Tracing the Linguistic Roots of Malay and Batak Languages in Sumatra Island: A Historical Comparative Study

Riska Meliana[1,*], Manna Maria Sopiana Manalu[2], Sulis Triyono[3]

[1]*Applied Linguistics Department, Faculty of Languages, Arts, and Culture, Universitas Negeri Yogyakarta, Yogyakarta 55281, Indonesia*
[2]*Applied Linguistics Department, Faculty of Languages, Arts, and Culture, Universitas Negeri Yogyakarta, Yogyakarta 55281, Indonesia*
[3]*Applied Linguistics Department, Faculty of Languages, Arts, and Culture, Universitas Negeri Yogyakarta, Yogyakarta 55281, Indonesia*

ARTICLE INFO

Keywords:
Batak language;
Glottochronology;
Lexicostatistics;
Linguistics historical comparative;
Malay language

Article History:
Received : 12/03/2024
Revised : 14/05/2024
Accepted : 25/05/2024
Available Online: 27/05/2024

ABSTRACT

Previous research in comparative historical linguistics has traditionally focused on languages within a single region, overlooking cognate languages in other areas. This study seeks to rectify this by quantitatively and qualitatively describing the kinship between Rejang, Serawai, Lembak (Bengkulu), and Toba, Mandailing, and Nias (North Sumatra) languages. It aims to unearth empirical evidence regarding the timing of divergence between Malay and Batak languages, as well as the grouping of languages and the percentage of kinship between Bengkulu Province and North Sumatra Province. Utilizing Morris Swadesh's lexicostatistics and glottochronology methods, the research evaluates word kinship based on a fundamental 150-word list. Results indicate significant differences among the six languages, particularly with Rejang and Nias displaying low similarity levels, falling below 30% and not even reaching 10%, respectively. The percentage of kinship between local language pairs in Bengkulu and North Sumatra Province averages at 22.66%, classifying them under the "Family stock" category, indicating identical word correlations despite differing phonetic elements. Glottochronological calculations estimate the separation time between Malay and Batak languages to range from 419 to 3,289 BC. This research significantly enhances understanding of regional language kinship and linguistic diversity.

How to cite (in APA style): Meliana, R., Manalu, M. M. S., & Triyono, S. (2024). Tracing the Linguistic Roots of Malay and Batak Languages in Sumatra Island: A Historical Comparative Study. *OKARA: Jurnal Bahasa dan Sastra, 18*(1), 142-164. https://doi.org/10.19105/ojbs.v18i1.12865

1. INTRODUCTION

Language plays a vital role in understanding cultural heritage and diversity in a region and passing on knowledge from one generation to another (Mailani et al., 2022). The movement of people from one region to another causes the language to become separated from the parent language or mother tongue because it adapts to social, natural, and environmental contacts where the community lives. The language developed 100,000 years ago (Mahriyuni et al., 2023). The evolution of language in this world has gone through a

***Corresponding Author**: Riska Meliana riskameliana.2022@student.uny.ac.id

very long stage. The linguistic development of a language is inseparable from the kinship or similarity of one language with another (Istanti et al., 2020). Language kinship is a collection of languages from a language group with similarities and the same development history (Setiawan, 2020). Thus, languages that are related or have similar vocabularies are the same proto-language.

The vocabulary across those languages exhibits notable similarities and differences. For example, languages spoken in Bengkulu province and North Sumatra. Some of the spoken words that show similarities and differences are exhibited in the word "work," which is spoken as [kulaghan] in Serawai, Bengkulu province, and [pakarejoan] in Toba, North Sumatra. Nevertheless, certain parallels exist, as observed in the word for "water," which is denoted as [aiak] in Serawai and [aek] in Toba and Mandailing. However, not all languages spoken in the Bengkulu and North Sumatra provinces could be included in this study due to the limited number of sources and the extensive scope of the discourse. These languages cover a wide range of speaker numbers, from those with tens of millions of speakers such as Malay or Indonesian, Javanese, and Tagalog (Tryon, 2006).

Indonesia is a multicultural and multilingual country that studies the languages of various parts of the world, and experts have classified these languages into 13 families. The grouping is based on the criteria of phonology and vocabulary, and one of the language families is Austronesian (Martius, Hasbi, & Rehayati, 2023). Experts have categorized this Austronesian family into two sub-groups: the West Austronesian sub-group and the East Austronesian sub-group (Pawley, 2007; Ross, 2020). This Western Austronesian family includes Malay or Nusantara languages (Martius et al., 2023). Thus, along with development, people move from one region to another, causing the language to become separated from the parent language or mother tongue because it adapts to social contact, nature, and the environment in which the community lives.

Research on linguistic diversity in Indonesia, particularly on the island of Sumatra, significantly reflects Indonesia's cultural and linguistic complexity (Collins, 2019; Zein, 2020; Espree-Conaway, 2022). Classifying these languages into specific families, such as the Austronesian family, helps understand their origins and relationships. In this context, identifying the languages in Bengkulu Province and North Sumatra Province becomes relevant because these provinces have unique cultural and linguistic heritages (Tarigan, 2016; Samsudin, 2017). Through language mapping conducted by *Badan Pengembangan and Pembinaan Bahasa* (LPPB), we can understand how these languages have evolved and adapted to their social and environmental surroundings (Mahsun et al., 2017). Bengkulu Province, for example, is one of the regions in Sumatra that hosts a variety of indigenous languages, which may have undergone separation from their parent languages due to social interactions and environmental influences (Samsudin, 2017).

Several contextual issues were considered when selecting data on the languages of Bengkulu province. Firstly, Bengkulu has a rich history in culture and linguistics, so it is essential to understand linguistic diversity (Adelaar et al., 1996; Omar, Jaafar, & Mat, 2015). For one, the celebration of the origins of the Tabot ritual in Bengkulu demonstrates kinship and sociocultural systems (Wahyuni et al., 2021). Second, language mapping can assist in the preservation and revitalization efforts of minority languages that might threatened with extinction (Rahayu, 2018; Zabadi et al., 2023; Pramuniati, Mahriyuni, & Syarfina, 2024). By understanding the geographical distribution and number of speakers for these languages, steps can be taken to promote their use and development. In addition, this research can

also provide insight into the social dynamics and human migration that influence language development in the region.

Thirty-three languages were identified on the island of Sumatra in 2019. There are fifteen regions with six original languages in Bengkulu Province and nine regions with five Batak languages spread across North Sumatra Province that have been identified through the Language Mapping study in Indonesia by Badan Pengembangan dan Pembinaan Bahasa (formerly Pusat Bahasa/Center for Language) (Mahsun et al., 2017). The following is the identification of languages in Table 1 and the number of regions indicated in Table 2.

Table 1
Language Identification

Bengkulu Province		**North Sumatra Province**	
Indigenous language	*Not indigenous language*	*Indigenous language*	*Not indigenous language*
Bengkulu language	Jawa language	Batak language	Jawa language
Enggano language	Minangkabau language	Nias language	Melayu language
Rejang language	Sunda language	-	Minangkabau language

Source : (Mahsun et al., 2017)

Table 2 lists the language distribution areas in Bengkulu and North Sumatra provinces. It includes specific regions and sub-districts where various languages are used. This information is valuable for understanding the linguistic diversity in these areas and can be used to promote language preservation efforts and cultural diversity.

Table 2
Language Distribution Areas

No	**Bengkulu Province Region**	**No**	**North Sumatra Province**
1.	Ipuh sub-district	1.	Asahan Regency
2.	Teluk Segara sub-district	2.	Tanjung Balai City
3.	Muara Bangkahulu sub-district	3.	Simalungun Regency (especially the west coast)
4.	Bengkulu City		
5.	Pelalo Village	4.	Dairi Regency
6.	Taba Tinggi village, Padang Ulak Tanding area	5.	Central Tapanuli Regency
7.	Rejang Lebong district	6.	North Tapanuli Regency
8.	Tanjung Betuah Village	7.	Karo Regency
9.	Merpas area	8.	Langkat Regency
10.	Southern Bengkulu	9.	Deli Serdang Regency
11.	Kepahiang area		
12.	Ketahun Village (Air Lelangi) and Southern Muko-Muko		
13.	South Kaur (Jembatan Dua and Tanjung Bunga)		
14.	Central Kaur (Lubuk Gung)		
15.	Gading Cempaka Village (Tanah Patah)		

Source : (Mahsun et al., 2017))

Tables 1 and 2 show the distribution of language areas found in several Bengkulu Province and North Sumatra regions. Bengkulu Province has three indigenous languages, such as Bengkulu, Enggano, and Rejang, and three non-indigenous languages (Javanese, Minangkabau, and Sundanese). Meanwhile, North Sumatra Province has two indigenous languages (Nias and Batak) and three non-indigenous languages (Javanese, Malay, and Minangkabau).

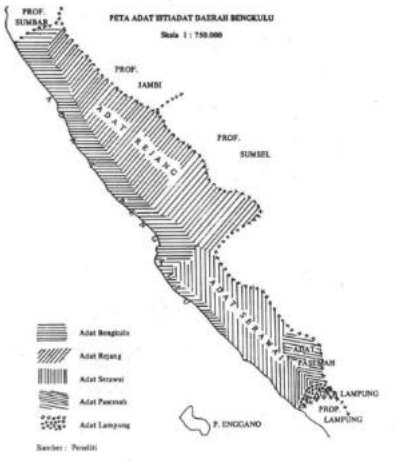

Fig 1. Customs Map of Bengkulu Province (Departemen Pendidikan dan Kebudayaan, 1977)

According to Figure 1, the distribution of the regional customs map shows five customs spread across Bengkulu Province, and two traditions dominate, namely Adat Rejang and Adat Serawai. Customs encapsulate a region's cultural and linguistic aspects, leading to linguistic diversity within local communities. This diversity prompts inquiries into the distinctions and resemblances among these languages. Addressing this inquiry, researchers undertook a mapping study of regional languages in Sumatra.

2. LITERATURE REVIEW

2.1 Comparative-Historical Linguistics

The comparative method was developed and used successfully in the 19th century to reconstruct this parent language, Proto-Indo-European, and has since been applied to the study of other language families (Cho, 2020). With the methods of comparative linguistics, it is possible to chart the phonemic. Changes by which contemporary languages have developed from a common parent language and reconstructed some of the vocabularies of the parent language (Gudschinsky, 1956). Comparative historical linguistic research begins by tracing sound correspondences to see to what degree the set of sound devices in related words are reflected by one proto-language (Mahsun et al., 2017).

The field of study that questions language in the realm of time and changes in language elements that occur within a certain period is called historical-comparative linguistic studies (Kumala & Lauder, 2021). Historical linguistics is concerned with what has been described as “Linguistic Archeology.” Its primary goal is to identify how different languages are related to each other, or as it is often called, the comparative method or comparative reconstruction. Another objective of comparative and historical linguistics is identifying alternative ways languages can be classified or placed in different linguistic typologies (Reagan, 2021).

Theoretically, the languages are expected to exhibit similarities in both form and meaning, particularly in a set of words known as cognates. These cognates are believed to originate from a common proto-language and are considered the ancestors of the respective languages. Comparative-historical linguistics analyzes the diachronic aspect of related languages, identifying general ideas, categories, and elements to show the resemblance and kinship of the researched languages' shared origin (Mohinur, 2023). Keraf (1966) argues that comparative historical linguistics has goals and interests, namely questioning cognate languages by comparing and classifying elements that show kinship in a language family (A'laikum & Ermanto, 2023). The new definition of the "language family" concept has shifted how cognates are defined, identified, and analyzed. The classic comparative-historical method assumes that sound change in spoken languages is regular (Hale, 2014; Nefaa, 2023).

The study utilizes Comparative Historical Linguistics to analyze cognate sets and shared characteristics among related languages, asserting that linguistic similarities are systematic inheritances from a common proto-language, thereby deepening our understanding of linguistic evolution (Dardanila, Widayati, & Gustianingsih, 2023). Comparative linguistics, or comparative-historical linguistics (formerly comparative philology is a branch of historical linguistics concerned with comparing languages to establish their historical relatedness (Zafar, 2024). Comparative historical methods are the primary research tool in restoring (reconstructing) the history of languages. This method primarily focuses on the creation of comparative-historical grammars (covering primarily phonetics) and etymological dictionaries (representing vocabulary) (Allamuratova, 2024).

2.2 Lexicostatistics and Glottochronology

One of the significant contributions of Morris Swadesh to the field of linguistics, and the one with which his name is most often associated, is the lexicostatistic method for determining the time-depth of divergence between related languages, to which he gave the name glottochronology (Troike, 1969). In historical linguistics, lexicostatistics is a quantitative method to estimate the percentage of lexical cognates between languages, and glottochronology is further suggested to calculate the approximate date of separation between two languages (Tao et al., 2023). Lexicostatistics and glottochronology are connected methods that use vocabulary to make historical inferences about language relationships (McMahon & McMahon, 2012). Given linguistic divergence over time, it is possible to estimate the age of linguistic lineages in the same way that biologists use molecular sequence divergence to determine the age of biological lineages (Gray, Atkinson, & Greenhill, 2011).

Lexicostatistics was applied by finding out the cognates of compared languages to obtain the cognate percentage statistically (Keraf, 1996; Crowley & Bowern, 2010; Parmini et al., 2023). In an attempt to do just that, Swadesh developed a historical linguistic approach called lexicostatistics and its derivative 'glottochronology' using the basic vocabulary of the language. McMahon and McMahon argue that quantitative approaches such as lexicostatistics and glottochronology have been widely applied to detect hypothesized genetic relationships among languages (McMahon & McMahon, 2012). Lexicostatistics is historically one of the most widely used and most heavily criticized quantitative approaches in historical-comparative linguistics (Nefaa, 2023; Reagan, 2021). The broad objective of the lexicostatistical method is to determine how distinct languages

are genetically related based on the proportion of cognates they share (Zhang & Gong, 2016).

Table 3 presents a classification of language kinship systems based on time-depth and percentage of cognates based on lexicostatistical and glottochronological calculations. It categorizes languages into different levels, starting from dialects with a time-depth of 0-5 centuries and a cognate percentage of 81-100%, to mesophyla of acrophylum with a time-depth of 100 centuries and above and a cognate percentage of 1% to less than 1%. This classification system helps to understand the relationships between languages at various levels of relatedness and time-depth. It provides a framework for categorizing languages based on their historical connections and the percentage of cognates they share. Researchers can use this classification to study language evolution, historical linguistics, and language preservation efforts.

Table 3
Classification of Language Kinship Systems

Language level	Time-depth (in centuries)	Cognate percentage %
Dialect of language	0-5	81-100 %
Language of family	5-25	36-81 %
Families of stock	25-50	12-36 %
Stock of microphilum	50-75	4-12 %
Microphyla of esophyulum	75-100	1-4 %
Mesophyla of acrophylum	100-and above	1 - less than 1%

Source : (Anayati, Wardana, Mayasari, & Purwarno, 2022)

In statistics, sampling methods are developed using a portion of the sample to represent the population (Wu-Urbanek, 2023). Sampling can also apply to historical linguistics, such as core or basic vocabulary. The following procedure in analyzing the data used formulas for calculating the percentage of kinship (Muhammad & Hendrokumoro, 2022), namely the lexicostatistical formula and calculating the time of separation between languages using the glottochronology formula proposed by Keraf as below :

C = $\frac{Vt}{Vd}$ **x 100 %**

Description:
C : Percentage of language kinship
Vt : Dependent variable
Vd : Basic variable

W = $\frac{log.C}{2\ log.r}$

Description:
W : Separation time (*time in depth*)
C : Percentage of language kinship
r : Retention in 1.000 years, retention 80,5% rounded to 81%
log : Logarithm

Previous research focused on comparing several languages in one province or analyzing phonological aspects only. However, some relatively many studies are relevant to this research, including (1) research by Martius and colleagues entitled "The Analysis of Kindness of Malay Language in Riau Island, Jambi, and Palembang: A Lexicostatistical and Glotocronological Study." The results of this study indicate that The results of this study indicate that Jambi Malay, Palembang Malay, and Riau Islands Malay are in one language family or one lineage with a kinship percentage of above 77% and a range of years of separation from the language since 280 - 555 years ago (Martius et al., 2023). Research

by Dewi Ayu Lestari and colleagues is entitled "Dialectological and Lexicostatistical Study of Serawai Language in Padang Capo Village, South Bengkulu Regency and Puding Village, South Bengkulu Regency." The results also show that the level of similarity of the basic vocabulary of the Serawai language dialect "o" with dialect "au" is 62% kinship words (Mahsun et al., 2017). Additionally, Lilis and Yolanda's research on "Lexicostatistical Studies on Toba Batak and Angkola Mandailing Batak" demonstrates that the kinship between BT and BAM comprises 55% of 150 Swadesh vocabularies, with proto-languages beginning to diverge around 640 AD (calculated from 2020) (Napitupulu & Silaban, 2020).

This study fills the void by observing and analyzing the kinship relations of indigenous languages in two provinces located in the Sumatra archipelago, namely Bengkulu and North Sumatra Provinces. Thus, the effort to renew the topic of this study was oriented to gain deeper insights into the comparison and kinship relations of languages in the wider Sumatra region. Previous studies on "Lexical Kinship Analysis using Lexicostatistics and Glotochronology Methods" have significantly advanced our understanding, yet a literature review exposes a research gap. This research aims to 1) describe the kinship quantitatively-qualitatively between the Rejang Tribe (RT), Serawai Tribe (ST), Lembak Tribe (LT) (Bengkulu), and Batak Toba Tribe (BTT), Batak Mandailing Tribe (BMT), and Batak Nias Tribe (BNT) (North Sumatra) languages; 2) uncovers empirical evidence concerning the timing of the divergence between Rejang Tribe (RT), Serawai Tribe (ST), Lembak Tribe (LT) (Bengkulu) and Batak Toba Tribe (BTT), Batak Mandailing Tribe (BMT), and Batak Nias Tribe (BNT) (North Sumatra) languages; and 3) the grouping of languages and the percentage of kinship between Bengkulu Province and North Sumatra Province.

3. METHOD

3.1 Research Design

This research applied an inductive thinking stage of design, where several linguistic phenomena obtained in the field were analyzed using theories and methods that follow the objectives achieved in this study. This research used a qualitative and quantitative approach, with data collected in two different places. Rejang, Serawai, and Lembak languages are the original languages that are still used daily by native speakers who live in Bengkulu Province, Selebar District, Bengkulu City, Indonesia. The Batak languages of Toba, Mandailing, and Nias are indigenous languages of North Sumatra spoken by Batak tribes living in several places in Medan, North Sumatra. The informants were selected according to the criteria set in the study. The main informants in this study were six people (4 women and 2 men), and supporters were six people. Conditions for determining informants: 1) male or female; 2) between 30-75 years old (not senile); 3) the informant was born and raised in the village and used the language; 4) knowledgeable about the language; 5) can understand Indonesian, and 6) be physically and mentally healthy.

3.1 Data Sources and Data Collection

This research used interview and note-taking techniques to collect data. Interviewing is a fundamental technique of qualitative inquiry. Scholars have described the qualitative interview as central to data collection (Gill et al., 2008). Creswell outlined critical steps for conducting interviews, including selecting interviewees based on study criteria, determining the most suitable interview type to address research questions, and specifying details such

as location, timing, recording method, and participant consent for robust data collection (Virella & Woulfin, 2024). The following is a note-taking technique used to capture data collected from interviews and select data according to research needs (Sugiyono, 2019).

In research, linguists usually perform a collection of roots 100 to 200 words or Swadesh based on small-scale meaning lists (Tao et al., 2023). The interview technique (via telephone) was used to obtain 150 Swadesh vocabulary from the respondents, who are native speakers of the language. In contrast, the tabular recording technique was used to detail and organize the data systematically. This study used qualitative and quantitative data analysis methods (Schoonenboom, 2023).

3.2 Data Analysis

Lexicostatistics prioritizes the statistical observation of words to determine language groupings based on similarities and differences. On the other hand, glottochronology focuses on calculating the time depth or age of kinship languages for grouping in historical linguistics (Makkawaru & Hendrokumoro, 2022). Data analysis employed the comparative historical method, a branch of historical linguistics called comparative linguistics. It compares languages to establish historical connections. Gorys Keraf notes that it facilitates analyzing data across at least two periods, enabling observation of language changes or developments (Candria, 2022). Genetic relatedness implies a common origin or proto-language, and comparative linguistics aims to construct language families, reconstruct proto-languages, and specify the changes that have resulted in the documented languages (Zafar, 2024).

The procedure of this research consists of the following steps: 1) Creation of a data table, which consists of 150 Swadesh vocabularies; 2) Collection of data from sources (native speakers) via telephone to obtain relevant data; 3) analysis and grouping of language pairs based on the data obtained, by classifying based on: a) word pairs that correspond identically, b) word pairs that correspond phonemically, c) phonetically similar word pairs, d) word pairs that differ by one phoneme, 4) Gloss calculation using Lexicostatistical and Grotochronological methods, 5) Calculating the error term to evaluate the accuracy of the data obtained during the translation and data collection process, and 6) Similar vocabulary is marked based on the classification of kinship systems, such as language, family, stock, micropyle, mesophylum, and macrophylum (Anayati et al., 2022).

4. RESULTS

The discussion of this research begins by displaying a list of 150 basic Swadesh vocabularies consisting of language groupings in Bengkulu Province and North Sumatra Province with three pairs of languages that dominate in the province, such as Rejang Tribe (RT), Serawai Tribe (ST) and Lembak Tribe (LT) as well as Batak Toba Tribe (BTT), Batak Mandailing Tribe (BMT) and Batak Nias Tribe (BNT). It indicates the percentage of kinship to the six languages. To get a percentage of these languages using the lexicostatistical method, look for similarities or similarities of lexicon words both in form and meaning. The percentage of kinship in these languages using the lexicostatistical method is explained as follows.

Table 4
Result of Language Classification Calculation

Language/Tribe Name	Kinship Word	Cognate Percentage %	Kinship Relationship
Rejang (RT) – Serawai (ST)	37	25%	Families of stock
Serawai (ST) – Lembak (LT)	78	52%	Language of family
Rejang (RT) – Lembak (LT)	41	27,3%	Families of stock
Toba (BTT) – Mandailing (BMT)	58	39%	Language of family
Mandailing (BMT) – Nias (BNT)	7	5%	Stock of microphilum
Toba (BTT) – Nias (BNT)	6	4%	Microphyla of esophyulum

4.1 The Percentages of Cognate (RT-ST, ST-LT, and RT-LT)

After the calculation using the formula, the results obtained from the percentage of language kinship between RT and ST are 37 pairs of words (24.66%) or rounded up to 25% are related and 113 words that do not have a kinship relationship. In the percentage of language kinship between ST and LT, the results show 78 (52%) pairs of words that are related and 72 (48%) pairs of words that are not connected. Meanwhile, regarding the percentage of language kinship between RT and LT, after comparing the basic vocabulary using the formula, the results show 41 (27.33%) related word pairs and 109 (72.66%) rounded or 73% unrelated word pairs. The 150-word pairs can be found in Table 4, which consists of several categories or kinship criteria, namely:

4.1.1 Identically Correlated Word Pairs

Identically correlated word pairs are word pairs that sound the same and have the same meaning (Rehayati, Hasbi, & Martius, 2023). This means the basic vocabulary compared is built from the same phoneme elements and contains the same meaning. In the language kinship comparison between RT and ST, 17 (11.33%) pairs of words are related identically. In the kinship comparison between ST and LT, 41 (27.33%) were identically related word pairs. Then, the language kinship comparison between RT and LT resulted in 15 pairs of words with a 10% percentage of identical kinship.

Table 5
Identical Word Pairs

No. Data	Gloss	Translation	Language Pairs 1		Language Pairs 2		Language Pairs 3	
			RT	ST	ST	LT	RT	LT
5	*Angin*	Wind	[aŋin]	[aŋin]				
36	*Debu*	Dust	[dəbu]	[dəbu]				
62	*Kelakuan*	Behavior			[pəraŋai]	[pəraŋai]		
65	*Kiri*	Left			[kidau]	[kidau]		
18	*Berdiri*	Stand					[təgak]	[təgak]
26	*Bersih*	Clean					[bərsih]	[bərsih]

4.1.2 Word pairs that have phonetically similar correspondences

There are 8 (5.33%) related word pairs between RT and ST are phonetically similar. In comparing language vocabulary between ST and LT, there are pairs of related words because they are phonetically related to as many as 9 (6%) pairs of words. Meanwhile, regarding word pairs related because they have phonetic similarities in the comparison between RT and LT, there are 6 (4%) word pairs. The phonetic similarity can be seen in the phoneme correspondence of the following word pairs to clarify the meaning of the phonetic similarity. Table 6 shows the correspondence of the phonemes pairs below to clarify the meaning of the phonetic similarity.

Table 6
Phonetically Similar Word Pairs

Gloss	Translation	Phoneme Correspondence	Language Pairs 1		Language Pairs 2		Language Pairs 3	
			RT	ST	ST	RT	ST	ST
Berpikir	Thinking	/e/ ↔ /i/	[peker]	[piker]				
Memegang	Holding	/o/ ↔ /a/	[pegoŋ]	[pegaŋ]				
Asap	Smoke	/ə/ ↔ /a/	[asəp]	[asap]			[asep]	[asap]
Cantik	Beautiful		[aləp]	[alap]			[aləp]	[alap]
Motor	Motorcycle	/u/ ↔ /o/	[mutor]	[motor]				
Menonton	Watching						[nuton]	[noton]
Tajam	Sharp	/e/ ↔ /a/	[tajem]	[tajam]				
Angin	Wind	/i/ ↔ /e/			[aŋin]	[aŋen]	[aŋin]	[aŋen]
Lupa	Forget	/o/ ↔ /e/			[lupo]	[lupe]		
Basah	Wet	/a/↔ /e/			[basah]	[besah]		
Anak perempuan	Girl	/o/ ↔ /ə/			[tino]	[tinə]		

Table 6 examples of word pairs in the three languages, we can see that the phoneme /e/, which is in the middle position of the initial or final syllable in the words thinking, sharp, and hand or [peker], [tajem], and [taŋen], phonetically corresponds to the phonemes /i/ and /a/ in the words [piker], [tajam], and [taŋan]. Then it is phonetically corresponding to the phonemes /o/, /u/, and /ə/ as in the words hold, motorcycle, and smoke or [pegoŋ], [mutor], and [asəp] in RT and ST. Furthermore, the phonemes /e/ and /ə in the final position on the words forget, ears, and girl or [lupe], [teliŋe], and [tinə] correspond phonetically with /o/ on the words [lupo], [teliŋo], and [tino], the phoneme /i/ which is in the penultimate position of the word [aŋin] or wind and [diŋin] or cold, then corresponds phonetically with /e/in words wind and cold or [aŋen] and [deŋen].

Furthermore, the phoneme /a/, which is in the penultimate position in the words wet, lake, and salt or [basah], [danau], and [gaham], corresponds phonetically with the phoneme /e/ in the words [besah], [denau], and [geham] in ST and LT. As for RT and LT, the phonemes /i/ and /e/ in the penultimate position on the words wind, smoke, and rain or [aŋin], [asep] and [ujen] in RT are phonetically equivalent to the phonemes /e/ and /a/ on the words [aŋen], [asap] and [ujan] in LT, then the phoneme /ə/ which is in the penultimate position of the word beautiful and hand [aləp] and [taŋən] in RT is phonetically equivalent to the phoneme /a/ of the word [alap] and [taŋan] in LT. Then, the phoneme /u/ in the penultimate position in the word watching [nuton] is phonetically equivalent to the phoneme /o/ in the word [noton] in LT.

4.1.3 Word pairs related by repeated phoneme correspondences

In repeated word pairs with a kinship relationship in RT and ST languages, only 22 (14.66%) word pairs are related because they correspond regularly. Then, ST and LT language pairs are only 11 (7.33%) word pairs. Meanwhile, comparing basic vocabulary and the percentage of language kinship between RT and LT resulted in 10-word pairs and 6.66% rounded up to 6.7% related to repeated phoneme correspondence. Examples of these word pairs can be seen in the following section of Table 7.

Table 7
Repeated Phoneme Word Pairs

No. Data	Gloss	Translation	Language Pairs 1		Language Pairs 2		Language Pairs 3	
			RT	ST	ST	LT	RT	LT
36	*Debu*	Dust	[dəbu]	[dəbu]	[dəbu]	[dəbu]	[dəbu]	[dəbu]
70	*Licin*	Slippery	[licin]	[licin]	[licin]	[licin]	[licin]	[licin]
111	*Pantai*	Beach	[pantai]	[pantai]			[pantai]	[pantai]
58	*Kanan*	Right			[kanan]	[kanan]		
95	*Mendung*	Cloudy					[mənduŋ]	[mənduŋ]

4.1.4 Word pairs are related because there is only one different phoneme

The comparison of language vocabulary between RT and ST obtained 10 (6.66%), ST and LT 6 (4%), and RT and LT 2 (1.33%) pairs of words that are related because there is only one different phoneme. Examples of these word pairs can be seen in Table 8.

Tabel 8
Different Phoneme Word Pairs

No. Data	Gloss	Translation	Language Pairs 1		Language Pairs 2		Language Pairs 3	
			RT	ST	ST	LT	RT	LT
26	*Bersih*	Clean	[bərsih]	[bərsiah]				
140	*Tidur*	Sleep	[tidu'a]	[tiduak]				
116	*Perut*	Stomach			[peghut]	[pehut]		
133	*Tampar*	Slap			[təpuak]	[təpuk]		
48	*Ikan asin*	Salted Fish					[kan'asin]	[ikan asin]
85	*Membeku*	Frozen					[bəkou]	[bəku]

In the examples in Table 8, it is understood that the phonemes in the words sleep or [tidu'a] and [tiduak] in RT and ST in the word [tidur] are completely zero-matched. The abbreviating mark or apostrophe on the phoneme [tidu'a] is marked to indicate the omission of a word part in a particular context. In the phoneme [tidu'a], the letter /k/ is omitted. In addition to related vocabulary, there is also vocabulary that is not related at all in the comparison between RT and LT, with a total of 113 (75.33%) word pairs. Meanwhile, it is understood that the phonemes /u/ and /m/ that are in the initial position in the words mother and frozen or [umak], [mbəku] in ST and LT have zero correspondence with the phoneme pairs in the words [mak] and [bəku].

Furthermore, the phoneme /g/ in the middle of the word stomach or [peghut] has zero correspondence with the phoneme pair in the word stomach or [pehut] in ST and LT. In addition to having related vocabulary, in the comparison between ST and LT, there are also unrelated vocabulary and no similarity, which amounted to 72 (48%). Besides having related vocabulary, the comparison between RT and LT includes unrelated vocabulary. One hundred nine pairs of words (72.66%) were rounded up, and 73% had no kinship relationship.

In addition to related vocabulary, there is also unrelated vocabulary in the comparison between RT and ST, which found 113 (75.33%); ST and LT also found unrelated vocabulary and have no similarity as much as 72 (48%). RT and LT also include unrelated vocabulary, obtaining 109 pairs of words and (72.66%) rounded or 73% pairs of words. The following examples of these word pairs can be seen in the description in Table 9.

Tabel 9
Unrelated Word Pairs

No. Data	Gloss	Translation	RT	ST	LT
22	*Berkelahi*	Fight	[lago]	[cukuan]	[bəlegeh]
9	*Baik*	Good	[baəs]	[iluak]	[padek]
66	*Kotor*	Dirty	[kutak]	[pekaŋ]	[rəŋai]

Furthermore, it will discuss the percentage of kinship and language grouping in North Sumatra Province with three pairs of languages, which are Batak Toba (SBT), Batak Mandailing (SBM), and Batak Nias (SBN). Only one language pair, SBT and SBM, was found in North Sumatra Province, with 39 pairs of words. SBM and SBN obtained only one pair of words, and then, for SBT and SBN, no kinship relationship was found between 150 words of the language. The following is a presentation of the results of the data obtained from the Province of North Sumatra.

4.2 The Percentage of Cognate (BTT-BMT, BMT-BNT, and BTT-BNT)

The percentage of language kinship between BTT and BMT obtained results showed 58 (38.66%) pairs of words, which rounded up to 39%. Meanwhile, BMT and BNT found 7 (4.6%) or rounded up to 5% kinship percentage, then BTT and BNT found only six pairs of related words or 4% kinship percentage and 92 (61.33%) pairs of unrelated words. The 150 pairs of related words consist of four types of kinship criteria as follows:

4.2.1 Identically correlated word pairs

Identically related word pairs are basic vocabulary words that are compared and built from the same phoneme elements and contain the same meaning. In the language kinship comparison between BTT and BMT, 39 (26%) pairs of words are identically related. The kinship comparison between BTT and BMT, and between BMT and BNT, has 2 (1.33%) identically related word pairs. However, there are no words that are identically related in the BTT and BNT language pairs. The following are examples of identically related word pairs in Table 10.

Table 10
Identically Word Pairs

No. Data	Gloss	Translation	Language Pairs 1		Language Pairs 2	
			BTT	BMT	BMT	BNT
8	Ayam	Chicken	[manuk]	[manuk]		
10	Baju	Clothes	[abit]	[abit]		
34	Danau	Lake			[danau]	[danau]
59	Kaya	Rich			[kayo]	[kayo]

4.2.2 Word pairs that are related because they are phonetically similar

The pairs of related words between BTT and BMT that have phonetic similarity are 2 (1.33%) pairs of words. The comparison of language vocabulary between BMT-BNT and BTT-BNT has similarities, namely only 0.66% or one pair of words that are related because they are phonetically like the word "land," namely [tano] and [tane]. The phonetic similarity in phoneme correspondence in word pairs can be seen in Table 11 below.

Table 11
Phonetically Similar Word Pairs

No. Data	Gloss	Translation	Phoneme Correspondence	Language Pairs 1		Language Pairs 2		Language Pairs 3	
				BTT	BMT	BMT	BNT	BTT	BNT
138	*Telur*	Egg	/o/ ↔ /e/	[tolor]	[telor]				
121	*Sabar*	Patience	/a/ ↔ /o/	[sabar]	[sobar]				
134	*Tanah*	Land	/o/ ↔ /e/			[tano]	[tane]	[tano]	[tane]

Table 11 shows examples of word pairs in the three languages: BTT-BMT, BMT-BNT, and BTT-BNT. We can see that the phoneme /e/ is in the second position at the beginning of the word egg or [telor] phonetically corresponds to the phoneme /o/ in the word [tolor]. Furthermore, the phoneme /a/ is in the second position at the beginning of the word patience, or [sabar] corresponds phonetically to /o/ in the word [sobar]. Then, the phoneme /o/, in the penultimate position in the word [tano] in BTT and BMT, corresponds phonetically with /e/ in the word land or [tane] in BNT.

4.2.3 Word pairs related by repeated phoneme correspondences

In repeated word pairs with a kinship relationship in BTT and BMT languages, only 3 (2%) word pairs are related because they correspond regularly. Then, for BMT and BNT language pairs, there are only 2 (1.33%) word pairs. Meanwhile, comparing word pairs in the language between BTT and BNT did not find pairs of words that corresponded to repeated phonemes. Examples of these word pairs can be seen in the following section of Table 12.

Table 12
Repeated Phoneme Word Pairs

No. Data	Gloss	Translation	Language Pairs 1		Language Pairs 2	
			BTT	BMT	BMT	BNT
71	*Lupa*	Forget	[lupa]	[lupa]		
75	*Manis*	Sweet	[manis]	[manis]		
34	*Danau*	Lake			[danau]	[danau]

BTT, BMT, and BNT are also vocabularies with repeated phoneme correspondences in one language in particular. Examples of words with repeated phonemes can be seen in Table 13.

Tabel 13
Repeated Phoneme

No. Data	Gloss	Translation	BTT	BMT	BNT
143	*Tulang*	Bone	[holi-holi]	[oli-oli]	
16	*Berbisik*	Whispering	[usip-usip]		
14	*Berbaring*	Lying		[guluŋ – guluŋ]	
36	*Debu*	Dusk			[khabwu-khabwu]

4.2.4 Word pairs are related because there is only one different phoneme

The comparison of language vocabulary between BTT and BMT obtained 18 (12%); in BMT and BNT, there are 4 (2.66%) or rounded up 3%, and in BTT and BMT, 6 (4%) pairs of words that are related because there is only one different phoneme. Examples of these word pairs can be seen in Table 14.

Table 14
Different Phoneme Word Pairs

No. Data	Gloss	Translation	Language Pairs 1		Language Pairs 2		Language Pairs 3	
			SBT	SBM	SBM	SBN	SBT	SBN
15	*Berbelok*	Turn	[marbelok]	[mambelok]				
25	*Berpikir*	Thinking	[marpikkir]	[marpikir]				
8	*Ayam*	Chicken			[manuk]	[manu]	[manuk]	[manu]
127	*Sepatu*	Shoes			[sipatu]	[sifatu]	[sipatu]	[sifatu]

The example in Table 14 shows that the word thinking or [marpikkir] phoneme in BTT indicates that two letters are similar or have something in common. In addition to having related vocabulary, comparing language pairs between BTT-BMT, BMT-BNT, and BTT-BNT also includes completely unrelated vocabulary because there are no four criteria for the percentage of kinship in the two languages compared. There are 92 pairs of words, and (61.33%) have no relationship. Examples of these word pairs can be seen in Table 15.

Tabel 15
Unrelated Word Pairs

No. Data	Gloss	Translation	BTT	BMT	BNT
31	Cantik	Beautiful	[bagak]	[deges]	[bagas], [siga]
110	Panas	Hot	[mohop]	[milas]	[awuhkhu]
116	Perut	Stomach	[butuha]	[boltok]	[bhetua']

4.3 Time-depth (W1)

The development of lexicostatistics is the science of glottochronology, which was developed in the 1950s, and it proposes mathematical formulas to determine the timing when two languages separate. Based on the percentage of core vocabulary of culturally independent words (Zafar, 2024). After the percentage of kinship between the two languages is determined using the lexicostatistics formula, the next step is to find the time of separation between the two languages using the following glottochronology formula:

4.3.1 Separation time between RT-ST, ST-LT, and RT-LT

Rejang – Serawai (RT-ST)	Serawai – Lembak (ST-LT)	Rejang – Lembak (RT-LT)
$W = \frac{\log.C}{2\ \log.r}$	$W = \frac{\log.C}{2\ \log.r}$	$W = \frac{\log.C}{2\ \log.r}$
$= \frac{\log.25}{2\ \log.81}$	$= \frac{\log.52}{2\ \log.81}$	$= \frac{\log.27{,}3}{2\ \log.81}$
$= \frac{-0{,}602}{-0{,}183}$	$= \frac{-0{,}283}{-0{,}183}$	$= \frac{-0{,}563}{-0{,}183}$
= 3,289 (x 1000)	= 1,546 (x 1000)	= 3,076 (x 1000)
= 3.289 years	= 1.546 years	= 3.076 years

The separation time is multiplied by 1000, resulting in RT-ST (3,289 years), ST-LT (1,546 years), and RT-LT (3,076 years). So, the calculation of the initial separation time of RT, ST, and LT is shown in the previous calculation. In other words, the calculation of the initial separation time between Rejang, Serawai, and Lembak languages can expressed as follows: (a) Rejang and Serawai languages are estimated to have become one language about 3,289, Serawai and Lembak languages 1,546, and Rejang and Lembak languages 3. 076 years ago; (b) Rejang and Serawai languages are estimated to have separated from their parent languages around 1,266 AD, Serawai and Lembak languages around 477 AD, and Rejang and Lembak languages around 1053 AD (calculated in 2023).

4.3.2 Separation time between BTT-BMT, BMT-BNT, and BTT-BNT

Toba - Mandailing (BTT-BMT)	Mandailing – Nias (BMT-BNT)	Toba – Nias (BTT-BNT)
$W = \frac{\log.C}{2\ \log.r}$	$W = \frac{log.C}{2\ log.r}$	$W = \frac{log.C}{2\ log.r}$
$= \frac{\log.25}{2\ \log.81}$	$= \frac{log.5}{2\ log.81}$	$= \frac{log.4}{2\ log.81}$
$= \frac{-0{,}602}{-0{,}183}$	$= \frac{-0{,}698}{-0{,}183}$	$= \frac{-0{,}602}{-0{,}183}$
= 3,289 (x 1000)	= 0,515 (x 1000)	= 0,419 (x 1000)
= 3.289 years	= 515 years	= 419 years

The separation time is multiplied by 1000, resulting in BTT-BMT (1,408 years), BMT-BNT (515 years), and BTT-BNT (419 years). So, the calculation of the initial separation time of BTT, BMT, and BNT is seen in the previous calculation. In other words, the calculation of the initial separation time between Toba, Mandailing, and Nias languages can stated as follows: (a) Toba and Mandailing Batak languages are estimated to become one language around 1. 408, Mandailing and Nias Batak languages around 515 years, and Toba and Nias Batak languages around 419 years ago; (b) Toba and Mandailing Batak languages are estimated to separate from their parent languages around the 615th century AD, Mandailing and Nias Batak languages around the 1508th century AD, and Mandailing and Nias Batak languages around the 1604th century AD (calculated in 2023).

4.4 Range of Time-depth Error and New Cognate Percentage

The previous calculation is not an exact calculation of the year of separation of the two languages. Therefore, a more specific calculation was carried out to avoid errors in the initial calculation. Thus, the continued statistical technique is still needed. The following technique calculates the error period. In other words, the next step is to calculate the error range to determine a more precise separation time using the following formula.

4.4.1 Calculating the error range between RT-ST, ST-LT, and RT-LT

Rejang – Serawai (RT-ST)	**Serawai – Lembak (ST-LT)**	**Rejang – Lembak (RT-LT)**
Period of error:	**Period of error:**	**Period of error:**
$W_1 - W_2$	$W_1 - W_2$	$W_1 - W_2$
= 3.289 – 3.016	= 1.546 – 1.371	= 3.076 – 2.830
= 273	= 175	= 246

4.4.2 Calculating the error range between BTT-BMT, BMT-BNT, and BTT-BNT

Toba - Mandailing (BTT-BMT)	**Mandailing – Nias (BMT-BNT)**	**Toba – Nias (BTT-BNT)**
Period of error:	**Period of error:**	**Period of error:**
$W_1 - W_2$	$W_1 - W_2$	$W_1 - W_2$
= 1.408 – 2.054	= 515 – 6.672	= 419 – 7.109
= - 646	= - 6157	= - 6.690

The results of the calculation of the error range to determine a more appropriate separation time concluded at the age between Rejang, Serawai, and Lembak languages can be expressed as follows:

1. It is estimated that the single language has formed into one language in RT-ST around (3289 - 3016), ST-LT (1546 - 1371), RT-LT (3076 - 2,830), BTT-BMT (1408-2054), BMT-BNT (515-6672), and BTT-BNT (419-6690) years ago.
2. It is estimated that the languages were one language family between RT-ST about (3289 - 3016), ST-LT (1546 - 1371), RT-LT (3076 - 2,830), BTT-BMT (1408-2054), BMT-BNT (515-6672), and BTT-BNT (419-6690) years ago.
3. It is estimated that it began to separate from Proto in both language pairs around RT-ST (1266 - 993) AD, ST-LT (477- 652) BC, RT-LT (1053-807) BC, BTT-BMT (615-31), BMT-BNT (1508-4649), and BTT-BNT (1604-5086) BC.

4.5 The percentage of Cognate between Bengkulu and North Sumatra Province

After calculating with the lexicostatistical formula to see the results of the percentage of kinship between Bengkulu and North Sumatra Provinces, the results obtained from the percentage of language kinship between Bengkulu and North Sumatra Provinces are 34 pairs of words (22.66%) or rounded up 23% and fall into the category of "Families of stock" which are related to pairs of words that correlate identically, correspond phonetically and differ by one phoneme. The 150-word pairs can be seen in Table 16, which consists of several categories or kinship criteria, namely:

Table 16
Identical Word Pairs Between Bengkulu and North Sumatra Province

No. Data	Gloss	Translation	RT	ST	LT	BTT	BMT	BNT
5	Angin	Wind	[aŋin]	[aŋin]			[aŋin]	
107	Motor	Motorcycle		[motor]	[motor]	[motor]		
34	Danau	Lake		[danau]			[danau]	[danau]
59	Kaya	Rich		[kayo]	[kayo]		[kayo]	[kayo]

5. DISCUSSION

The study of linguistic kinship between Malay and Batak languages on Sumatra Island presents interesting results that significantly contribute to our understanding of language diversity and historical relationships in the region. The results indicate a substantial kinship between the two languages, with a certain percentage of cognate relationships identified between the different language pairs. These findings support the hypothesis that there are common linguistic roots among the languages studied, highlighting these languages' historical relationships and evolution over time.

This research reveals the kinship relationship and age calculations on cognate languages between Rejang, Serawai, and Lembak languages in Bengkulu Province and Toba, Mandailing, and Nias languages in North Sumatra, which are reviewed using a historical-comparative linguistic approach and inter-ethnic relationship based on the history of a common ancestor. After analyzing the six languages in the result section, the following important conclusions can be highlighted. Firstly, the six languages on the island of Sumatra have an arts and humanities genetic relationship, as shown by language kinship relationships supported by historical and sociocultural evidence (Samsudin, 2017; Yanti, 2017; Afria et al., 2021; Rajagukguk & Widayati, 2022; Mulyani & Nasution, 2022; Hendrokumoro et al., 2024).

Secondly, based on the research results, the six languages show significant differences between the Rejang language (RT), Nias language (BNT), and other languages. Rejang language shows a low level of similarity, below 30%, compared to Serawai (ST) or Lembak (LT). On the other hand, the Nias language shows a more striking difference, especially when compared to Toba (BTT) and Mandailing (BMT), where the three language pairs do not even reach a similarity level of 10%.

This research is also an effort to preserve the six original languages in Bengkulu and North Sumatra Provinces, one of which is research on the vitality of the Rejang Language (RT). This study found that the Rejang Language is one of the regional languages originating from Bengkulu Province, which has a script, namely the Kaganga script (Sudarmanto, Taher, & Khanif, 2020). Then, research the dynamics of the Rejang

language's historical origin and the language preservation problems (Asmahasanah, Zulela, & Marini, 2020).

The historical comparative study of the roots of Sumatra Island languages provides new insights into linguistic diversity, inter-ethnic relations, language development, and regional sociolinguistic dynamics. The implications of this study include preserving local languages, strengthening inter-ethnic relations, supporting language development, and exploring the social dynamics that shape the linguistic landscape of the island. By utilizing the results of this study, stakeholders can contribute to the preservation of linguistic diversity, enrich cultural understanding, and encourage sustainable language development initiatives in each of these regions.

The unique findings of this study highlight the importance of preserving linguistic heritage, a deeper understanding of cultural interactions, social factors influencing language development, and language evolution, where languages have experienced shifts and declines in usage due to the influence of external languages entering Bengkulu Province and North Sumatra. In the example of the word 'work' or [kulaghan] in the Serawai language in Bengkulu Province, there is a change in usage by the community, especially by young people, which switches from [kulaghan] to [gawean]. This research shows that lexicostatistical and glottochronological methods are relevant in analyzing languages in one region and can be extended to other areas.

6. CONCLUSION

The study's main conclusions on the linguistic kinship between Malay and Batak languages in Sumatra Island underscore the significant kinship and shared linguistic heritage between these language groups. The research findings highlight the historical connections and evolutionary paths of Malay and Batak languages, revealing a deep-rooted linguistic kinship that traces back to Proto-Austronesian and Proto-Malay origins. Through lexicostatistical and glottochronological analysis, the study quantifies the linguistic similarities and divergence between these languages, providing valuable insights into the linguistic evolution of Sumatra Island. This study's importance and relevance lie in its contribution to historical linguistics and language evolution. The research enhances our understanding of Sumatra's linguistic diversity and cultural heritage by uncovering Malay and Batak's linguistic roots and kinship relationships. The findings offer a nuanced perspective on the shared origins and historical connections between these language groups, enriching our knowledge of language evolution in the region.

However, it is essential to acknowledge the study's limitations, such as focusing on a specific set of languages and needing a more comprehensive sociolinguistic analysis. Future research could expand the analysis scope to include a broader range of languages in Sumatra and delve deeper into the sociocultural factors influencing language development. Additionally, exploring additional linguistic aspects beyond lexicostatistics and glottochronology, such as phonological and morphological features, could provide a more holistic understanding of language divergence and evolution in the region.

Acknowledgment
Not applicable

Availability of Data and Materials
Not applicable

Competing Interests
The authors declare that they have no competing interests.

Funding
The authors receive no financial support for the research, authorship, and or publication of this article.

Authors' Contribution
Riska Meliana worked on the project and the main conceptual ideas. She wrote on the analysis theory, collected, and analyzed data for Bengkulu Province, translated the article, and checked the grammar. Manna Maria Sopiana Manalu collected and analyzed data for North Sumatra Province. Sulis Triyono proofread the manuscript.

Authors' Information
RISKA MELIANA is a student in the Applied Linguistics Study Program at the Faculty of Language, Arts and Culture at Yogyakarta State University. Her research interests include applied linguistics, forensic linguistics, and cultural study.
Email: riskameliana.2022@student.uny.ac.id
MANNA MARIA SOPIANA MANALU is a student in the Applied Linguistics Study Program at the Faculty of Language, Arts and Culture at Yogyakarta State University.
Email: mannamaria.2022@student.uny.ac.id
SULIS TRIYONO is a lecturer at the Applied Linguistics Study Program, Faculty of Languages, Arts, and Culture, Universitas Negeri Yogyakarta. His research interests include Applied Linguistics.
Email: sulis@uny.ac.id; ORCID https://orcid.org/0000-0002-2795-757X

REFERENCES

Adelaar, K. A., Prentice, D. J., Grijns, C. D., Steinhauer, H., & Engelenhoven, A. V. (1996). Malay: Its history, role and spread. In S. A. Wurm, P. Mühlhäusler, & D. T. Tryon (Eds.), *Atlas of Languages of Intercultural Communication in the Pacific, Asia, and the Americas* (pp. 673–694). De Gruyter Mouton. https://doi.org/10.1515/9783110819724.2.673

Afria, R., Izar, J., Anggraini, R. D., & Fitri, D. H. (2021). Analisis Komparatif Bahasa Bengkulu, Rejang, dan Enggano. *Lingua Franca: Jurnal Bahasa, Sastra, dan Pengajarannya*, *5*(1), 1-10. https://doi.org/10.30651/lf.v5i1.4274

A'laikum, A., & Ermanto. (2023). Kekerabatan Bahasa Minangkabau dI Nagari Mungo Kecamatan Luak Kabupaten Lima Puluh Kota dan Bahasa Melayu Riau di Desa Buantan Besar Kecamatan Siak Sri Indrapura Kabupaten Siak. *PERSONA: Language and Literary Studies*, *2*(2), 166–176.

Allamuratova, N. D. (2024). Comparative Linguistics and Translation Studies. *World of Scientific News in Science*, *2*(2), 812–817. https://worldofresearch.ru/index.php/wsjc/article/view/310

Anayati, W., Wardana, M. K., Mayasari, M., & Purwarno, P. (2022). Lexicostatistics of Malay and Malagasy Languages: Comparative Historical Linguistic Study. *English Review: Journal of English Education*, *10*(3), 875–882. https://doi.org/10.25134/erjee.v10i3.6690

Asmahasanah, S., Zulela, & Marini, A. (2020). Dinamika Asal Mula Bahasa Rejang dan Problematika Upaya Pelestarian di Sekolah Dasar Bengkulu Utara. *Prosiding Seminar Nasional Pascasarjana*, *1*, 203–210. Jakarta: Universitas Negeri Jakarta. https://journal.unj.ac.id/unj/index.php/semnas-ps/article/view/16879

Candria, M. (2022). Telaah Linguistik Historis Komparatif terhadap Bahasa Indonesia, Jawa, Madura, dan Bali. *Endogami: Jurnal Ilmiah Kajian Antropologi*, *6*(1), 67–75. https://doi.org/10.14710/endogami.6.1.67-75

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

Collins, J. T. (2019). Global Eras and Language Diversity in Indonesia: Transdisciplinary Projects Towards Language Maintenance and Revitalization. *Paradigma: Jurnal Kajian Budaya*, *9*(2), 103-117. https://doi.org/10.17510/paradigma.v9i2.302

Crowley, T., & Bowern, C. (2010). *An Introduction to Historical Linguistics*. Oxford: Oxford University Press.

Dardanila, Widayati, D., & Gustianingsih. (2024). Language Kindship of Jamee, Gayo, and Malay. *Migration Letters*, *21*(2), 901–912. https://doi.org/10.59670/ml.v21i2.6310

Departemen Pendidikan dan Kebudayaan. (1977). *Adat Istiadat Daerah Bengkulu*. Proyek Penelitian dan Pencatatan Kebudayaan Daerah Departement Pendidikan dan Kebudayaan. Retrieved from https://repositori.kemdikbud.go.id/7688/1/ADAT%20ISTIADAT%20DAERAH%20BENGKULU.pdf

Espree-Conaway, D. A. (2022). *Evolutionary Cartographies of Language Diversification: Quantitative Approaches to the Geolinguistic Mapping of the Kayanic Languages (Central Borneo)* (Theses and Dissertations, The University of North Dakota). The University of North Dakota, Grand Forks, North Dakota.

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of Data Collection in Qualitative Research: Interviews and Focus Groups. *British Dental Journal*, *204*(6), 291–295. https://doi.org/10.1038/bdj.2008.192

Gray, R. D., Atkinson, Q. D., & Greenhill, S. J. (2011). Language Evolution and Human History: What a Difference a Date Makes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1090–1100. https://doi.org/10.1098/rstb.2010.0378

Gudschinsky, S. C. (1956). The ABC's of Lexicostatistics (Glottochronology). WORD, *12*(2), 175–210. https://doi.org/10.1080/00437956.1956.11659599

Hale, M. (2014). The Comparative Method: Theoretical issues. In *The Routledge Handbook of Historical Linguistics*. London: Routledge.

Hendrokumoro, Darman, F., Nuraeni, N., & Ma'shumah, N. K. (2024). The Genetic Relationship Between Alune, Lisabata, Luhu, and Wemale (Western Seram, Indonesia): A Historical-Comparative Linguistics Approach. *Cogent Arts & Humanities*, *11*(1), 2306718. https://doi.org/10.1080/23311983.2024.2306718

Istanti, W., Seinsiani, I. G., Visser, J. G., & Lazuardi, A. I. D. (2020). Comparative Analysis of Verbal Communication Vocabulary Between Indonesian-Afrikaans for Foreign Language Teaching. *International Journal of Language Education*, *4*(3), 389–397. https://doi.org/10.26858/ijole.v4i3.15106

Keraf, G. (1996). *Linguistik Bandingan Historis*. Jakarta: PT. Gramedia Pustaka Utama

Kumala, S. A., & Lauder, M. R. (2021). Makna Toponim di Tangerang sebagai Representasi Keberadaan Etnis Cina Benteng: Sebuah Kajian Linguistik Historis Komparatif. *Ranah: Jurnal Kajian Bahasa*, *10*(2), 304-313. https://doi.org/10.26499/rnh.v10i2.4048

Mahriyuni, Pramuniati, I., & Maftuhah, R. A. (2023). Lexicostatistics of Javanese and Sasak Languages: Comparative Historical Linguistic Studies. *Mimbar Ilmu*, *28*(1), 124–130. https://doi.org/10.23887/mi.v28i1.59797

Mahsun, Fernandez, I. Y., Laksono, K., Lauder, M. R., & Nadra. (2017). *Bahasa dan Peta Bahasa di Indonesia*. Jakarta: Badan Pengembangan dan Pembinaan Bahasa.

Mailani, O., Nuraeni, I., Syakila, S. A., & Lazuardi, J. (2022). Bahasa Sebagai Alat Komunikasi dalam Kehidupan Manusia. *Kampret Journal*, *1*(2), 1–10. https://doi.org/10.35335/kampret.v1i1.8

Makkawaru, & Hendrokumoro. (2022). *The Genetic Relationship between Bugis and Kaili. Journal Educational Verkenning*. *3*(1), 17–27. https://hdpublication.com/index.php/jev/article/download/143/177

Martius, M., Hasbi, M. R., & Rehayati, R. (2023). The Analisis of Kindness of Malay Language in Riau Island, Jambi and Palembang: A the Lexicostatistical and Glotocronological Study. *International Journal of Humanities and Social Science Invention (IJHSSI)*, *12*(8), 1-9. https://doi.org/DOI: 10.35629/7722-12080109

McMahon, A., & McMahon, R. (2012). Lexicostatistics and Glottochronology. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (1st ed.). Wiley. https://doi.org/10.1002/9781405198431.wbeal0701

Mohinur, J. (2023). Theoritical Aspects of Comparative Linguistics. *World Bulletin of Social Sciences*, *22*, 51–55.

Muhammad, S. R., & Hendrokumoro, H. (2022). Hubungan Kekerabatan Bahasa Aceh, Bahasa Devayan, Bahasa Sigulai, dan Bahasa Jamee. *Diglosia: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya*, *5*(4), 897–920. https://doi.org/10.30872/diglosia.v5i4.511

Mulyani, R., & Nasution, K. (2022). Leksikostatistik Bahasa Karo, Bahasa Devayan dan Bahasa Mandailing. *TALENTA Conference Series: Local Wisdom, Social, and Arts (LWSA)*, *5*(2), 158–166. https://doi.org/10.32734/lwsa.v5i1.1341

Napitupulu, L. H., & Silaban, Y. N. (2020). Kajian Leksikostatistik pada Bahasa Batak Toba dan Batak Angkola Mandailing. *Jurnal Bahasa Indonesia Prima (BIP)*, *2*(2), 82–90. https://doi.org/10.34012/bip.v2i2.1336

Nefaa, A. (2023). Genetic relatedness of Tunisian Sign Language and French Sign Language. *Frontiers in Communication*, *8*, 1201148. https://doi.org/10.3389/fcomm.2023.1201148

Omar, A. H., Jaafar, S., & Mat, S. R. C. (2015). Contact of Dialect Clusters: The Malay Peninsula and Sumatra. *Open Journal of Modern Linguistics*, *05*(05), 459–469. https://doi.org/10.4236/ojml.2015.55040

Parmini, N. P., Mawa, I. W., Soper, I. W., Suparta, I. M., Sueni, N. M., & Temaja, I. G. B. W. B. (2023). The Genetic Relationship Between Balinese and Madurese. *International Journal of Education, Vocational and Social Science*, *2*(1), 283–295. https://doi.org/10.99075/ijevss.v2i01.169

Pawley, A. (2007). *Genes, Language, and Culture History in the Southwest Pacific*. 198 Madison Avenue, New York: Oxford University Press.

Pramuniati, I., Mahriyuni, M., & Syarfina, T. (2024). The Vitality of Malay Language in North Sumatra, Indonesia. *Research Journal in Advanced Humanities*, *5*(1). 223-245 https://doi.org/10.58256/h2vqkg98

Rahayu, N. (2018). Studi Awal Sebaran Bahasa-Bahasa Etnik di Provinsi Bengkulu. *Wacana: Jurnal Penelitian Bahasa, Sastra, dan Pengajarannya*, *16*(1), 26–35. https://doi.org/10.33369/jwacana.v16i1.6693

Rajagukguk, D. L., & Widayati, D. (2022). Relationship of Batak Karo, Batak Toba, and Nias Comparative Historical Linguistic Study. *International Journal of Humanities Education and Social Sciences (IJHESS)*, *1*(6), 1037-1046. https://doi.org/10.55227/ijhess.v1i6.192

Reagan, T. (2021). Historical Linguistics and the Case for Sign Language Families. *Sign Language Studies*, *21*(4), 427–454. https://doi.org/10.1353/sls.2021.0006

Rehayati, R., Hasbi, M. R., & Martius, M. (2023). An Exploration of Local Wisdom: Rites of Passage in Malay Culture in Riau and Palembang. *International Journal of Social Science Research and Review*, *6*(9), 106–117. https://doi.org/10.47814/ijssrr.v6i9.1476

Ross, M. (2020). Narrative Historical Linguistics: Linguistic Evidence for Human (Pre)history. In R. D. Janda, B. D. Joseph, & B. S. Vance (Eds.), *The Handbook of Historical Linguistics* (1st ed., pp. 468–499). Wiley. https://doi.org/10.1002/9781118732168.ch22

Samsudin. (2017). *Sosiologi Perkotaan: Studi Perubahan Sosial dan Budaya (Bengkulu)*. Bengkulu: Pustaka Pelajar & IAIN Bengkulu Press. Retrieved from http://repository.iainbengkulu.ac.id/5115/

Schoonenboom, J. (2023). The Fundamental Difference Between Qualitative and Quantitative Data in Mixed Methods Research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *24*(1), Art. 1. https://doi.org/10.17169/FQS-24.1.3986

Setiawan, L. G. I. P. S. (2020). Hubungan Kekerabatan Bahasa Bali dan Sasak dalam Ekoleksikon Kenyiuran: Analisis Linguistik Historis Komparatif. *Jurnal Inovasi Penelitian*, *1*(1), 27–30. https://doi.org/10.47492/jip.v1i1.44

Simanjuntak, T. (2015). Progres Penelitian Austronesia di Nusantara. *AMERTA, Jurnal Penelitian dan Pengembangan Arkeologi* *33*(1), 25-44. https://doi.org/10.24832/amt.v33i1.211

Sudarmanto, P. M. S., Taher, M. D. S., & Khanif, A. (2020). Vitalitas Bahasa Rejang: Pergeseran dan Pemertahanan Bahasa Daerah pada Suku Rejang Bengkulu. *PKM : Penelitian Sosial & Humaniora*, 304. https://simbelmawa.kemdikbud.go.id/prosiding/pkm/article/view/413

Sugiyono, S. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.

Tao, Y., Wei, Y., Ge, J., Pan, Y., Wang, W., Bi, Q., … Zhang, M. (2023). Phylogenetic Evidence Reveals Early Kra-Dai Divergence and Dispersal in the Late Holocene. *Nature Communications*, *14*(1), 6924. https://doi.org/10.1038/s41467-023-42761-x

Tarigan, B. (2016). *Kebertahanan dan Ketergeseran Leksikon Flora Bahasa Karo: Kajian Ekolinguistik* (Dissertation, Universitas Sumatera Utara). Universitas Sumatera Utara, Medan. Retrieved from http://sipus.usu.ac.id/opac2.2/buku/134231/Kebertahanan-dan-ketergeseran-leksikon-flora-Bahasa-Karo-:-kajian-ekolinguistik.html

Troike, R. C. (1969). The Glottochronology of Six Turkic Languages. *International Journal of American Linguistics*, *35*(2), 183–191. https://doi.org/10.1086/465053

Tryon, D. (2006). Proto-Austronesian and the Major Austronesian Subgroups. In P. Bellwood, J. J. Fox, & D. Tryon (Eds.), *The Austronesians: Historical and Comparative Perspectives* (1st ed.). ANU Press. https://doi.org/10.22459/A.09.2006.02

Virella, P., & Woulfin, S. (2024). Tell Me About Your Trauma: An Empathetic Approach-Based Protocol for Interviewing School Leaders Who Have Experienced a Crisis. *Qualitative Research Journal*. https://doi.org/10.1108/QRJ-09-2022-0121

Wahyuni, D., Maulina, Y., Mulia, A., & Sunardi, S. (2021). Cultural Discourse in Reading Texts of Indonesian Language Proficiency Test. *International Journal of Language Education*, *5*(4), 356-371. https://doi.org/10.26858/ijole.v5i4.23590

Wu-Urbanek, M.-S. (2023). *A Computer-Assisted Approach to the Comparison of Mainland Southeast Asian Languages*. Thuringia: Friedrich-Schiller-Universität Jena.

Yanti, N. (2017). Hubungan Kekerabatan Bahasa Rejang, Serawai, dan Pasemah dengan Menggunakan Teknik Leksikostatistik. *GENTA BAHTERA: Jurnal Ilmiah Kebahasaan Dan Kesastraan*, *3*(2), 178–189. https://doi.org/10.47269/gb.v3i2.14

Zabadi, F., Darmawati, B., Wahyuni, D., Winahyu, S. K., Lestaningsih, D. N., & Abduh, A. (2023). Revealing the Kafoa Language Vitality through the Basic Cultural Vocabulary Mastery: Implications for Language Education. *International Journal of Language Education*, *7*(4), 686–701. https://doi.org/10.26858/ijole.v7i4.53017

Zafar, D. (2023). Analysis the Studies into Comparative Linguistics. *European Journal of Artificial Intelligence and Digital Economy*, *1*(2), 19–22. https://doi.org/10.61796/jaide.v1i2.238

Zein, S. (2020). *Language Policy in Superdiverse Indonesia* (1st edition). London ; New York: Routledge.

Zhang, M., & Gong, T. (2016). How Many Is Enough?—Statistical Principles for Lexicostatistics. *Frontiers in Psychology*, *7*, 1916. https://doi.org/10.3389/fpsyg.2016.01916