P-ISSN 2338-0446
E-ISSN 2580-376X
GERAM
Gerakan Aktif Menulis
JURNAL PENDIDIKAN, BAHASA, DAN SASTRA
Volume 14, Issue 3, September 2026

EXPLORING SOUND CHANGE AND LANGUAGE KINSHIP: A LITERATURE STUDY FROM THE PERSPECTIVE OF HISTORICAL AND COMPARATIVE LINGUISTICS

Nadra[*1)], Reniwati[2)], Muhammad Yusdi[3)], Aldrizi Salsabila[4)]

[1]Universitas Andalas, Padang, Indonesia, nadra@hum.unand.ac.id
[2]Universitas Andalas, Padang, Indonesia, Indonesia, reniwati@hum.unand.ac.id
[3]Universitas Dharma Andalas, Padang, muhammad_yusdi@yahoo.co.id
[4]Universitas Andalas, Padang, Indonesia, aldrizi.unand@gmail.com

[*]*Correspondence to: nadra@hum.unand.ac.id*

Article History:	Received	Revision	Accepted	Published
	May 8, 2026	June 27, 2026	June 28, 2026	September 7, 2026

Abstract
Research on sound change, language kinship, and protolanguage reconstruction in Indonesia remains limited, given the number of existing languages. The purpose of this study is to explore the results of studies published over the last 10 years (2015–2025) and to identify the application of theory and methodology, as well as their consequences. The PRISMA method employs multiple pertinent keywords for its search. Articles were retrieved from Sinta and Publish or Perish (POP). After screening and eligibility, 24 of 245 publications were selected for comprehensive analysis. The results indicate that sound change occurs regularly and has a pattern. The analysis of language kinship, based on lexicostatistical and glottochronological methods, reveals various levels of language grouping, from the language-family level to the macrophylum, with an estimated separation time of thousands of years. Numerous issues were identified in comprehending the theory and methodology, including inaccuracies in calculating the percentage of cognates that resulted in genealogical misclassification, errors in identifying cognate words and sound correspondences, reliance on unsuitable data sources, flawed transcription, and concerns regarding the validity of the separation time, all of which significantly impacted the interpretation of the research findings. For enhanced accuracy, it is advisable to employ a combination of quantitative and qualitative approaches and to consider external factors, such as migration, language contact, and geography. This study is expected to enhance Indonesian historical and comparative linguistic studies and elucidate the evolution of language kinship.

Keywords
Historical and Comparative Linguistics, Language Kinship, Proto-Language, Reconstruction, Sound Change

DOI : https://doi.org/10.25299/geram.2026.28148
Citation : Nadra, Reniwati, Yusdi, M., Salsabila, A. (2026). Exploring Sound Change and Language Kinship: A Literature Study from The Perspective of Historical and Comparative Linguistics. GERAM (Gerakan Aktif Menulis), 14(3), 392–405 https://doi.org/10.25299/geram.2026.28148

INTRODUCTION

Historical and comparative linguistics is a branch of linguistics that studies language kinship and its historical development using linguistic evidence available today. Campbell (1998) stated that the comparative method is the most effective instrument for determining cross-language kinship, demonstrating that it is the most fundamental tool for scientifically identifying language kinship. By systematically retracing sound change, the researcher may reconstruct the proto-language and describe the evolutionary course of derived languages. Sound change is a central aspect of the comparative method because it is both relatively regular and predictable within the framework of phonological correspondence. Crowley & Bowern (2010) stated that the nature of sound change is regular, and the findings of consistent sound correspondence are the primary basis for the comparative reconstruction. Processes such as sound lenition/fortition, omission, addition, metathesis, fusion, assimilation, and dissimilation provide concrete evidence for mapping the phonological change from a proto-language to modern forms.

In Indonesia, research on language kinship, proto-language reconstruction, and sound change remains limited, given the number of languages it has, which is the second-most-spoken language in the world (Ethnologue, 2023; *Kementerian Pendidikan dan Kebudayaan* (Ministry of Education and Culture), 2019; Nadra, 2006; Purwo, 2003). The research already done by Rizqi & Widayati (2021), for instance, shows six sound changes from Proto-Austronesian (PAN) into Sumatran Javanese Dialects (SJD), which are metathesis (phonemic transposition), apheresis (initial sound omission), syncope (medial sound omission), apocope (final sound omission), prothesis (initial sound addition), and paragoge (final sound addition). Additionally, based on kinship, Sulistiyarini & Hendrokumoro (2023) also found that the Javanese, Sundanese, and Makassarese preserve different levels of kinship, even though the three languages originate from the same proto-language, with the following percentages: 43% (Javanese–Sundanese), 20% (Javanese–Makassarese), and 16% (Sundanese–Makassarese). Furthermore, Nadra et al. (2022) found variations in the reflex of the Proto-Malayic vowel phoneme in Malay, particularly in Merangin Regency, Jambi Province. The most significant variation occurred in the final (ultimate) syllable at Renah Alai, or observation point 1 (PO 1), due to language contact with Kerinci, Javanese, and Minangkabau influences. Although there have been studies on sound change and language kinship in Indonesia, their articles are fragmented and spread across multiple journals. There has been no thorough synthesis of the typology of sound change, lexical retention, language kinship, and the age or time of separation, including reconstruction of these languages, so they have not been comprehensively mapped. This study aims to bridge this gap.

The literature review serves as a strategic method for mapping and synthesizing research findings. By collecting and comparing the results of previous studies, authors can identify common patterns, methodological differences, and gaps that have not been addressed in earlier research. For instance, a study by Arlini & Saddhono (2024) discovered a kinship between Ogan and Javanese of only 0.18%, asserting that the two languages belong to different language families. In contrast, earlier classifications by several experts, including Dyen (1965), Nothofer (1975), and Adelaar (1992), categorized both languages within the same language family, specifically the Austronesian family, under the West Austronesian language group. This discrepancy has implications for glottochronology, leading to significantly different estimates of the ages or separation times for the two languages.

Another study showed a kinship between Geser and Ambonese Malay at 37%, which can be categorized as a language family (Mayangsari et al., 2025). In this study, primary data obtained from questionnaires distributed to informants were used for Geser, while secondary data obtained from an Ambonese Malay dictionary were used for Ambonese Malay. The data-collection method used could not capture sounds, preventing the determination of sound correspondences. In addition, comparing field vocabulary data (primary) with formal dictionary data (secondary) would compromise the validity of lexicostatistical calculations due to differences in register and the antiquity of words.

This study is essential for elucidating the reasons for discrepancies in research outcomes and for examining the evolution of comparative-historical linguistic studies in Indonesia. This study aims to address two primary questions: (1) To what extent have linguistic studies conducted in Indonesia over the past decade been grounded in diachronic linguistics or comparative historical linguistics? (2) What are the theoretical and methodological applications, as well as the research findings and their implications for understanding sound change and linguistic relationships in Indonesia? This study seeks to investigate sound evolution and linguistic affiliation through the lens of comparative historical

linguistics. This study specifically aims to (1) analyze linguistic research conducted in Indonesia over the past decade from the perspective of diachronic linguistics or comparative historical linguistics, and (2) identify the application of theories and methodologies, as well as research findings and their implications for understanding sound change and language relationships in Indonesia.

The study of historical and comparative linguistics is structured to investigate languages with kinship, establish groupings, reconstruct proto-language forms, and discover the mechanisms of language change (Crowley & Bowern, 2010). The primary assumption is that languages descended from a common ancestor will exhibit systematic similarities in phonological, morphological, syntactic, and lexical aspects.

Sound changes are one of the language changes in phonology that occur from one period to another. Campbell (1998)) and Crowley & Bowern (2010) classified the types of sound change into several main categories, among others: (1) Lenition and fortition. Lenition is the sound type occurring from strong to weak, such as *bulan 'moon'→ *fulan*. Weak sounds rarely become stronger (fortition); (2) Sound loss is the change in form of one or more sounds disappearing in a language, such as *niuR 'coconut' → *niu*. There are several types of sound loss, such as (a) apheresis (initial sound loss), (b) apocope (final sound loss), (c) syncope (medial sound loss), (d) cluster reduction (loss of one of the consonants in a syllable or a cluster reduction), and (e) haplology (loss of one syllable when it is also found in the following syllable or loss of a very similar syllable); (3) Sound addition from the original language to the derived one can occur in the form of (a) excrescence (addition of a consonant between the two consonants in a word), (b) epenthesis (a vowel addition to avoid a consonant sequence), and (c) prothesis (initial sound addition). Keraf (1984) added the final sound (paragoge); (4) Metathesis is a sound change in the form of the transposition of two phonemes; (5) Fusion, fission, and vowel Breaking. (a) Fusion is the sound change in which two original sounds merge into one sound, and the result of the single sound brings about some phonetic characteristics of the two original sounds; (b) Fission is the sound change that is opposite to fusion, in which a single sound splits into two, with each resulting sound preserving characteristics of the original; and (c) Vowel breaking, or diphthongization, is a sound change where a single vowel sound transforms into a diphthong, maintaining the original vowel while adding a glide either before or after it; (6) Assimilation is a sound change in which one sound replaces another. In this process, the different sounds are produced identically or similarly. When the phoneme changes, it is influenced by the phoneme that precedes it; therefore, this type of change is known as regressive assimilation. When the following changes and modifications are made to the proposed one, such assimilation is called progressive. Assimilation, too, may be differentiated into partial assimilation and total assimilation. If a sound change minimally retains one previous property, it is called partial assimilation. If all properties undergo changes and are modified by other sounds, then such assimilation is called total; (7) Dissimilation is the opposite of assimilation: a sound change in which identical or similar sounds are made different; and (8) Unusual sound change deviates from the categories discussed above.

To determine language kinship, the 200-word Swadesh list is commonly used, as it is well-suited for measuring lexical similarity. The traditional comparative method relies on regular sound correspondence, whereas the lexicostatistical method calculates the lexical similarity level to determine the kinship percentage between the compared languages. Furthermore, the chronological method is used to measure the time separating the compared languages.

The purpose of language reconstruction is to reconstruct the proto-language form reflected in the languages and modern dialects spoken by people. The reconstructed form is marked with an asterisk. Bynon (1979) defined protolanguage as a theoretical construct intended to correlate language or dialect systems with a historical relationship by describing the fundamental norms (Mahsun, 1995). By reconstructing a protolanguage, one can determine whether the contemporary language or dialect spoken by speakers in the present moment has undergone innovation or retention. In other words, the historical development of a language or dialect can be traced through language reconstruction. The above theories form the basis for analyzing the literature reviewed from a historical-comparative linguistic study perspective.

RESEARCH METHODS

This study draws on data from a relevant literature review on sound changes and language kinships. The data collection process using PRISMA (Preferred Reporting Items for Systematic

Reviews and Meta-Analyses) is divided into four major steps, depicted in a flow diagram. The first step, identification, is carried out by searching academic databases through the Sinta portal and Publish or Perish (PoP) software, using the keywords "language kinship", "lexicostatistics", "glottochronology", "historical-comparative linguistics", "Proto-Austronesian", "sound change", "assimilation", and "reconstruction", both in English and Indonesian. In the second step, screening, articles unrelated to the study's core issue are removed. The third step, eligibility, is determined by inclusion and exclusion criteria, which include (a) publishing between 2015—2025; (b) a direct connection to the topic of sound change and language kinship; and (c) the study rejecting general, non-scientific literature or literature with no apparent theoretical basis. This step involved a critical examination of the literature to identify central concepts, analytical methodologies, and key discoveries about sound change, linguistic kinship, and the process of language reconstruction. The fourth step was determining the final qualifying items.

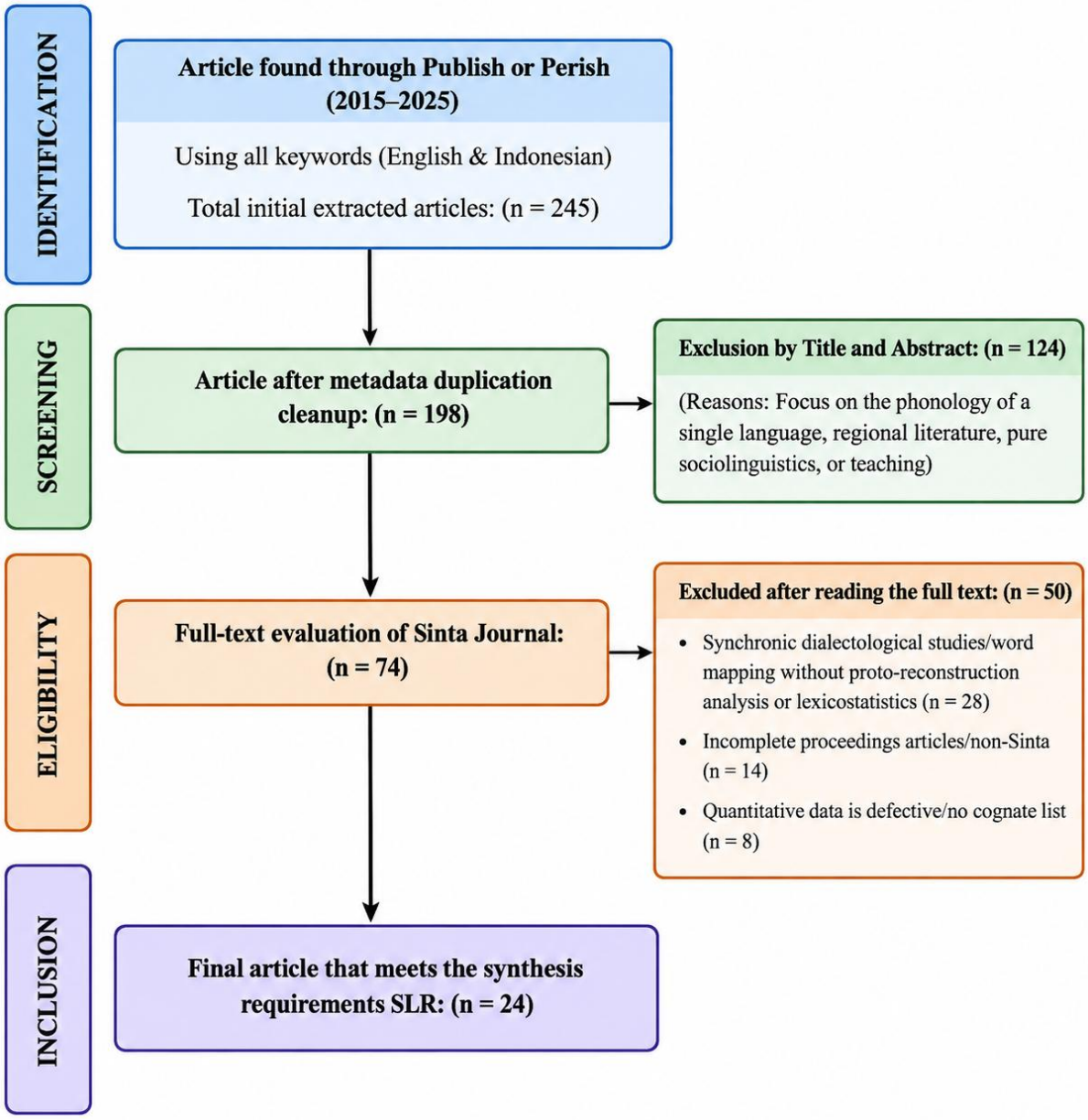


Figure 1. PRISMA Flowchart

There are 24 articles that meet the criteria and comprise the data. The results of the literature search and selection are presented in a table listing the number, journal name, author, year of publication, article title, and main findings. Subsequently, each study is analyzed thoroughly to achieve a holistic understanding aligned with the defined research objectives.

RESULTS AND DISCUSSION

Results

The articles are grouped into three parts: studies on sound change, kinship, and reconstruction. Each group is presented in a table. Each table contains a number, the journal name, the author, the year of publication, the article title, and the main finding presented in the article. Table 1 contains the studies on sound change.

Table 1. Studies on Sound Change

No.	Name of Journal	Author & Year	Article Title	Main Information
1.	EduCompassion: Jurnal Integrasi Pendidikan Islam dan Global	Mukramah (2025)	Sound Change of Austronesian Protolanguage to Minangkabau	There are some sound changes in metathesis, apheresis, syncope, apocope, prothesis, ephenthesis, and paragoge in Minangkabau at Kampung Padang, Medan.
2.	Jurnal Pendidikan Tambusai	Prasatya et al. (2024)	Sound Change of Austronesian Protolanguage into Palembangese Malay Dialect: Historical Comparative Study	There are some sound changes in forms of assimilation, metathesis, apheresis, syncope, apocope, ephenthesis, and paragoge.
3.	Deskripsi Bahasa	Sukma et al. (2024)	Sound Change of ProtoAustronesian into the Riau Malay Dialect of the Riau Language at Pasir Pengaraian	Sound Changes (assimilation, metathesis, apheresis, syncope, apocope, prothesis, ephenthesis, paragogue, and monophthongization). Semantic change (pejoration, widening, narrowing, and bifurcation).
4.	Linguistik Indonesia	Rahman et al. (2024)	DataLexical Base: Sound Change of Minangkabau Sangir Jujuan Isolect	Vowel lenition, consonant lenition, fortition, apheresis, apocope, haplology, syncope, ephenthesis, diphthongization, and metathesis
5.	Kajian Linguistik dan Sastra	Siregar et al. (2022)	Proto-Austronesian Sound Change (PAN) in Karonese, Tobanese, Pakpakse, Simalungunese, Mandailingnese, and Angkolan: Study of Historical Comparative Linguistics and Phonology	These six languages, Karonese, Tobanede, Pakpakse, Simalungunese, Mandailingnese, and Angkolan, belong to a Bataknese cognate language.
6.	Kulturistik: Jurnal Bahasa dan Budaya	Rizqi & Widayati (2021)	Sound Change of Proto-Austronesian into Javanese of Sumatran Dialect: A Study of Historical Comparative Linguistics	Javanese–Sumatran branch of Austronesian cognate within the particular local transformation (metathesis, apheresis, syncope, apocope, prothesis, and paragoge).
7.	Jurnal Bahasa and Sastra	Adhiti, (2019)	Study of Historical Comparative Linguistics on Sound Change Pattern	Sound changes are merging, breaking, uniting, clustering, sound loss, phoneme addition, and metathesis. Sound loss in apheresis, syncope, and apocope. Then there is a phoneme addition in prothesis, ephenthesis, and paragoge.
8.	Samudra Bahasa	Halimatussakdiah &	Sound Change of Proto-Austronesian into Malay of Tamiang Dialect (BMDT)	Metathesis, apheresis, syncope, apocope, prothesis, and paragoge.

No.	Name of Journal	Author & Year	Article Title	Main Information
		Widayati (2019)		
9.	Sociohumaniora – Jurnal Ilmu-ilmu Sosial dan Humaniora	Nadra et al. (2022).	Proto Malayic Vowel Phoneme Reflex in the Malay Language	The reflex of the Proto-Malayic vowel in Meranginese Malay varies in the final word syllable, with the greatest variation in OP1 Renah Alai due to the influence of Kerinci, Javanese, and Minangkabau.

Based on the data in Table 1, it appears that the language changes studied are sound changes from Proto-Austronesian to various regional languages or dialects in Indonesia (articles 1, 2, 3, 5, 6, and 8). One article discusses sound changes from Proto-Malayic (article 9), one from Proto-Minangkabau (article 4), and one does not explain them (article 7). The languages or dialects studied are the following: a) Malay (Palembang Malay, Riau Malay, Tamiang Malay, and Merangin Malay); b) Minangkabau (Minangkabau in Medan and Minangkabau in Sangir Jujuan); c) Batak (Karo, Toba, Pakpak, Simalungun, Mandailing, and Angkola); and d) Javanese (Javanese dialects in Sumatra).

Data from the studies above indicate that the process of sound change is systematic and consistent, but it varies across languages. Historical and comparative linguistics are the primary approaches for tracing the development of phonemes from protolanguage to modern languages and dialects. Mukramah (2025) research on the Minangkabau language, located in Kampung Padang, Medan, has identified seven types of phonological change: metathesis, apheresis, syncope, apocope, prothesis, epenthesis, and paragoge. These findings align with those of Prasatya et al. (2024) in Palembang, who also identified similar patterns of change, indicating sound adaptation to environmental sounds. Furthermore, Sukma et al. (2024) examined the Riau Malay dialect of Pasir Pengaraian and identified eight major sound change processes, followed by semantic changes, including pejoration, broadening, narrowing, and bifurcation. This emphasizes the close relationship between phonological evolution and semantic evolution. Various studies, such as those by Rahman et al. 2024), have shown that the Sangir Jujuan isolect in Minangkabau undergoes vowel softening, consonant softening, strengthening, and haplological processes, indicating a tendency to reduce and strengthen sounds in accordance with their phonological context.

Siregar et al. (2022) proved the kinship of the Batak language through a comparative analysis of six languages: Karonese, Tobanese, Pakpaknese, Simalungunese, Mandailingse, and Angkolan, highlighting the role of syncope, metathesis, apheresis, apocope, and paragoge in modern word formation. Other research, such as that conducted by Rizqi & Widayati (2021) supports the position that Javanese-Sumatra is part of the Austronesian language family through specific vowel sound transformations, but it does not explain what is called the Sumatran dialect of Javanese. Adhiti (2019) found complex patterns of sound change in Alorese, Kabola, and Hamap, consisting of mergers, groupings, phonemic additions, and metathesis. Furthermore, (Halimatussakdiah & Widayati Dwi, 2019) identified key sound change mechanisms in the Tamiangse Malay dialect, including prothesis, syncope, apocope, and paragoge, indicating significant phonological innovation. Another study, by Nadra et al. (2022), which examined Proto-Malay vowel reflexes in Merangin Malay, revealed phonemic variations in word-final syllables (the final syllable) resulting from language contact with Kerinci, Javanese, and Minangkabau. This emphasizes that sound change is not only phonologically internal but is also influenced by external factors and socio-cultural dynamics. Studies on language kinship are presented in Table 2.

Table 2. Studies on Kinship

No.	Name of Journal	Author & Year	Title of Article	Main Information
1.	Madah: Bahasa dan Sastra	Sulistiyarini & Hendrokumoro (2023)	Language Kinship of Javanese, Sundanese, and Makassarese: A Study of Historical	Kinship of Javanese–Sundanese 43%, Java–Makassarese 20%, and Sundanese–Makassarese

No.	Name of Journal	Author & Year	Title of Article	Main Information
			Comparative Linguistics	16%, separating period 2737 SM up to 231 M.
2.	Metamorfosa	Lestari & Mulyono (2023)	Kinship of Indonesian and Lowest Javanese Languages in Phonology and Vocabulary: A Study of Historical Comparative Linguistics	Kinship of Indonesian and lowest Javanese 24% at the cognate level, estimated to have separated approximately 3.386 million years ago.
3.	Masyarakat Linguistik Indonesia	Lailiyah & Wijayanti (2022)	Kinship of Javanese, Balinese, and Bima Languages: Historical Comparative Linguistics Perspective	Javanese, Balinese, and Bima are not closely cognate (<36%, submicrophylum); the closest kinship is Javanese–Balinese at the sub-cognate level, whereas Javanese–Bima and Balinese–Bima are at the microphylum level.
4.	Masyarakat Bahasa & Sastra Nusantara	Khairumi et al., (2024)	The Kinship of Minangkabau and Javanese Languages: A Study of Historical Comparative Linguistics	Kinship of Minangkabau and Javanese languages 28,3% level subgrouping families of stock, estimated separating time 2.664–3.164 years ago or approximately 640–1140 BC.
5.	Metamorfosa	Arlini & Saddhono (2024)	The Kinship of Oganese and Javanese Languages: A Study of Historical Comparative Linguistics	Language kinship is at 0,18%, showing the furthest relationship due to different cognate branching of Austronesian and Malay-Polynesian languages.
6.	Bahtera Indoneia: Jurnal Penelitian Pendidikan Bahasa dan Sastra Indonesia	Fadhila et al. (2024)	The Kinship of Teringinese and Kampung Baru Languages, West Kotawaringin Regency: Study of Historical Comparative Linguistics	Kinship of Teringinese and Kampung Baru languages 75%, belonging to a language family.
7.	Titian	Fitrah & Afria (2017)	Language Kinship of Malay Ethnicity, Bataknese, Sundanese, Bugisese, and Javanese in Jambi Province: a Study of Historical Comparative Linguistics	Kinship Percentage: Malay–Sundanese 44%, Malay–Bataknese 31,5%, Malay–Javanese 26%, and Malayu–Bugise 15%.
8.	Fondatia: Jurnal Pendidikan Dasar	Hilmi (2017)	Lexicostatistics of Sasambo, Sasak, Sumbawa/Samawa, and Bima/Mbojo languages: Study of Historical Comparative Linguistics	The languages of Sasak, Samawa, and Mbojo preserve a high degree of kinship and are classified as a language family.
9.	Deiksis	Hafizah (2018)	Lexicostatistics of Indonesian and Minangkabau: Dialect	Both the Indonesian and the Minangkabau languages have a language kinship rate of 82.8%,

No.	Name of Journal	Author & Year	Title of Article	Main Information
			of Bukit Tinggi (Study of Historical Comparative Linguistics)	with an estimated separation time of approximately 442 years.
10.	Ranah: Jurnal Kajian Bahasa	Mayangsari et al. (2025)	Rate of Cognate Geser and Ambonese Malay Languages	Based upon the lexicostatistical calculation method, it is found that the kinship percentage of Geser and Ambonese Malay ranges from 37% can be categorized into a language family. Both languages are separate from their parent languages approximately 561 BC up to. 23 BC.

All articles in Table 2 focus on regional languages in Indonesia belonging to the Proto-Austronesian family, such as Javanese, Sundanese, Malay, Minangkabau, Banjar, and Bima. The studies can be divided into two groups: those that focus on kinship and the time of language separation (articles 1, 2, 4, 9, and 10) and those that focus on kinship alone, without any time separation (articles 3, 5, 6, 7, and 8). Based on these studies, the closest kinship relationship is found between Indonesian and the Minangkabau language spoken in Bukittinggi, while the most distant relationship is found between Ogan and Javanese.

The kinship between Javanese, Sundanese, and Makassarese varies in degree of closeness. Sulistiyarini & Hendrokumoro (2023) found that the kinship percentages for Javanese–Sundanese, Javanese–Makassarese, and Sundanese–Makassarese were 43%, 20%, and 16%, respectively, over a time span estimated at 2737 BC to 231 AD. Furthermore, Lailiyah & Wijayanti (2022) confirmed that the kinship of Javanese–Balinese is sub-cognate, with a closer kinship than Javanese–Bima and Balinese–Bima, reaching only the micro-phylum level (<36%). These findings align with the geographical conditions and the history of language dissemination in the areas where the Balinese migrated, as well as with cultural ties to the Javanese. In contrast, Bima is closer to the Sasambo's group in the eastern region.

The research findings within the Malay-language group demonstrate a high level of internal cohesion. Hafizah (2018) noted that the Indonesian and Minangkabau languages share a high degree of kinship, with a similarity index of 82.8%, suggesting they diverged approximately 442 years ago. This relationship highlights the historical connections between the languages, shaped by extensive contact and the influence of modern standardization through Indonesian. In the Jambi region, Fitrah & Afria (2017) reported the following kinship percentages: Malay–Sundanese at 44%, Malay–Batak at 31.5%, Malay–Javanese at 26%, and Malay–Bugis at 15%. These figures indicate varying degrees of linguistic distance, from close to more distant relationships. Furthermore, Oktariza & Sufiyandi (2017) identified 11 systematic phonemic correspondences between the Jambi Seberang Malay and Bungo Malay isolects, suggesting patterns of regular divergence and innovation within the Malay language group. This is supported by Alfikri et al. (2023), who reconstructed 26 protophonemes in Merangin Malay, including 5 vowels, 18 consonants, and 3 diphthongs, offering valuable insights into the relict phonology of the Malay group.

The Sasambo languages (Sasak, Samawa/Sumbawa, and Mbojo/Bima) also show high cohesion at the subgroup level. Hilmi (2017) stated that these languages are closely related and derive from a proto-language, which is why they are grouped into a single language family. This finding is consistent with patterns of phonological divergence and common innovations, thus indicating a historical connection in Lombok–Sumbawa. Fadhila et al. (2024) found that the Teringin and Kampung Baru languages in West Kotawaringin are 75% related, meaning they are both part of a language family with a very high level of basic vocabulary kinship. On the other hand, Arlini & Saddhono (2024) found a kinship between the Ogan and Javanese languages of only 0.18%. Geographically, the Ogan language is located in South Sumatra and belongs to the Austronesian language family, while the Javanese language is located in Java and belongs to the Malayo-Polynesian language family. Arlini &

Saddhono (2024) statement indicates that the Ogan language belongs to the Austronesian language family, and the Javanese language belongs to the Malayo-Polynesian language family, implying that the two languages belong to different language families. Other research shows a kinship between Geser and Ambonese Malay, with a kinship percentage of 37%, which can be categorized as a language family Mayangsari et al. (2025).

Several studies have identified the inter-language separation time, providing insights into the chronology of language divergence. Lestari & Mulyono (2023) said that both Indonesian and Low Javanese languages possess a kinship rate as 24% under a separation time estimated at as long as 3.386 years ago, meanwhile Khairumi et al. (2024) found that both the Minangkabau and Javanese languages have a kinship rate of 28.3% under a separation time estimated at 2.664–3.164 years ago, or approximately 640–1140 BC. Mayangsari et al. (2025) said that both the Geser and the Ambonese Malay are separate from their parent language from 561up to 23 BC. These historical insights, though fascinating, require meticulous interpretation, as they also depend upon the lexical retention rate and the effect of language contact Additionally, research on protolanguage reconstruction is presented in Table 3.

Table 3. Studies on Protolanguage Reconstruction

No.	Name of Journal	Writer & Year	Article Title	Main Information
1.	Krinok: Jurnal Linguistik Budaya	Oktariza & Sufiyandi (2017)	Austronesian Phonemic Correspondence in Jambi Malay Language: Comparative Analysis over Jambi Malay Seberang and Bungo Malay Isolects	There are 11 systematic phonemic correspondences across the Jambi Malay, Seberang Malay, and Bungo Malay isolects, based on 100 questionnaires.
2.	Pendidikan Bahasa dan Sasta	Alfikri et al. (2023)	Phonemic Reconstruction of Malay Language in Merangin Regency, Jambi Province	Twenty-six forms of proto phonemes were found in the research: 5 vowels, 18 consonants, and 3 diphthongs
3.	Journal of Nusantara Studies (JONUS)	Aman et al. (2018)	Internal Reconstruction of Banjar Beriah Purba Variant	The changes of both vowel and consonant phonemes at the word initial and final. The reconstructed relic forms showed a close kinship to the protolanguage form in the Malay cognate.
4.	Arbitrer	Budasi (2021)	The Reconstruction of Etymon Proto of Sambori and Teta Isolects Spoken in Bima Regency, NTB, Indonesia	Etymon Reconstruction of proto-Sambori and Teta languages demonstrated their kinship and reconstructed preliminary phonologies for both languages, with conservation and mapping in East Indonesia.
5.	Arbitrer	Cho (2020)	A Review of Family Context and Reconstruction Problems in the Austronesian Language Family	This literature-based article examines the widespread development of Austronesian cognate languages and the problems in reconstructing the Austronesian languages in the Archipelago. The reconstruction of proto-Austronesian faces challenges in the form of dialectal variation, lexical inconsistency, and instability in sound correspondences.

The studies presented in Table 3 focus on proto-language reconstruction. Besides the kinship percentage, the phonological evidence, through sound correspondences and proto-phoneme reconstruction, provided a stronger genealogical foundation. The etymon reconstruction of Sambori and Teta Protolanguage by Budasi (2021) proved and mapped out preliminary phonology for the conservation of languages in Bima. Aman et al. (2018) also reconstructed the relic forms of Banjar Beriah variants, showing changes in both vowels and consonants in word-initial and word-final positions, as well as kinship proximity to the protolanguage form of the Malay cognate. Cho (2020) underlined the challenges in reconstructing the Austronesian languages, including dialectal variation, instability in sound correspondences, and lexical inconsistency. The findings indicated that analyzing language kinship requires methodological triangulation, combining lexicostatistics, reconstruction, and phonological evidence to achieve accurate results.

Discussion

Based on this literature review, historical and comparative linguistics studies in Indonesia not only show a diversity of findings regarding sound changes and language kinship but also reveal several methodological issues that influence the interpretation of research results. Therefore, the evaluation focuses on three aspects, namely (1) methodological trends in phoneme reconstruction and language kinship determination, (2) misclassification of language genealogies, and (3) the validity of determining the period of language separation.

In Indonesia, comparative historical linguistics typically uses lexicostatistical methods for language classification. This method is based on the Swadesh vocabulary list. This trend is evident in studies conducted by Hafizah (2018), Fitrah & Afria (2017), (Hilmi, 2017), Lailiyah & Wijayanti (2022), Sulistiyarini & Hendrokumoro (2023), Lestari & Mulyono (2023), Khairumi et al. (2024), and Mayangsari et al. (2025). Some studies go beyond lexicostatistics; others continue with glottochronology. In general, these studies use the percentage of related words as the primary basis for determining language kinship. Applying this method is indeed easy because it relies on data from only 200 basic vocabulary words. However, the dominant use of this technique demonstrates the dependence of comparative historical linguistics research on this methodology. A basic vocabulary, as the name suggests, consists of very basic words that all language communities in the world share. Borrowing basic vocabulary from other languages is extremely rare. Therefore, this vocabulary is relatively stable. This assumption is not entirely fulfilled. Language communities, such as the Nusantara community, are dynamic. Their ancestors were renowned as brave sailors who sailed the oceans, crossed islands, and established trade and migration networks. In this case, language contact is unavoidable. For example, the Banjar Hulu language includes the words *ŋara*n 'name' and *hiraŋ* 'black'. Both words are included in the basic Swadesh vocabulary, which originates from Javanese (Adelaar, 1992). This means that the realization of concepts in the basic Swadesh vocabulary sometimes undergoes changes through borrowing. In such cases, these words must be excluded from the comparison.

Moreover, the majority of historical and comparative linguistics studies have aimed to determine the percentage of cognates utilizing 200 fundamental Swadesh vocabulary terms. A criterion for identifying cognates is that the compared words exhibit sound correspondences (Crowley & Bowern, 2010). Sound correspondences are insufficiently robust when derived from a single word pair. Additional word pairs must be incorporated. These pairs of words will demonstrate consistent sound correspondences. Phonetic transcriptions are required to ascertain sound correspondences. Sound correspondences are essential for reconstructing proto-phonemes. These proto-phonemes will constitute the proto-lexicon. Transcription via orthographic methods, as demonstrated by (Arlini & Saddhono, 2024), may encounter issues when a single phoneme is represented by two letters; for instance, the phoneme *ŋ* is orthographically transcribed as *ng*.

In addition, research by Alfikri et al. (2023) and Aman et al. (2018) both examined the reconstruction of protolanguage phonemes, one for the Malay language spoken in Merangin, Jambi Province, and the other for Banjar Beriah. Besides that, Budasi (2021) investigated the correlation between Proto-Sambori-Teta and its progenitor, Proto-Austronesian, elucidating the phonetic alterations that transpired between the two proto-languages. The three studies show that phoneme reconstruction provides a more robust basis for tracking linguistic history than calculating related-word percentages. Research in comparative historical linguistics should employ quantitative methods that

incorporate lexicostatistical and glottochronological techniques, alongside a qualitative approach that involves correspondence analysis and the reconstruction of phonemes and lexicons.

Conversely, Adhiti (2019) asserted that the approach employed to ascertain sound change patterns is a syncomparative strategy, namely, identifying similarities and differences across related contemporary languages. Sound change patterns cannot be determined by syncomparative methods. Changes can be observed diachronically by contrasting historical and contemporary forms. The study delineates protophoneme forms denoted by an asterisk (*). Nonetheless, there is no clarification about whether these protophoneme types stem from reconstructions based on extant evidence or from reconstructions conducted by other scholars.

Lexicostatistics is a method for categorizing or classifying languages. The foundation for determining this classification level is the 200 or 100 fundamental vocabulary terms established by Morris Swadesh (Crowley & Bowern, 2010); Keraf, 1984). These fundamental vocabulary terms are universal, meaning they are present in all languages. Arlini & Saddhono (2024) conducted a comparison of Ogan and Javanese, quantifying the percentage of cognate terms. Their report indicated that these 200 fundamental vocabulary words were utilized. Nonetheless, their examination inadequately encompassed fundamental words, such as radio, coffee, closet, luxury, flashlight, buying food, vinegar, and Eid. Certain terms constitute cultural vocabulary and represent secondary needs. Such terms are prone to cultural influences and borrowing. Consequently, they are unsuitable for assessing genealogical ties among languages. In assessing kinship, similarities arising from borrowing must be eliminated from the comparison (Keraf, 1984), including loanwords from their own cognate languages.

Arlini & Saddhono (2024) focused on determining kinship words based more on form similarity rather than conducting a sound correspondence analysis. Relying on a single word pair to establish the sound correspondence set is insufficient. Therefore, it is necessary to include additional word pairs to strengthen the sound correspondence set. This procedure is not clearly outlined in the text. Another issue that needs to be criticized in the study by Arlini & Saddhono (2024) there is an inconsistency in the calculation of kinship. The percentage of related words is 0.18%. The level of subgrouping is the mesophyla of a macrophylum. This figure indicates a calculation error. The result of the calculation should be 18%. It means the level of subgrouping is at the family level for a stock. This computational inaccuracy substantially affects the assessment of language levels or classifications. The misclassification would have been avoided had verification been conducted using credible historical classifications, such as those proposed by Adelaar (1992), Dyen (1965), and Nothofer (1975), which states that both languages belong to the Austronesian language family, specifically the West Austronesian group. Moreover, verification may also be conducted by sound correspondence analysis of the languages under comparison.

Establishing the temporal separation of languages through the glottochronology approach is a primary objective of historical and comparative linguistics research. Determining this separation period requires the proportion of kinship terms (cognates) derived from lexicostatistical approaches, as evidenced in the studies of Hafizah (2018) and Sulistiyarini & Hendrokumoro (2023). This approach has limitations, such as its assumption that the rate of vocabulary change is relatively constant across all languages. In reality, these changes can result from various external factors, such as migration, trade, population mobility, and language contact. These factors can accelerate or slow down language change, so the level of lexical retention does not always reflect the actual separation age.

This constraint is evident in the research by Mayangsari et al. (2025) on Geser and Ambonese Malay. Both languages evolved in regions characterized by a history of language contact. In some instances, lexical similarities arise not solely from genealogical connections but also from borrowings. Consequently, calculating the percentage of cognate words has the potential to produce inaccurate estimates of separation time if it does not adequately account for the influence of language contact. In addition to this issue, the data sources in the study also differed: field data for Geser and dictionary data for Ambonese Malay. This disparity may lead to prejudice due to the distinct attributes of the documentation from the two sources. This disparity can be elucidated through data veracity, degree of uniformity, and temporal recording. Field data reflect the language actually used by the community, so the phonetic transcription matches the informants' speech. This advantage of field data is not found in dictionary data. Moreover, dictionary data has been selected, edited, and standardized. The language varieties are not necessarily the same or comparable. In terms of documentation time, the data collection time may differ. Another advantage of field data is that clarification can be provided directly with

informants. This elucidation cannot be achieved with data derived from dictionaries. These differences can affect the identification of cognate words and, consequently, the calculation of language separation time.

Evidence of language contact was also identified in a study by Hafizah (2018) about the Indonesian and Minangkabau. The two languages share a close genealogical relationship and are in intensive contact through education, mass media, administration, and formal communication. This condition allows for lexical borrowing from Indonesian into Minangkabau. Consequently, some vocabulary similarities do not entirely reflect retention of the same parent language, but rather the result of this language contact. This situation can lead to high lexical similarities and similarities between the two languages today. This will affect the percentage of cognate words and the time elapsed since the two languages were separated. Therefore, the high percentage of cognate words needs to be interpreted carefully. Variations in the level of kinship depend not only on geographic distance but also on migration, language contact, and the internal evolution of languages.

CONCLUSION

The results of this study show that common phonological changes include metathesis, apheresis, syncope, apocope, prothesis, epenthesis, paragoge, monophthongization, and diphthongization. In some cases, they are accompanied by changes in meaning, such as pejoration, expansion, or narrowing. These confirmed that the sound does not change randomly but follows a given constraint that can be historically retraced. Analysis of language kinship, utilizing lexicostatistics and glottochronology, uncovers subgrouping levels of language affiliation, extending from language level to macrophylum, with estimated separation times thousands of years ago. The results of this study also revealed several methodological issues, misclassifications, and challenges in determining the language separation period, including data discrepancies, inaccurate transcriptions, errors in identifying cognate words and sound correspondences, and calculation errors. This resulted in misclassification of groupings and kinship relationships, as well as in the determination of the separation period of the compared languages, even though the results were illogical. Furthermore, the studies were often not supported by qualitative analysis.

To achieve more precise results, the assessment of language kinship should incorporate both quantitative methods, such as lexicostatistics and glottochronology, and qualitative approaches, like sound correspondence analysis and protolanguage reconstruction. Additionally, it is essential to consider the patterns of sound change in the languages being compared. Migration, language contact, and geographic isolation need to be identified as external factors that can influence subgrouping levels.

This study is expected to contribute not only methodologically and theoretically to the development of historical and comparative linguistics but also to uphold the practical value of preserving language history and strengthening cultural identity. Most importantly, historical linguistics is a continuous journey of discovery, and continued research in this field is crucial for enriching language documentation and deepening understanding of the history and evolution of cognate languages.

ACKNOWLEDGEMENT

The authors would like to thank Universitas Andalas, through LPPM Universitas Andalas, for funding the research project under Contract Number 458/UN16.19/PT.01.03/PLB/2025, dated 14 April 2025.

REFERENCES

Adelaar, K. A. (1992). *Proto Malayic: the Reconstruction of its Phonology and Parts of its Lexicon and Morphology*. Pacific Linguistics, C-119

Adhiti, I. A. I. (2019). Kajian Linguistik Historis Komparatif pada Pola Perubahan Bunyi. *KULTURISTIK: Jurnal Bahasa dan Budaya*, *3*(2), 75. https://doi.org/10.22225/kulturistik.3.2.1203

Alfikri, M., Nadra, N., & Marnita, R. (2023). Rekonstruksi Fonem Bahasa Melayu di Kabupaten Merangin Provinsi Jambi. *Jurnal Pendidikan Bahasa dan Sastra*, *23*(1), 91–108.

Aman, R., Othman, N. N., A.H, S., Iswary, E., & O.D, A. (2018). Rekonstruksi Dalaman Varian Banjar

Beriah Purba. *F Nusantara Studies 2018*, *Vol 3(2) 9*(2), 92–106.
Arlini, E. W., & Saddhono, K. (2024). Kekerabatan Bahasa Ogan dan Bahasa Jawa: Kajian Linguistik Historis Komparatif. *Jurnal Metamorfosa*, *12*(1), 30–44. https://doi.org/10.46244/metamorfosa.v12i1.2424
Budasi, I. G. (2021). The Reconstruction of Etymon Proto of Sambori and Teta Isolects Spoken in Bima Regency-NTB-Indonesia. *Jurnal Arbitrer*, *8*(1), 14. https://doi.org/10.25077/ar.8.1.14-24.2021
Bynon, T. (1979). *Historical Linguistics*. Cambridge University Press.
Campbell, L. (1998). *Historical Linguistics: An Introduction*. Edinburgh University Press.
Cho, M. (2020). A Review About Family Context and Reconstruction Problems in the Austronesian Languages Family. *Jurnal Arbitrer*, *7*(2), 210–220. https://doi.org/10.25077/ar.7.2.210-220.2020
Crowley, T., & Bowern, C. (2010). *An Introduction to Historical Linguistics* (Fourth Ed.). Oxford University Press.
Dyen, I. (1965). *The Lexicostatistical Classification of the Austronesian Languages*.
Ethnologue. (2023). *The Languages of the Word.* https://www.ethnologue.com/language/ind/
Fadhila, A., Saddhono, K., & Suwandi, S. (2024). Kekerabatan Bahasa Teringin dan Bahasa Kampung Baru Kabupaten Kotawaringin Barat: Kajian Linguistik Historis Komparatif. *Bahtera Indonesia; Jurnal Penelitian Bahasa dan Sastra Indonesia*, *9*(2), 604–612. https://doi.org/10.31943/bi.v9i2.775
Fitrah, Y. & Afria, R. (2017). Kekerabatan bahasa-bahasa etnis Melayu, Batak, Sunda, Bugis, dan Jawa di Provinsi Jambi: Sebuah kajian linguistik historis komparatif. *Titian: Jurnal Ilmu Humaniora*, *1*(2), 204–2018. https://online-journal.unja.ac.id/index.php/titian
Hafizah, H. (2018). Leksikostatistik Bahasa Indonesia dengan Bahasa Minang Dialek Bukittinggi (Kajian Linguistik Historis Komparatif). *Deiksis*, *10*(03), 247. https://doi.org/10.30998/deiksis.v10i03.2757
Halimatussakdiah, & Widayati. D. (2019). Perubahan Bunyi Bahasa Proto-Austronesia ke dalam Bahasa Melayu Dialek Tamiang (Bmdt). *Jurnal Samudra Bahasa*, *2*(2), 33–39. http://ejurnalunsam.id/index.php/JSB
Hilmi, M. (2017). Leksikostatistik Bahasa (Sasambo) Bahasa Sasak, Bahasa Sumbawa/Samawa, dan Bahasa Bima/Mbojo : Kajian Linguistik Historis Komparatif. *Fondatia: Jurnal Pendidikan Dasar*, *1*(1), 165–176.
Kementerian Pendidikan dan Kebudayaan (Ministry of Education and Culture). (2019). *Bahasa dan Peta Bahasa di Indonesia* (6th ed.). Kementerian Pendidikan dan Kebudayaan.
Keraf, G. (1984). *Linguistik Bandingan Historis*. Gramedia.
Khairumi, N., Suyuti, M. W., & Cahyaningrum, R. (2024). Kekerabatan Bahasa Minangkabau dan Bahasa Jawa: Kajian Linguistik Historis Komparatif. *Mabasan* (Vol. 18, Issue 2, pp. 235–252). https://doi.org/10.62107/mab.v18i2.906
Lailiyah, N., & Wijayanti, F. I. (2022). Kekerabatan Bahasa Jawa, Bali dan Bima: Perspektif Linguistik Historis Komparatif. *Linguistik Indonesia*, *40*(2), 327–345. https://doi.org/10.26499/li.v40i2.308
Lestari, W. F., & Mulyono, S. (2023). Kekerabatan Bahasa Indonesia dan Bahasa Jawa Ngoko Bidang Fonologi dan Kosakata: Kajian Linguistik Historis Komparatif. *Journal Metamorfosa*, *11*(2), 128–144. https://ejournal.bbg.ac.id/metamorfosa
Mahsun. (1995). *Dialektologi Diakronis: Sebuah Pengantar*. Gadjah Mada University Press.
Mayangsari, D., Hendrokumoro, H., Kusumawicitra, N., & Madani, A. (2025). *Taraf Kognat Bahasa Geser dan Bahasa Melayu Ambon (Cognate Frequency of Geser Language and Ambonese Malay Language Dewi Mayangsari)*. 14*, 30–44.
Mukramah, M. (2025). Perubahan Bunyi Bahasa Proto Austronesia ke Bahasa Minangkabau. *Educompassion: Jurnal Integrasi Pendidikan Islam dan Global*, *2*(1), 100–104. https://doi.org/10.63142/educompassion.v1i3.120
Nadra. (2006). *Rekonstruksi Bahasa Minangkabau*. Andalas University Press.
Nadra, N., Marnita, R., Alfikri, M., & Masni, A. K. (2022). Proto Malayic Vowel Phoneme Reflex in Malay Language. *Sosiohumaniora*, *24*(3), 452. https://doi.org/10.24198/sosiohumaniora.v24i3.29612
Nothofer, B. (1975). *The Reconstruction of Proto-Malayo-Javanic*. Martinus Nijhoff.
Oktariza, D., & Sufiyandi. (2017). Korespondensi Fonem Austronesia dalam Bahasa Melayu Jambi: Telaah Komparatif terhadap Isolek Melayu Jambi Seberang dan Isolek Melayu Bungo. *Jurnal*

Linguistik Budaya (Vol. 2, Issue 1). http://ojs.umb-bungo.ac.id/index.php/Krinok/index

Prasatya, R. F., Dardanila, & Bangun, P. (2024). Perubahan Bunyi Bahasa Proto Austronesia ke dalam BahasaPalembang Dialek Melayu Palembang: Kajian Linguistik Historis Komparatif. *Jurnal Pendidikan Tambusai*, *8*, 17830–17856.

Purwo, B. K. (2003). *Research on Nusantara Languages in Indonesia. International Mother Language Day Seminar Paper*.

Rahman, F., Rina, N., & Kurniati, S. (2024). Basis Data Leksikal: Perubahan Bunyi Bahasa Minangkabau Isolek Sangir Jujuan. *Linguistik Indonesia*, *42*(1), 185–198. https://doi.org/10.26499/li.v42i1.572

Rizqi, F. A., & Widayati, D. (2021). Perubahan Bunyi Bahasa Proto Austronesia ke dalam Bahasa Jawa Dialek Sumatera: Kajian Linguistik Historis Komparatif. *Kulturistik: Jurnal Bahasa dan Budaya*, *5*(2), 29–35. https://doi.org/10.22225/kulturistik.5.2.3380

Siregar, E. D., Ernanda, E., & Afria, R. (2022). Perubahan Bunyi Bahasa Proto Austronesia (PAN) pada Bahasa Karo, Bahasa Toba, Bahasa Pakpak, Bahasa Simalungun, Bahasa Mandailing dan Bahasa Angkola: Kajian Linguistik Historis Komparatif dan Fonologi. In *Kajian Linguistik dan Sastra* (Vol. 1, Issue 2). https://doi.org/10.22437/kalistra.v1i2.20294

Sukma, Widayati, D., Dardanila, & Lubis, T. (2024). Perubahan Bunyi Bahasa Proto Austronesia ke dalam Bahasa Melayu Riau Dialek Pasir Pengaraian. *Deskripsi Bahasa*, *7*(2), 99–121.

Sulistiyarini, S., & Hendrokumoro, H. (2023). Hubungan Kekerabatan Bahasa Jawa, Sunda, dan Makassar: Kajian Linguistik Historis Komparatif. *Madah: Jurnal Bahasa dan Sastra*, *14*(2), 187–202. https://doi.org/10.31503/madah.v14i2.633