjemahan Budaya (Penerbit USM)*.
- Shariff, Z. A. M., Othman, M., & Maiden, M. I. A. (1990). Recognition of standard malaysian language pronunciation. In *Proc. ICSLP 1990* (pp. 277-280).
- Sulaiman, F. B., Yanti, N. K. S., Lesmanawati, D. A. S., Trent, M., Macintyre, C. R., & Chughtai, A. A. (2019). Language-specific gaps in identifying early epidemic signals – a case study of the Malay language. *Global Biosecurity*, 1. <https://doi.org/10.31646/gbio.33>
- Suryadi. (2015). A linguistic kaleidoscope of the Malay letter: The case of the eighteenth century official letters from the Sultanate of Buton. *Wacana*, 16(2), 304–338. <https://doi.org/10.17510/wacana.v16i2.380>
- Sze, H. K., Ping, T. T., Kong, T. E., & Yu-n, C. (2010). Malay Language Modeling In Large Vocabulary Continuous Speech Recognition With Linguistic Information. *2nd Workshop on Spoken Language Technologies for Under-Resourced Languages, SLTU 2010*, 56 – 61.
- Tan, C. (1981). On the entropy of the Malay language (Corresp.). *IEEE Transactions on Information Theory*, 27(3), 383-384.
- Teeuw, A. (1959). The history of the Malay language. *Bijdragen Tot de Taal-, Land-En Volkenkunde*, 2de Afl, 138–156.
- Tuuk, H. N. van der (1864). Art. VIII. On the existing Dictionaries of the Malay Language. *Journal of the Royal Asiatic Society*, 1(1), 181-186.
- Veiga, J. D. S., Seddon, G. M., & Jusoh, I. (1981). Testing the understanding of pictorial spatial relationships using the Portuguese and Malay languages. *Educational Psychology*, 1(4), 305-317.
- Yaqin, L. N., Bilad, M. R., Yusof, B., Hassan, H., Damit, A. R., & Chuchu, I. F. (2025). Mapping research evolution in higher education: a scientometric analysis of Brunei Darussalam (1986–2024). *Discover Sustainability*, 6(1). <https://doi.org/10.1007/s43621-025-00917-3>
- Yaqin, L. N., Yusof, B., & Damit, A. R. (2024). “ Status Quo” of Malay Linguistics Studies: A Bibliometric Analysis. *Journal of Language and Literature Studies*, 4(2), 296–312.

ABOUT THE AUTHORS

Lalu Nurul Yaqin, PhD., is an Assistant Professor of Malay Language, Linguistics and Literature at the Faculty of Arts and Social Sciences, Universiti Brunei Darussalam. His scholarly work spans cultural linguistics, comparative politeness studies, sociolinguistics, and pragmatics. He also specialises in bibliometric and scientometric analysis, as well as the application of artificial intelligence in language and linguistics.

Yabit Alas, PhD., is a Senior Assistant Professor of Malay Language, Linguistics, and Literature at the Faculty of Arts and Social Sciences. His expertise lies in comparative Austronesian linguistics, with a research focus on the Dusunic languages of North Borneo, their origins, dispersal, and classification. He also studies the Malay language in the context of globalisation.

Salinah Ja'afar, PhD, is an Associate Professor at the Academy of Malay Studies, University of Malaya, and is the Vice President of the Malaysian Translators Association. A former translator at Bank Negara Malaysia, she is also a lecturer, author of books, speaker, and academic administrator. Among her published works are *Dirgahayu Bahasa Istana Johor*, *Semantics*, *The Academic Biography of Nik Safiah Karim*, *Malay Pantun Discourse*, and several others.