

## MAKASSAR LANGUAGE AND SELAYAR LANGUAGE COMPARATIVE HISTORICAL COMPARATIVE LINGUISTIC REVIEW: LEXICOSTATISTICS

Ilmi Solihat<sup>1)</sup>, Muldawati<sup>2)</sup>

<sup>1,2</sup>Sultan Ageng Tirtayasa University

<sup>1,2</sup>Jl. Ciwaru Raya, Cipare, Serang District, Serang City, Banten 42117

<sup>1</sup>Email: [ilmisolihat@untirta.ac.id](mailto:ilmisolihat@untirta.ac.id)

<sup>2</sup>Email: [7771230002@untirta.ac.id](mailto:7771230002@untirta.ac.id)

### Abstrak

*Bahasa Makassar dan bahasa Selayar diasumsikan memiliki kesenjangan dalam kekerabatan. Sedangkan, dalam realitas kedua bahasa tersebut tidak sepenuhnya identik. Tujuan penelitian yang menganalisis bahasa Makassar dan bahasa Selayar dari tingkat kekerabatan dan waktu pisah, faktor yang mempengaruhi, dan keterbatasannya. Peneliti menggunakan mixed methods. Metode kuantitatif melalui teknik leksikostatistik dan teknik grotokronologi, sedangkan metode kualitatif digunakan untuk teknik pengumpulan data yang menggunakan teknik simak, catat, dan rekam serta wawancara terhadap informan. Sumber primer berupa informan penutur MS berumur 60 tahun, DM berumur 58 tahun, dan berumur 50 tahun, dari informan tersebut memvalidasi dari sumber sekunder yaitu daftar kosakata Swadesh dan Kamus bahasa Makassar dan Selayar. Analisis data dalam penelitian ini dengan menerapkan langkah-langkah yang harus digunakan dalam penelitian historis komparatif. Hasil kekerabatan 64,5 % tingkat kekerabatan masuk dalam kategori tingkatan bahasa keluarga (family). Sedangkan, hasil perhitungan waktu pisah dalam abad masuk dalam kategori tingkatan Dialek Bahasa (Dialect of language) karena waktu pisah 360 tahun pada klasifikasi 0-5 abad. Faktor yang mempengaruhi kekerabatan seperti asal usul dan sejarah, pengaruh budaya, perbedaan struktural, penggunaan dan pemertahanan bahasa. Keterbatasan memerlukan pertimbangan seperti peminjaman bahasa, analisis yang cermat terhadap kata serumpun dan pertimbangan kontak bahasa historis, serta perubahan bahasa.*

**Kata Kunci:** linguistik historis komparatif, bahasa makassar, bahasa selayar

### Abstract

Makassar and Selayar languages are assumed to have a gap in kinship. Whereas, in reality, the two languages are not completely identical. The purpose of the study is to analyze Makassar and Selayar languages from the level of kinship and time of separation, influencing factors, and their limitations. The researcher uses mixed methods. Quantitative methods through lexicostatistics techniques and grotokronology techniques, while qualitative methods are used for data collection techniques using listening, recording, and recording techniques and interviews with informants. Primary sources in the form of MS speaker informants aged 60 years, DM aged 58 years, and aged 50 years, from these informants validated from secondary sources, namely the Swadesh vocabulary list and the Makassar and Selayar language dictionaries. Data analysis in this study by applying the steps that must be used in comparative historical research. The results of kinship 64.5% of the kinship level falls into the category of family language levels (family). Mean while, the results of calculating the separation time in centuries are included in the category of Language Dialect levels. (Dialect of language) because the separation time is 360 years in the 0-5 century classification. Factors that influence kinship such as origin and history, cultural influences, structural differences, language use and maintenance. Limitations require considerations such as language borrowing, careful analysis of cognates and consideration of historical language contact, as well as language change.

**Keywords:** comparative historical linguistics, makassar language, selayar language

## 1. INTRODUCTION

Austronesian languages share a common ancestor, the Proto-Austronesian language spoken thousands of years ago. Over time, Proto-Austronesian split into several sublanguages due to geographical separation, migration, and language contact. According to recent research, “the Austronesian language family is the second largest language family in the world in terms of number of languages, and was the most geographically extensive before European colonial expansion in the last five centuries” (Blust & Trussel, 2013). Until a comparative method was developed to analyze sound correspondences, grammatical features, and shared vocabulary (cognates) in Austronesian languages by identifying systematic patterns of change, linguists were able to reconstruct the ancestral language and group related languages into subfamilies (Blust, 2018).

One of the studies is using the lexicostatistic method using a standard word list or swadesh word list to calculate the percentage of cognates between languages. The spread of Austronesian languages has spread widely in various regions, one of which is in South Sulawesi, Indonesia. Recent studies show that “quantitative analysis using lexicostatistical techniques can produce the highest percentage of 68% between cognates in a group of closely related language families” (Anggayana et al., 2020). A higher percentage indicates a closer relationship and a shorter time since divergence.

The Makassar language, which belongs to the Austronesian language group, is widely used by the Makassar people who live in various regions in South Sulawesi, Indonesia. The area includes Gowa, Sinjai, Maros, Takalar, Jeneponto, Bantaeng, Pangkajene and the Islands, Bulukumba, the Selayar Islands, and Makassar City. Contemporary research confirms that “the Makassar language is an Austronesian language belonging to the West Malayo-Polynesian subgroup in the South Sulawesi family spoken by

around two million people in the province of South Sulawesi” (Jukes, 2021). The Makassar language is categorized as an Austronesian language belonging to the Malayo-Polynesian subgroup of the South Sulawesi branch. Specifically, it is under the Makassar or Makassarik group which includes the Konjo language, including the Mountain and Coastal varieties, and the Selayar language.

In linguistic discussions, Konjo and Selayar are generally seen as variations or dialects of Makassar, illustrating the linguistic relationships in the region. Along with Bugis, Mandar, and Sa'dan (Toraja), Makassar has close linguistic ties within the South Sulawesi language group. A recent study on the empowerment of the Makassar language shows that “Makassar influences Indonesian, especially in morphological aspects, namely the presence of clitics and time markers” (Abbas, 2021).

From a lexical point of view, the Makassar language group stands out as the most distinctive among the languages spoken in South Sulawesi. The average percentage of vocabulary similarity between the Makassar language group and other South Sulawesi languages is around 43% (Charles & Barbara, 1987). The Gowa or Lakiung dialect shows the highest level of divergence with vocabulary similarity around 5-10 percentage points lower than the Konjo and Selayar languages when compared to other South Sulawesi languages. However, research led by linguist Ulo Sirk found a higher percentage of vocabulary similarity ( $\geq 60\%$ ) between Makassar and other South Sulawesi languages. These quantitative data are in line with qualitative assessments that place Makassar as an integral part of the South Sulawesi language group (Sirk, 1989).

Selayar Regency is located in South Sulawesi Province as one of 24 regencies that stretch from the northern tip to the southern tip of Sulawesi Island. This special regency offers a unique language that distinguishes it from other

administrative areas in South Sulawesi. The local language used in the Selayar language known as the Selayar language belongs to the Austronesian language family and functions as the mother tongue on Selayar Island and various surrounding islands in the Selayar Islands Regency, South Sulawesi, Indonesia. Recent research explains that “the Selayar language is a Malayo-Polynesian language spoken by about 100,000 people on Selayar Island in the province of South Sulawesi, Indonesia” (Finer & Klemme, 2024).

The linguistic landscape of Selayar language bears traces of Malay, Makassarese, and Bugis, indicating a rich tapestry of influences. In addition, there is a noteworthy linguistic relationship between Selayar and the Coastal Konjo language, prevalent in Ujung Loe Regency and Bulukumba Regency in South Sulawesi Province, Indonesia. Language interactions that underscore the diverse and culturally vibrant linguistic ecosystem that exists in this region.

Makassar language is identified with kinship with its closest areas around Makassar, as well as Selayar language which is identical to Konjo language. As if the two languages have a gap in kinship. Making researchers in the field of language believe that the assumption is true. Meanwhile, in reality, both languages are used by their speakers in everyday life, identical glosses are often found, differences in sound correspondence from the vowel and consonant levels. From these findings, this research is increasingly crucial to prove and find assumptions that are not in line with reality.

Relevant research related to the study of analyzing kinship and separation time such as the research of (Zulham et al., 2022) entitled “Kinship of Makassar and Selayar Languages: Lexicostatistical and Glotochronological Analysis” obtained the results of the kinship of the two languages 60% and a separation time of 378 years. However, in this study there were no other factors that influenced the kinship, except for phonological, morphological, and syntactic

factors. Further research by (Sulistiyaning & Hendrokumoro, 2023) entitled “Kinship Relations of Javanese, Sundanese, and Makassar Languages: Comparative Historical Linguistic Study” obtained the results of the kinship of Javanese – Makassar 20% with a separation time between 2131 BC – 1463 BC, and Sundanese – Makassar 16% with a separation time of 2737 BC – 1917 BC in the study did not put forward the factors and limitations of considering kinship between languages. Then, research by (Dian Astuti, 2023) entitled “Kinship Relations between the Konjo Dialect and the Lakiung Dialect of the Makassar Language: A Dialectological Approach” obtained a kinship result of 51.5% with a discussion of the differences in phonology and lexicon.

From previous research, this study will analyze from linguistic factors alone, but also from other factors such as history, culture, structure, language use, and language maintenance as well as the limitations of considering kinship between languages. Recent comparative studies emphasize the importance of “using appropriate statistical principles for lexicostatistics in comparing languages based on the proportion of cognates in standard basic vocabulary lists” (Zhang & Gong, 2016).

The purpose of this study is to analyze the Makassar language and the Selayar language from the level of kinship and separation time and discuss the factors that influence kinship to the stage of its limitations. Until the percentage and year of the two languages are found in the current year search. This study is also a significant update from previous studies, presenting the latest phenomena or information that can improve and strengthen previous research. Until the discovery of a language that is assumed not to have a strong kinship, it will now be tested by comparing it directly with lexicostatistics techniques. Lexicostatistics techniques for statistical calculations with the help of a swadesh

vocabulary list and validation on respondents strengthen the results of this study.

The field of comparative linguistics continues to make use of lexicostatistics, a powerful method for uncovering relationships between languages. By analyzing shared vocabulary, lexicostatistics allows researchers to reconstruct language families, estimate divergence times, and identify potential zones of language contact. A core tenet of lexicostatistics remains the assumption that languages with a common ancestor share a core vocabulary through inheritance. This core vocabulary, or cognates, are less susceptible to borrowing or substitution than culturally specific terms (Bouckaert et al., 2018). Recent years have seen advances in lexicostatistical methods. Standardized word lists continue to be used, but researchers are applying more sophisticated statistical techniques to account for potential similarities and refine estimates of divergence times (Chang et al., 2021). The word lists in question are known as swadesh word lists.

## 2. LITERATURE STUDY

Language grouping techniques usually prioritize statistical analysis of words (lexicon) and then attempt to determine groupings based on the percentage of similarities and differences between one language and another, called lexicostatistics. The purpose of this method is to determine the percentage of relatedness between two or more languages. The lexicostatistics approach is used to determine the degree of relatedness between two languages. A contemporary definition asserts that “lexicostatistics is a comparative linguistic method that involves comparing the percentage of lexical cognates between languages to determine their relationship” (Hoffmann et al., 2021).

Lexicostatistics in question relates to a method of language categorization that places emphasis on the statistical evaluation of the lexicon, then categorizing it according to the extent of the

percentage of similarities and differences observed between different languages as stated by (Keraf, 1996) that this technique involves a systematic approach to analyzing linguistic data with a focus on word structures and patterns to determine relationships and differences between languages that ultimately lead to a more comprehensive understanding of language evolution and diversity, to the stage of calculating the age of language kinship. Recent research explains that “lexicostatistics compares languages for phylogenetic affinity based on the proportion of cognate words in a standard basic vocabulary list, where each slot in the list is a concept (meaning), and the collected items (words) that occupy the same slot are compared across linguistics” (McMahon & McMahon, 2021).

Vocabulary comparisons are essential in the field of linguistics because they play a vital role in identifying and establishing the degree of lexical similarity that exists between two different languages, a process that aids linguists in the analysis of linguistic structures and patterns. (Crowley, 1992) highlights the importance of such comparisons in explaining the complexity and nuances of language systems, thereby contributing to a broader understanding of language evolution and diversity. Researchers use this lexicostatistical method to try to determine exactly how old a language is, when it emerged, and how it relates to its relatives.

This is in line with the opinion of Grimes (Ino, 2015) lexicostatistics a method used in linguistics involves grouping languages or dialects based on statistical analysis that places emphasis on quantitative calculations to determine the extent of kinship that exists between them. This technique aims to identify the relative frequency of shared lexical items across different linguistic entities. Recent studies have shown that “lexicostatistics as a method of language classification has certain limitations but remains useful when applied with appropriate statistical

principles and considering the possibility of coincidental similarities” (Hoffmann et al., 2021).

Lexicostatistics, a linguistic method used to compare languages, consists of compiling a basic vocabulary for each language being examined. In this field, Morris Swadesh introduced 200 universal basic vocabularies covering various categories such as pronouns, numbers, body parts (including properties and activities), elements related to nature and its surroundings, and everyday cultural objects. The application of lexicostatistics in language classification is widely adopted by linguists around the world because of its effectiveness.

This method relies on the use of numerical data in statistics as a basis for its organization and analysis. The main goal of this approach is to establish relationships between different languages by identifying similarities in the lexicon, thus shedding light on potential relationships and linguistic affiliations. Through the examination of shared vocabulary items, lexicostatistics aims to uncover underlying patterns and correlations between two or more languages, facilitating a deeper understanding of their historical and structural relationships. Contemporary research asserts that “modern lexicostatistical methods employ more sophisticated statistical techniques to account for possible similarities and refine estimates of divergence times” (Rama et al., 2022)

Highlighted the advantages of the lexicostatistical method, such as the rapid identification of linguistic relationships through basic vocabulary lists, grouping of related languages, and initial language classification. In addition, outlined three main assumptions of the lexicostatistical method, including simultaneous replacement of basic vocabulary in all languages, the possibility of changes in all basic vocabulary at the same time, and the existence of a common basic vocabulary in all languages throughout the world.

Here are the steps of the lexicostatistics technique and the grottochronology technique as follows.

#### 1) Collecting basic vocabulary

When comparing two or more languages, the most important factor is to create a list of basic vocabulary of the language. The list that is often used in this method is the list compiled by Morris Swadesh which contains 200 words. Data validated by Makassar and Selayar informants are entered into a table along with 200 basic vocabulary of each language. These words do not refer to a particular language level, but are based on the formulation of comparative language by grouping 200 Morris Swadeshi words that sound and mean similar things.

#### 2) Selecting words to be used as research data from each language

The words that have been registered are then selected for the purpose of finding terms that can serve as research materials. This selection process is carried out by considering the criteria that the selected words show similarities or are identical in both languages being compared so that the words selected for analysis show a level of equivalence or correlation between the two contrasting languages.

#### 3) Calculating the degree of language kinship

Constant or index using Morris Swadesh’s proposal, which is 81% (Keraf, 1996). To calculate the percentage of kin words, the following formula is used (Keraf, 1996).

$$\frac{Kognat \times 100\%}{Gloss/vocabulary}$$



**Table 1 Classification of Language Groupings according to Crowley (1992:179)**

Language Level	Pissat downRelatives Words	Time Separated in Centuries
Language		
Dialects ( <i>Dialect of language</i> )	100 – 81	0 – 5
Family ( <i>family</i> )	81 – 36	5 – 25
Clump ( <i>stock</i> )	36 – 12	25 – 50
Microfilum	12 – 4	50 – 75
Meosophilum	4 – 1	75 – 100
Macrophyllum	< 1 %	100.

#### 4) Calculating language separation time

The time of separation between two related languages for which the proportion of related words is known can be calculated using the following formula (Keraf, 1996).

$$t = \frac{\log c}{2 \log r}$$

Information:

t = time of separation in thousands (millennia) years ago

r = retention or constant percentage in 1000 or also called index

c = percentage of relatives

log = logarithma

#### 5) Calculating the age of a language

To calculate the error term, a standard error is used which is usually 70% of the true estimate. The standard error is calculated using the following formula (Keraf, 1996).

$$S = \sqrt{\frac{c(1-c)}{n}}$$

Information:

S = standard error in the percentage of relative words

c = percentage of relative words

n = number of words compared (both relatives and non-relative)

### 3. METHOD

Researchers use *mixed methods*. Quantitative methods through lexicostatistics techniques and glottochronology techniques. The quantitative method of comparative historical linguistics uses a comparative historical linguistics approach. This approach was introduced by American linguist Morris Swadesh in the late 1940s. The primary sources in this study were informants who were native speakers of Makassar named MS aged 60 years, native speakers of Selayar named DM aged 58 years, and native speakers who understood Makassar and Selayar named BR aged 50 years, from these informants validating from secondary sources, namely the Swadesh vocabulary list and the Makassar and Selayar language dictionary. Qualitative methods were used for data collection techniques that used listening, note-taking, and recording techniques as well as interviews with informants. Data analysis in this study by applying the steps that must be used in comparative historical research, namely (1) Collecting basic vocabulary, (2) Selecting words that will be used as research data from each language, (3) Calculating the level of language kinship, (4) Calculating the time of language separation, and (5) Calculating the age of the language.

### 4. RESULTS AND DISCUSSION

List of 200 Lexicostatistics of Morris Swadesh vocabulary in Makassar and Selayar languages, the following shows the data found by researchers presented in the following table.

**Table 2 Consanguineous Pairs (*Cognate*) Identical to Makassar and Selayar languages**

No	Gloss	Makassar	Selayar
1.	Dia	<i>Ia</i>	<i>Ia</i>
2.	Kalian	<i>Ikau</i>	<i>Ikau</i>
3.	Siapa	<i>Inai</i>	<i>Inai</i>
4.	Apa	<i>Apa</i>	<i>Apa</i>
5.	Berapa	<i>Sikura</i>	<i>Sikura</i>
6.	Lain	<i>Maraeng</i>	<i>Maraeng</i>
7.	Dua	<i>Rua</i>	<i>Rua</i>
8.	Tiga	<i>Tallu</i>	<i>Tallu</i>
9.	Lima	<i>Lima</i>	<i>Lima</i>
10.	Pendek	<i>Bodo</i>	<i>Bodo</i>
11.	Manusia	<i>Tau</i>	<i>Tau</i>
12.	Anak	<i>Anak</i>	<i>Anak</i>
13.	Burung	<i>Jangang-jangang</i>	<i>Jangang-jangang</i>
14.	Kutu	<i>Kutu</i>	<i>Kutu</i>
15.	Cacing	<i>Gallang-gallang</i>	<i>Gallang-gallang</i>
16.	Ranting	<i>Tangke</i>	<i>Tangke</i>
17.	Biji	<i>Batu</i>	<i>Batu</i>
18.	Bunga	<i>Bunga</i>	<i>Bunga</i>
19.	Daging	<i>Assi</i>	<i>Assi</i>
20.	Tulang	<i>Buku</i>	<i>Buku</i>
21.	Bulu	<i>Bulu</i>	<i>Bulu</i>
22.	Kepala	<i>Ulu</i>	<i>Ulu</i>
23.	Telinga	<i>Toli</i>	<i>Toli</i>
24.	Mata	<i>Mata</i>	<i>Mata</i>
25.	Gigi	<i>Gigi</i>	<i>Gigi</i>
26.	Lidah	<i>Lila</i>	<i>Lila</i>
27.	Kuku	<i>Kanuku</i>	<i>Kanuku</i>
28.	Kaki	<i>Bangkeng</i>	<i>Bangkeng</i>
29.	Tangan	<i>Lima</i>	<i>Lima</i>
30.	Leher	<i>Kallong</i>	<i>Kallong</i>
31.	Hati	<i>Ate</i>	<i>Ate</i>
32.	Minum (Meminum)	<i>Annginung</i>	<i>Annginung</i>
33.	Makan (Memakan)	<i>Annganre</i>	<i>Annganre</i>
34.	Gigit (Mengigit)	<i>Angngokkok</i>	<i>Angngokkok</i>
35.	Tidur	<i>Tinro</i>	<i>Tinro</i>
36.	Mati	<i>Mate</i>	<i>Mate</i>
37.	Bunuh (Membunuh)	<i>Ammuno</i>	<i>Ammuno</i>
38.	Potong (Memotong)	<i>Ammolong</i>	<i>Ammolong</i>
39.	Datang	<i>Battu</i>	<i>Battu</i>
40.	Diri (Berdiri)	<i>Ammenteng</i>	<i>Ammenteng</i>
41.	Beri (Memberi)	<i>Assare</i>	<i>Assare</i>
42.	Cuci (Mencuci)	<i>Assassa</i>	<i>Assassa</i>
43.	Dorong (Mendorong)	<i>Annyorong</i>	<i>Annyorong</i>
44.	Nyanyi (Bernyanyi)	<i>Akkelong</i>	<i>Akkelong</i>
45.	Matahari	<i>Mataallo</i>	<i>Mataallo</i>
46.	Bulan	<i>Bulang</i>	<i>Bulang</i>

47.	Bintang	<i>Bintoeng</i>	<i>Bintoeng</i>
48.	Hujan	<i>Bosi</i>	<i>Bosi</i>
49.	Laut	<i>Tamparang</i>	<i>Tamparang</i>
50.	Batu	<i>Batu</i>	<i>Batu</i>
51.	Tanah	<i>Lino</i>	<i>Lino</i>
52.	Angin	<i>Anging</i>	<i>Anging</i>
53.	Bakar (Membakar)	<i>Tunu</i>	<i>Tunu</i>
54.	Merah	<i>Eja</i>	<i>Eja</i>
55.	Hijau	<i>Moncong</i>	<i>Moncong</i>
56.	Kuning	<i>Didi</i>	<i>Didi</i>
57.	Malam	<i>Bangngi</i>	<i>Bangngi</i>
58.	Hari	<i>Allo</i>	<i>Allo</i>
59.	Tahun	<i>Taung</i>	<i>Taung</i>
60.	Dingin	<i>Dinging</i>	<i>Dinging</i>
61.	Penuh	<i>Rassi</i>	<i>Rassi</i>
62.	Lama	<i>Sallo</i>	<i>Sallo</i>
63.	Tajam	<i>Tarang</i>	<i>Tarang</i>
64.	Basah	<i>Basa</i>	<i>Basa</i>
65.	Dekat	<i>Ambani</i>	<i>Ambani</i>
66.	Kanan	<i>Kanang</i>	<i>Kanang</i>
67.	Kiri	<i>Kairi</i>	<i>Kairi</i>
68.	Di	<i>Ri</i>	<i>Ri</i>
69.	Dalam	<i>Lalang</i>	<i>Lalang</i>
70.	Dan	<i>Na</i>	<i>Na</i>
71.	Nama	<i>Areng</i>	<i>Areng</i>

**Table 3 Consanguineous Pairs (Cognate) Phonemic Correspondence of Makassar and Selayar Languages**

	Gloss	Makassar	Selayar
1.	Aku	<i>Inakke</i>	<i>Nakke</i>
2.	Kami	<i>Ikambe</i>	<i>Kambe</i>
3.	Kamu	<i>Ikau</i>	<i>Kau</i>
4.	Satu	<i>Sekre</i>	<i>Se're</i>
5.	Empat	<i>Appak</i>	<i>Appa'</i>
6.	Lebar	<i>Labbak</i>	<i>La'ba'</i>
7.	Tebal	<i>Kapalak</i>	<i>Ka~pala</i>
8.	Tipis	<i>Nipisik</i>	<i>Ni~pisi</i>
9.	Laki-laki	<i>Burakne</i>	<i>Tuburakne</i>
10.	Istri	<i>Baine</i>	<i>Bahine</i>
11.	Suami	<i>Burakne</i>	<i>Bura'ne</i>
12.	Binatang	<i>Olok-olok</i>	<i>Olo-olo</i>
13.	Ikan	<i>Jukuk</i>	<i>Juku'</i>
14.	Akar	<i>Akak</i>	<i>Aka'</i>
15.	Rumput	<i>Rukuk</i>	<i>Ru~ku'</i>
16.	Kulit	<i>Bukkuleng</i>	<i>Balulang</i>
17.	Tanduk	<i>Tanruk</i>	<i>Tandru'</i>
18.	Rambut	<i>Uk</i>	<i>Uhu'</i>



19.	Hidung	<i>Kakmuru</i>	<i>Ka'muru</i>
20.	Tungkai	<i>Palak bangkeng</i>	<i>Pala' bangkeng</i>
21.	Isap (Mengisap)	<i>Akngisuk</i>	<i>Angnginsu'</i>
22.	Ludah (Meludah)	<i>Appikru</i>	<i>Appe'ru</i>
23.	Tiup (Meniup)	<i>Attuik</i>	<i>Annui</i>
24.	Napas (Bernapas)	<i>Akmaik</i>	<i>A'mai</i>
25.	Dengar (Mendengar)	<i>Allanngerek</i>	<i>Allangngere</i>
26.	Tahu (Mengetahui)	<i>Anngisseng</i>	<i>Angngisse</i>
27.	Pikir (Berpikir)	<i>Appikkirik</i>	<i>Appikkiri</i>
28.	Cium (Mencium)	<i>Angngarak</i>	<i>Angngara'</i>
29.	Takut	<i>Mallak</i>	<i>Malla'</i>
30.	Hidup	<i>Tallasak</i>	<i>Attallasa</i>
31.	Renang (Berenang)	<i>Aklange</i>	<i>A'lange</i>
32.	Terbang	<i>Anribbak</i>	<i>Anri'ba'</i>
33.	Berbelok	<i>Abbiluk</i>	<i>A'biluk</i>
34.	Narik (Menarik)	<i>Abbesok</i>	<i>Ambeso'</i>
35.	Ikat (Mengikat)	<i>Assikkok</i>	<i>Annyikko'</i>
36.	Jahit (Menjahit)	<i>Ajjaik</i>	<i>Anjai'</i>
37.	Hitung (Menghitung)	<i>Akreng</i>	<i>A'rekeng</i>
38.	Beku (Membeku)	<i>Abbatu</i>	<i>A'batu</i>
39.	Air	<i>Jeknek</i>	<i>Je'ne</i>
40.	Garam	<i>Cekla</i>	<i>Se'la</i>
41.	Kabut	<i>Saliuk</i>	<i>Salihu'</i>
42.	Langit	<i>Langik</i>	<i>Langi'</i>
43.	Es	<i>Esek</i>	<i>Esi</i>
44.	Asap	<i>Rumbu</i>	<i>Ambu</i>
45.	Abu	<i>Ambu</i>	<i>Ahu</i>
46.	Hangat	<i>Kammuk</i>	<i>Kammu'</i>
47.	Baru	<i>Beru</i>	<i>Bau</i>
48.	Baik	<i>Bajik</i>	<i>Baji'k</i>
49.	Busuk	<i>Bottok</i>	<i>Botto'</i>
50.	Lurus	<i>Lambusuk</i>	<i>Lambusu</i>
51.	Tumpul	<i>Pokkolok</i>	<i>Pokkolo'</i>
52.	Licin	<i>Laccuk</i>	<i>Lassu'</i>
53.	Kering	<i>Kalotorok</i>	<i>Kalotoro'</i>
54.	Betul	<i>Tojeng</i>	<i>Toje'</i>
55.	Licin	<i>Lasu</i>	<i>Lassu</i>

**Table 4 Pairs of Cognates with One Different Phoneme in Makassar and Selayar Languages**

No	Gloss	Makassar	Selayar
1.	Mulut	<i>Bawa</i>	<i>Baba</i>
2.	Lutut	<i>Kulantuk</i>	<i>Kalantuk</i>
3.	Kotor	<i>Rakmasak</i>	<i>Rammasak</i>

From the 200 lists of Makassar and Selayar vocabulary compiled by Morris Swadesh, the pair

of relatives (*Cognate*) Identical in Makassar and Selayar languages as many as 71 cognate words.

Relative Pairs (*Cognate*) Phonemic Correspondence of Makassar and Selayar Languages as many as 55 cognate words. Relative Pairs (*Cognate*) One Different Phoneme in Makassar and Selayar Languages is 3 cognate words. Thus, the total number of cognate pairs is 129 vocabulary pairs.

**Table 5 Percentage of Relative Words according to Crowley (1992:179)**

Language Level	PisRelative Words percentage
Language Dialects ( <i>Dialect of language</i> )	100 – 81
Family ( <i>family</i> )	81 – 36
Clump ( <i>stock</i> )	36 – 12
Microfilum	12 – 4
Meosophilum	4 – 1
Macrophyllum	< 1 %

$$C = \frac{\text{Cognates (Number of Relative Words)}}{\text{Gloss (Word Count compare)}} \times 100\%$$

$$C = \frac{129}{200} \times 100\%$$

$$C = 64,5 \%$$

Based on the table of percentages of kin words according to (Crowley, 1992), the results of the calculation of kinship levels fall into the category of family language levels. (*family*) because the percentage of the word relatives is 81-36.

(1) Calculating First Separation Time (WP1)

$$t = \frac{\log c}{2 \log r}$$

$$t = \frac{\log 64,5 \%}{2 \log 81 \%}$$

$$t = \frac{-0.190}{2 (-0,0915)}$$

$$t = \frac{-0.190}{-0.183}$$

$$t = 1,038$$

$$= 1,038 \times 1000$$

$$= 1038$$

So, the first time the Makassar language and the Selayar language separated was 1038 years ago.

(2) Calculating Second Separation Time (WP2)

The dissociation of a language from its constituents does not occur instantaneously, but rather, it occurs gradually. Consequently, the divergence period indicates a temporal range, for example spanning from year X to year Y, rather than indicating a specific year. Thus, a second calculation of the separation time is necessary, through the following steps.

(a) Calculating the Error Range

$$S = \sqrt{\frac{c(1-c)}{n}}$$

$$S = \sqrt{\frac{0,645(1-0,645)}{200}}$$

$$S = \sqrt{\frac{0,645(0,355)}{200}}$$

$$S = \sqrt{\frac{0,228975}{200}}$$

$$S = \sqrt{0,001145}$$

$$S = 0,034$$

(b) Calculating the second kinship percentage (C1)

$$C1 = c + s$$

$$C1 = 0,645 + 0,034$$

$$C1 = 0,679 (67,9\%)$$

(c) Calculating the second kinship percentage (WP2)

$$t = \frac{\log c1}{2 \log r}$$

$$t = \frac{\log 0,679}{2 \log 0,81}$$

$$t = \frac{-0.168}{2 (-0,0915)}$$

$$t = \frac{-0.168}{-0.183}$$

$$t = 0,918$$

$$= 0,918 \times 1000$$

$$= 918 \text{ tahun}$$

Thus, the new calculation of the separation time is 918 years ago. Furthermore, to obtain the error period, the old time is subtracted from the

new time, which is  $1038 - 918 = 120$  years. So, the error period for calculating the separation time between the Makassar language and the Selayar language is 120 years.

### (3) Determining the Time of Language Separation in Thousands of Years

The temporal divergence of language is measured in terms of millennia or thousands of years, this calculation is done using the following formula.

WP = First Separation Time  $\pm$  Error Range

$$\begin{aligned} \text{WP1} &= 1038 + 120 \\ &= 1158 \end{aligned}$$

$$\begin{aligned} \text{WP2} &= 918 - 120 \\ &= 798 \end{aligned}$$

Thus, the separation time of the two languages is between 798 – 1158 years ago. Then the final calculation of the separation time of the Makassar language and the Selayar language WP1 minus WP2 ( $1158 - 798 = 360$  years). The calculation time when the research year was calculated backwards from the separation time results in order to find out the calculation of the separation time calculated from the research time.

$$\text{TP1} = 2024 - 1158 = 866$$

$$\text{TP2} = 2024 - 798 = 1226$$

**Table 6 Separation Time according to Crowley (1992:179)**

Language Level	Time Separated in Centuries
Language Dialects ( <i>Dialect of language</i> )	0 – 5
Family ( <i>family</i> )	5 – 25
Clump ( <i>stock</i> )	25 – 50
Microfilum	50 – 75
Meosophilum	75 – 100
Macrophyllum	100 >

Based on the separation period table according to (Crowley, 1992), the results of the separation time calculation in centuries fall into the category of Language Dialect level (*Dialect of language*) because the separation time is 360 years in the 0-5 century classification.

## DISCUSSION

Discussion of the historical comparison between Makassar and Selayar languages is an interesting topic in comparative historical linguistic studies. The results of the kinship of 64.5% according to (Crowley, 1992) the results of the calculation of the level of kinship fall into the category of family language levels (family). This finding is in line with research published in the Gema Wiralodra journal which revealed that lexicostatistical analysis showed a level of kinship between the Makassar language and the Selayar language of 60% which classifies both as language subfamilies of the same parent language (proto) (Zulham et al., 2022).

The results obtained prove (Crowley, 1992) understanding that this kind of comparison in explaining the complexity and nuances of the language system, thus contributing to a broader understanding of the evolution and diversity of languages. Research in Linguistics and Literature Studies confirms that comparative historical linguistic analysis allows tracing significant changes in language evolution, with particular emphasis on the evolution of spelling, vocabulary, and writing styles that reflect social and cultural dynamics (Lukman & Chattri Sigit Widyastuti, 2024).

Using this lexicostatistical method to try to find out exactly how old a language is, when it emerged, and how it relates to its relatives. (Rohman, 2021) in the UNISDA repository explains that change is an inevitability that occurs in language, where changes in language over a certain period of time accumulate and have an impact on the classification of language kinship using

percentages and separation times as the main indicators.

Another thing that is proven is that the results of previous studies show that the average percentage of vocabulary similarities between the Makassar language group and other South Sulawesi languages is around 43% (Charles & Barbara, 1987). The Gowa or Lakiung dialect shows the highest level of divergence with vocabulary similarities of around 5-10 percentage points lower than the Konjo and Selayar languages when compared to other South Sulawesi languages. (Andini, 2019) in her research on the comparison of the acoustic characteristics of Indonesian, Makassar, and Konjo languages revealed that some people prefer to use Makassar and Indonesian because according to native speakers, the two languages make it easier for them to communicate with people outside the region.

However, this study found a 64.5% higher kinship between Makassar and Selayar languages. This finding is consistent with the research of (Sulistiyarini & Hendrosumoro, 2023) who studied the kinship of Javanese, Sundanese, and Makassar through comparative historical linguistic studies, where complex kinship patterns were found with varying levels of kinship proportions and different separation times between languages.

Basically, these two languages come from the same language family, namely the Austronesian language family which also includes other languages in Indonesia and Southeast Asia. (Purwanti, 2020) research on the Austronesian language from Sumatra confirms that Austronesian is the largest ethnic group inhabiting the territory of Indonesia, and the study of the land of origin of the Malay-Polynesian ethnic group has attracted attention in comparative linguistic studies. This is reinforced by the research of (Sari et al., 2025) which shows that the influence of Austronesian languages on

regional languages in Central Java can still be traced through the remaining linguistic traces.

However, there are differences and historical influences that make both have unique characteristics. First, the origin and history, namely the Makassar language comes from the South Sulawesi language group, while the Selayar language is a dialect of the Bugis language spoken in the Selayar Islands, South Sulawesi. The history of migration and interaction between language groups in this region plays an important role in the formation and development of these language. (Wahyuni, 2024) research on the sound changes of Proto Austronesian into Palembang language shows that there is a systematic pattern of phonetic change in the evolution of Austronesian languages in various regions of the archipelago.

Second, cultural and historical influences, namely along with the history of trade and politics in South Sulawesi, these two languages have received influences from various other languages and cultures, such as Javanese, Arabic, and Dutch. This is reflected in the vocabulary, sentence structure, and special vocabulary used in traditional cultural and economic activities. Research in Linguistic and Literary Studies reveals that the influence of Dutch colonialism and the impact of globalization are the main focal points that highlight the adaptation of language to social and cultural dynamics (Lukman & Chattri Sigit Widyastuti, 2024).

Third, structural differences, namely, despite having the same roots, Makassar and Selayar languages can have significant differences in terms of phonology (sound system), morphology (word structure), syntax (sentence structure), and lexicon (vocabulary). For example, words used for traditional or local concepts may differ between the two languages. (Andini, 2019) findings show that classifying languages based on acoustic characteristics can reveal significant structural differences between closely related languages.

Fourth, use and maintenance, namely with the changing times and the flow of globalization, the use of local languages such as Makassar and Selayar may decline. However, efforts to maintain and promote the use of these languages as an important part of local cultural and historical identity continue to be carried out by caring communities. (Khaer et al., 2023) research on the context of speech in case reports at the Makassar District Court shows that the Makassar language is still used in formal and legal contexts, although with adaptations to Indonesian as the official language.

An in-depth study of the history, development, and interaction of the Makassar and Selayar languages can provide richer insights into the cultural and linguistic complexities of the South Sulawesi region. (Nadofah et al., 2024) research in the study of the kinship of Serang Javanese and Surabaya Javanese using a lexicostatistical approach shows that a comprehensive analysis of language kinship relationships can reveal linguistic patterns that reflect the history of migration and cultural contact. By paying attention to the differences and similarities between the two, we can better appreciate the richness of Indonesia's linguistic and cultural heritage.

Although lexicostatistics is a valuable tool, some limitations require consideration such as borrowing, namely borrowed vocabulary can increase the level of cognates, potentially indicating a closer relationship than is actually the case. Careful analysis of cognates and consideration of historical language contact are essential. (Rohman, 2021) research emphasizes that in analyzing the kinship of Austronesian languages, it is necessary to consider historical and geographical factors that influence language development.

Even the rate of language change, namely lexicostatistics, assumes a relatively constant rate of language change, but this may not always be the case. This can affect the accuracy of

divergence time estimates. (Wahyuni, 2024) research shows that the types of phonetic changes that occur in Proto-Austronesian into its descendant languages do not always follow a uniform pattern, so it requires a deeper analysis to understand the complexity of language evolution.

## 5. CONCLUSION

Makassar language and Selayar language are in the same family, namely Bugis. However, not all languages used are the same, but there are identical and non-identical from both languages. From the purpose of the study which analyzed the Makassar language and Selayar language from the level of kinship and separation time and discussion of factors that influence kinship to the stage of its limitations. The results obtained kinship 64.5% kinship level falls into the category of family language level (family). Meanwhile, the results of calculating the separation time in centuries are included in the category of Language Dialect levels. (Dialect of language) because the separation time is 360 years in the 0-5 century classification. Factors that influence kinship such as origin and history, cultural influences, structural differences, language use and maintenance. Limitations require considerations such as language borrowing, careful analysis of cognates and consideration of historical language contact, as well as language change.

## 6. REFERENCE

- Abbas, A. (2021). *Makassar Language Empowerment on the Use of Indonesian Language in Non-Formal Communication*. <https://doi.org/10.2991/assehr.k.211225.014>
- Andini, C. C. P. (2019). *Perbandingan ciri akustik bahasa Indonesia, bahasa Makassar, bahasa Konjo: Kajian linguistik komparatif*. Universitas Pendidikan Indonesia.
- Anggayana, I. W. A., I, N. S., Dhanawaty, N. M., & I, G. B. (2020). Lipang, Langkuru, Waisika Language Kinship: Lexicostatistics Study in Alor Island. *International Journal of*



- Psychosocial Rehabilitation*, 24(04), 301–319.  
<https://doi.org/10.37200/IJPR/V24I4/PR201010>
- Blust, R. (2018). *Metode komparatif dalam linguistik Austronesia*. De Gruyter Mouton.
- Blust, R., & Trussel, S. (2013). The Austronesian Comparative Dictionary: A Work in Progress. *Oceanic Linguistics*, 52(2), 493–523.  
<https://doi.org/10.1353/ol.2013.0016>
- Bouckaert, R., Lemey, P., & Drummond, A. (2018). Divergensi penanggalan dan proses evolusi dalam pohon filogenetik. *Ekologi Molekuler*, 27(2), 210–226.
- Chang, W., Zhengzhang, Z., & Calude, C. (2021). Leksikostatistik dan klasifikasi internal bahasa Tai-Kadai. *Linguistik Komputasi*, 47(3), 581–604.
- Charles, E. G., & Barbara, D. G. (1987). *Languages of South Sulawesi*. Department of Linguistics, Research School of Pacific Studies, Australian National University.
- Crowley, T. (1992). *An Introduction to Historical Linguistics*. Oxford University Press.
- Dian Astuti. (2023). Relasi Kekerabatan Dialek Konjo dan Dialek Lakiung Bahasa Makassar: Pendekatan Dialektologi. *Jurnal Edukasi Khatulistiwa: Pembelajaran Bahasa Dan Sastra Indonesia*, 6, 10–21.  
<https://doi.org/10.26418/ekha.v6i1.54199>
- Finer, S., & Klemme, A. (2024). Phonetic and phonological aspects of Selayarese. *Journal of Southeast Asian Linguistics*, 15(2), 112–135.
- Hoffmann, K., Bouckaert, R., Greenhill, S. J., & Kühnert, D. (2021). Bayesian phylogenetic analysis of linguistic data using BEAST. *Journal of Language Evolution*, 6(2), 119–135.  
<https://doi.org/10.1093/jole/lzab005>
- Ino, L. (2015). Pemanfaatan Linguistik Historis Komparatif Dalam Pemetaan Bahasa-Bahasa Nusantara. *RETORIKA: Jurnal Ilmu Bahasa*, 1(2), 365–378.  
<https://doi.org/10.22225/jr.1.2.41.365-378>
- Keraf, G. (1996). *Linguistik Bandingan Historis*. Gramedia.
- Khaer, F., Munirah, M., & Syahrudin, S. (2023). Konteks Ujaran Dalam Berita Acara Perkara Kasus Body Shaming Di Pengadilan Negeri Makassar: Tinjauan Linguistik Forensik. *Indonesia: Jurnal Pembelajaran Bahasa Dan Sastra Indonesia*, 4(2), 90.  
<https://doi.org/10.59562/indonesia.v4i2.44297>
- Lukman, & Chattri Sigit Widyastuti. (2024). Bahasa Indonesia Sebagai Produk Budaya dan Bagian Dari Bahasa Austronesia: Suatu Tinjauan Linguistik Historis Komparatif. *Kajian Linguistik Dan Sastra*, 3(1), 1–13.
- McMahon, A., & McMahon, R. (2021). Statistical principles for lexicostatistics revisited. *Diachronica*, 38(3), 385–412.
- Nadofah, N., Andriani, L., Yuliyanti, K., & Muhyidin, A. (2024). Kekerabatan Bahasa Jawa Serang dengan Bahasa Jawa Surabaya (Kajian Leksikostatistik). *GHANCARAN: Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 5(2).  
<https://doi.org/10.19105/ghancaran.v5i2.9653>
- Purwanti, R. (2020). Bahasa Austronesia Dari Sumatera. *Prosiding Balai Arkeologi Jawa Barat*, 63–70.  
<https://doi.org/10.24164/prosiding.v3i1.7>
- Rama, T., Kolachina, S., & Jäger, G. (2022). Phylogenetic networks and distances in linguistic classification. *Language Dynamics and Change*, 12(1), 156–189.
- Rohman, M. K. (2021). *Kekerabatan bahasa-bahasa Austronesia*. Universitas Islam Darul Ulum Lamongan.
- Sari, D. M., Pratiwi, N., & Kusuma, A. (2025). Pengaruh Bahasa Austronesia Terhadap Bahasa Daerah di Jawa Tengah. *Kajian Linguistik Dan Sastra*, 4(1), 23–35.
- Sirk, Ü. (1989). On the Evidential Basis for the South Sulawesi Language Group. *Nusa*, 3(1), 55–82.
- 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>
- Wahyuni, S. (2024). Perubahan Bunyi Bahasa Proto Austronesia ke dalam Bahasa Palembang Dialek Melayu Palembang: Kajian Linguistik Historis Komparatif. *Jurnal Pendidikan Tambusai*, 8(2), 234–247.
- Zhang, M., & Gong, T. (2016). How Many Is Enough? Statistical Principles for Lexicostatistics. *Frontiers in Psychology*, 7.  
<https://doi.org/10.3389/fpsyg.2016.01916>



Zulham, Rahim, Abd. R., & Agus, M. (2022).  
Kekerabatan Bahasa Makassar dan Bahasa  
Selayar: Analisis Leksikostatistik dan  
Glotokronologi. *Gema Wiralodra*, 13(1), 215–  
232. <https://doi.org/10.31943/gw.v13i1.215>