EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 10. No. 1, January 2026
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

Tracing Historical Linguistic Kinship: A Lexicostatistical Study on Dayak Ngaju and Banjar Kuala Languages in Kalimantan

Tiara Dwi Saputri, Misrita[2], Imam Qalyubi[3]
University of Palangka Raya[1,2], Islamic University of Palangka Raya[3]
Central Kalimantan - Indonesia
Email: [1]tiaradwisaa@gmail.com, [2]misrita@for.upr.ac.id ,
[3]imamqalyubi@yahoo.com

Article History:
Received: 20/06/2025
Accepted: 16/12/2025
Published:10/01/2026

Keywords*:*
Lexicostatistics, Dayak Ngaju, Banjar Kuala, Language Kinship

Abstract

This study explores the lexical kinship between the Dayak Ngaju and Banjar Kuala languages in Kalimantan using a lexicostatistical approach. The research aims to measure the degree of relatedness between the two languages based on core vocabulary and to classify their genealogical relationship. A total of 200 basic vocabulary items were obtained from Lexirumah and the Austronesian Basic Vocabulary Database using purposive sampling. The data were analyzed using Swadesh's formula to calculate the percentage of cognates. The results show a lexical similarity of 31%, placing the relationship at the "stock" level according to Keraf's classification. The findings suggest a historical connection supported by systematic phonological patterns, including apocope, apheresis, and metathesis. This study provides the first lexicostatistical comparison between Dayak Ngaju and Banjar Kuala and contributes to comparative historical linguistics as well as language preservation efforts in Kalimantan.

Kata Kunci:
Leksikostatistik, Dayak Ngaju, Banjar Kuala, Kekerabatan Bahasa

Abstrak

Penelitian ini mengkaji kekerabatan leksikal antara bahasa Dayak Ngaju dan Banjar Kuala di Kalimantan dengan menggunakan pendekatan leksikostatistik. Tujuan penelitian ini adalah mengukur tingkat hubungan antara kedua bahasa berdasarkan kosakata dasar dan menentukan klasifikasi kekerabatan secara genealogis. Sebanyak 200 kosakata dasar dikumpulkan dari Lexirumah dan Austronesian Basic Vocabulary Database dengan teknik purposive sampling. Data dianalisis menggunakan rumus Swadesh untuk menghitung persentase kosakata kerabat (cognates). Hasil penelitian menunjukkan tingkat kesamaan leksikal sebesar 31%, yang menempatkan hubungan kedua bahasa pada tingkat *stock* berdasarkan klasifikasi

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 10. No. 1, January 2026
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

Keraf. Temuan ini menunjukkan adanya hubungan historis yang diperkuat oleh pola fonologis sistematis seperti apokope, aferesis, dan metatesis. Penelitian ini merupakan studi pertama yang membandingkan kedua bahasa melalui leksikostatistik dan memberikan kontribusi pada linguistik historis komparatif serta upaya pelestarian bahasa di Kalimantan.

INTRODUCTION

Comparative historical linguistics has long served as a foundational approach to tracing the origins and relationships of languages. This approach allows scholars to identify similarities and divergences between languages based on consistent phonological and lexical changes over time (Fernandez, 1996; Salahuddin, 2023). One of the most widely accepted and effective techniques in this field is lexicostatistics, a statistical method for comparing core vocabulary items to determine the degree of linguistic kinship (Swadesh, 1952; Darman, 2022).

Lexicostatistical studies have been successfully applied to various Indonesian languages to uncover their linguistic connections. Mahriyuni et al. (2023), for example, showed that Javanese and Sasak share a 23.8% kinship based on 207 basic vocabulary items. Similarly, Salahuddin (2023) found that Bima and Manggarai share a 27% lexical similarity, suggesting they diverged from a common proto-language around 3,000 years ago, using glottochronological estimation.

More recent studies have also applied lexicostatistics specifically to languages in Kalimantan. Septian and Rahmat (2022) compared Banjar and Bakumpai and found a 35% lexical similarity, indicating a close genealogical link between the two. Likewise, Yuliana and Hasan (2021) analysed Dayak Bakumpai and Dayak Ngaju using a Swadesh list and identified systematic phonological correspondences supporting shared inheritance. Additionally, Sari and Arifin (2023) conducted a glottochronological study on Bornean Austronesian languages and emphasized that quantitative comparative linguistic research in Kalimantan remains limited. These studies demonstrate increasing scholarly interest in Kalimantan languages; however, none have examined the relationship between Dayak Ngaju and Banjar Kuala using lexicostatistical methods.

In the Kalimantan region, however, such quantitative comparative research remains scarce. Dayak Ngaju and Banjar Kuala spoken in Central and South Kalimantan have not been previously analyzed through systematic comparative methods. Yet, as Afria et al. (2020) and Setiawan (2020) note, linguistic comparison is essential in revealing migration patterns, cultural contact, and the evolution of local Austronesian languages.

Most earlier studies concerning the Dayak or Banjar languages have leaned toward descriptive or sociolinguistic analyses, rather than historical-comparative linguistics. For instance, Bustan et al. (2020) focused on the socio-historical aspects of Manggarai society, while Rizqi and Widayati (2021) examined variation in Sundanese and Baduy speech in a contemporary context. The lack of quantitative approaches highlights a significant gap in the research landscape.

The strength of lexicostatistics lies in its ability to present objective, quantifiable evidence. According to Humaidi and Kasmilawati (2023), this method not only yields lexical similarity percentages but also estimates the timeframe of language divergence. As Indonesia faces increasing threats of language extinction, this approach is particularly urgent and relevant (Erni et al., 2022).

Several other studies further underscore the method's potential. Aisyah and Widayati (2022) identified kinship among three dialects in Sumatra, while Erniati (2021) explored similarities between Banggoi and Hoti in the Maluku Islands. Even in areas with high cultural and geographical diversity, this approach continues to reveal meaningful linguistic relationships (Iqbal et al., 2022; Jamzaroh, 2020).

Lexicostatistics also enables analysis of phonological correspondences. Mahriyuni et al. (2023) classified cognate words into three main categories: identical pairs, phonemic correspondences, and single-phoneme differences. This categorization helps clarify whether similarities arise from shared inheritance or lexical borrowing (Ahya et al., 2022).

Databases such as Lexirumah (Kaiping & Klamer, 2020) and the Austronesian Basic Vocabulary Database (Kaiping et al., 2019) offer valuable resources for such research, although Kalimantan languages remain

underrepresented. As Hakim (2020) observes, there is still a pressing need to document and analyze the region's linguistic diversity systematically.

Beyond the theoretical contributions, comparative linguistic research also holds socio-cultural value. Tracing kinship between languages can strengthen ethnic identity and foster cultural awareness among local communities (Muhammad & Hendrokumoro, 2022). Such findings may inform local content curricula and support cultural revitalization efforts through evidence-based linguistic documentation.

Understanding the linguistic relationship between Dayak Ngaju and Banjar Kuala is crucial because both languages are currently under-documented and at risk of reduced intergenerational transmission. Identifying their lexical kinship provides empirical evidence needed for language maintenance, revitalization, and curriculum development in Kalimantan. Moreover, this research contributes to filling a knowledge gap in Austronesian comparative linguistics, where Kalimantan languages remain understudied despite their cultural and historical significance.

In summary, the literature shows that lexicostatistical analysis is not only theoretically grounded but also practically impactful. This study seeks to apply such a framework to Dayak Ngaju and Banjar Kuala two under-researched but culturally rich languages in Kalimantan thereby filling a significant gap and enriching our understanding of the Austronesian language landscape in Indonesia.

In the strategic competence and self-regulated learning (SRL) have recently received a lot of attention in the discussions surrounding language learning in an EFL setting. Strategic competence is defined as the ability of a learner to select, use, monitor and adjust learning strategies relevant to a particular situation and to a learning goal Oxford (2020). Self-regulated learning, on the other hand, focuses on a learner’s ability to control his or her progress by planning, monitoring, and evaluating their learning based on the set goals, self-reflection, and motivational control. The merger of these two concepts develops a profound perspective of how learners approach the management of learning a second or foreign language.

Research studies from recent years have analyzed the impact of motivation and strategy application on EFL learning results. The research conducted by An et al. (2020) Click or tap here to enter text. demonstrated that students who achieve high grades use more technology-based SRL strategies because they are intrinsically motivated. Bai and Wang (2023) demonstrated that students who maintained a growth mindset together with high self-efficacy levels showed stronger engagement in strategic learning activities. Research indicates that motivational elements significantly affect how learners behave strategically while performing in language tasks.

Other studies indicated the importance of teaching strategies for developing metacognition and improving academic performance. The works of (Chinpakdee and Gu 2024; Machili et al. 2020) validated the advantages of integrating strategic teaching within genuine learning activities. Furthermore, Habók et al. (2021) as well as Maftoon and Fakhri Alamdari (2020) pointed out that metacognitive techniques are much more effective when used alongside reflection and active feedback.

Nonetheless, these advancements have neglected numerous other aspects that still need further research. Most of the previous literature focused solely on quantifiable content, neglecting rich qualitative aspects of learners' experiences (Alabidi et al. 2022; Onah et al. 2020). There is also a lack of research on the development of strategic competence and self-regulated learning in diverse cultural and educational contexts (Kölemen 2021; Ma 2021a). These gaps underscore the need for comparative studies and exploratory analyses designed to integrate theoretical frameworks with tangible classroom approaches.

METHOD

This study applies a qualitative descriptive method with a lexicostatistical quantitative approach. Although the overarching framework is qualitative, the use of numerical calculations in determining lexical similarity renders this research methodologically hybrid. The qualitative aspect lies in the descriptive analysis of phonological correspondences and the classification of cognate words, while the

quantitative element is reflected in the lexicostatistical computation. As noted by Salahuddin (2023), this mixed approach is suitable for tracing historical linguistic kinship and estimating language divergence time. The method is designed to align with the research objective of uncovering the kinship level between the Dayak Ngaju and Banjar Kuala languages based on core vocabulary.

The data sources in this research consist of both secondary and primary data. Secondary data were obtained from established linguistic corpora such as the Lexirumah database (Kaiping & Klamer, 2020) and the Austronesian Basic Vocabulary Database (ABVD) (Kaiping et al., 2019). Both databases provide lexical items based on the 200-item Swadesh list, which represents culturally neutral, universal vocabulary such as kinship terms, body parts, natural elements, numerals, and common verbs (Swadesh, 1952; Darman, 2022). Supplementary sources, including regional dictionaries and previous descriptive works, were also consulted for phonological confirmation (Hakim, 2020; Iqbal et al., 2022).

To strengthen the accuracy of the lexical forms, primary data were also collected directly from native speakers. Two informants participated in this process: one native speaker of Dayak Ngaju and one native speaker of Banjar Kuala. Both informants are fluent first-language users representing older generational speech patterns, which reduces the influence of Indonesian or external contact languages. The primary Swadesh list was elicited through structured interviews, during which each lexical item was spoken, repeated when necessary, and clarified to avoid semantic and phonological ambiguity. Slow pronunciation and auditory verification were employed to prevent transcription errors. Verbal consent was obtained, and participation was voluntary, with data used solely for academic purposes.

The data collection technique employed was documentation and structured comparison. Vocabulary items from both languages were paired based on similarity in form and meaning. The word pairs were then categorized into three types: identical pairs, phonemic correspondence pairs, and minimal phonemic variation, following the classification system used by Mahriyuni et al. (2023). Only lexical items demonstrating systematic phonological patterns or historical continuity were

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 10. No. 1, January 2026
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

categorized as cognates, while forms suspected to be loanwords or accidental similarities were excluded (Erni et al., 2022; Ahya et al., 2022).

A purposive sampling technique was used in this study, where lexical items were selected based on three criteria: (1) belonging to the Swadesh core vocabulary list, (2) being present in both language data sets, and (3) demonstrating intergenerational stability. After classification, the percentage of cognate words was calculated using the lexicostatistical formula. The estimated time of language separation was further analyzed using glottochronological techniques (Humaidi & Kasmilawati, 2023; Lees, 1953). This systematic procedure ensures that the resulting classification accurately reflects the historical relationship between the Dayak Ngaju and Banjar Kuala languages.

FINDINGS AND DISCUSSION

Indonesia	Dayak Ngaju	Banjar Kuala	Reason	Notes
Aku	Aku	unda		
Jika	Amun	Amun	Identical in sound and meaning	Identical pair
Cepat	Ancap	Lakasi		
Hari	Andau	Hari		
Ayah	Apang	abah		
Api	Apui	Api	Apocope-Only differs in final vowel (-ui vs. -i)	Phonemic correspondence
Nama	Aran	Ngaran	Addition of /ŋ/ prefix in Banjar	Phonemic correspondence
Asap	Asep	Palak		
Anjing	Asu	Anjing		
Hati	Atei	Hati	Atei → Hati: /a/ ↔ /h/, /ei/ ↔ /i/ Identical in sound and meaning	phonemic correspondence.
Baik	Bahalap	Baik		
Tua	Bakas	Tuha		
Cantik/Gant eng	Bakena	Bungas		
Panas	Balasut	Panas		
Rambut	Balau	Rambut		
Tikus	Balawau	Tikus		
Suami	Bana	laki		
Putih	Baputi’	Putih	Apheresis (loss of “Ba-” prefix in Banjar)	One Phoneme Different
Dari	Bara	Tumat		
gendut	Baseput	Lamak		
Marah	Basingi	Sarik		
Tajam	Batajim	Landap		

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 10. No. 1, January 2026
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

Tidur	Batiruh	Guring		
Batu	Batu	Batu		Identical pair
Tumpul	Batumpul	Tumpul	Apheresis (loss of prefix "Ba-" in Banjar)	One Phoneme Different Pairs
Depan	Baun	Muka		
Perempuan	Bawi’	Binian		
Bukan	Beken	Lain		
Hidup	Belum	Hidup		
Hitam	Bilem	Hirang		
Basah	Bisa’	Basah		
Badan	Biti	Awak		
Buah	Bua’	Buah	Apocope in Dayak – Dayak Bua’ loses final /h/, represented as glottal stop /ʔ/ → Banjar keeps /h/	Phonemic Correspondence Pair
Pelit	Bukih	Pamalar		
Bulan	Bulan	Bulan		Identical pair
Bulu	Bulu	Bulu		Identical pair
Arang	Buring	Harang		
Darah	Daha’	Darah	Metathesis in Banjar – Banjar swaps Dayak /h/ and /r/ positions: Da-ha → Da-ra; glottal stop /ʔ/ omitted	Phonemic Correspondence Pair
Air	Danum	Banyu		
Daun	Dawen	Daun	Syncope in Banjar – Banjar drops medial vowel /e/ found in Dayak Dawen → Daun	Phonemic Correspondence Pair
Dengan	Dengan	Awan		
Tidak	Dia’	Kada		
Dua	Due’	Dua	Apocope in Dayak – Dayak Due’ has final glottal /ʔ* which is absent in Banjar Dua; vowel /e/ ~ /a/ variation	One phoneme different – Phonemic Correspondence Pair
Datang	Dumah	Datang		
Dahan	Edan	Dahan	Metathesis in Banjar – Banjar swaps Dayak’s /d/ and /h/: E-dan → Da-han; vowel /e/ → /a/ shift	2 Changes
Lemak	Enyak	gajih		
Empat	Epat	Ampat	Prothesis in Banjar – Banjar adds /m/ at onset: Epat → Am-pat; vowel /e/ → /a/ shift	One Phoneme Different
Siapa	Eweh	Siapa		
Mereka	Ewen	Buanya		
Tampak	gitan	Kalihatan		
Guntur	Guntur	Guntur		Identical pair
Kutu	Guti’	Kutu	Metathesis in Dayak – Dayak swaps /k/ → /g/, and /u/ and /i/ are	2 changes

			reordered: Kutu → Guti'; glottal stop /ʔ/ added in Dayak	
Sakit	Haban	Garing		
Sapi	Hadangan	Hadangan		Identical pair
Lari	Hadari	Bukah		
Besar	Hai'	Ganal		
Berbunyi	Hamauh	Bebunyi		
Kapan	Hamparea'	Pabila		
Berenang	Hanangui	Bakunyung		
Cacing	Handalai	Cacing		
Ular	Handipe'	Ular		
Pagi	Hanjewu	baisukan		
Sebentaer	Hanjulu	Satumat		
Delapan	Hanya	Dalapan		
Baru	Haru	Hanyar		
Kabut	Hasep	Kabut		
Bareng	hayak	Baimbai		
Dengar	Hining	Dangar		
Rumah	Huma'	Rumah	Metathesis in Banjar – Dayak Huma' has /h-u-m/, Banjar reorders to /r-u-m/ + Prothesis /r/ in Banjar	Phonemic Correspondence
Di	Hung	Wadah		
Satu	Ije'	Satu		
Kamu	Ikau	Nyawa		
Kami	Ikei	Kami		
Ekor	Ikuh	Buntut		
Pegang	Imbing	Pingkut		
Ibu	Indu	mama		
Daging	Isin	Daging		
Kita	Itah	Kita	Metathesis in Banjar – Reordering of /t/ and /k/, initial vowel /i/ changed to /k/ in Banjar	Phonemic Correspondence
Dia	Iye	Inya	Epenthesis in Banjar – Insertion of /n/ in medial position: Iye → Inya	One Phoneme Different Pairs
Enam	Jahawen	Anam		
Lempar	Jakah	Tawak		
Jatuh	Jatu'	Gugur		
Lidah	Jela'	Ilat		
Ini	Jetuh	nangini		
Itu	Jite'	Nintu		
Sendiri	Kabuat	Saurangan		
Bunga	Kambang	Kambang		Identical pair
Perut	Kanai'	Parut		
Jalan	Karatak	Kartak		Identical pair
Gigi	Kasinga	Gigi		
Mengetahui	Katawam	Tahu		
Abu	Kawu	Habu	Apheresis in Banjar – Initial /k/ in Dayak lost in Banjar; /w/ ~ /b/ shift	One Phoneme Different Pairs

Kering	Keang	Karing	Syncope in Dayak – Karing → Keang; loss of medial consonant /r/	Phonemic Correspondence
Jauh	Kejau	Jauh	Prothesis in Dayak – Dayak adds /ke-/ prefix: Jauh → Kejau	Phonemic Correspondence
Kalian	Ketun	Buhan Ikam		
Bagaimana	Kilenampi	Kayapa		
Malas	Kulas	Koler		
Makan	Kuman	Makan	Apheresis in Banjar – Loss of initial /ku-/ in Banjar: Kuman → Makan	
Penuh	Kuntep	Hibak		
Kecil	Kurik	Halus		
Langit	Langit	Langit		Identical pair
Ikan	Lauk	Iwak		
Tangan	Lenge’	Tangan		
Belakang	Likut	Balakang		
Lima	Lime	Lima	Apocope – final vowel /e/ in Dayak becomes /a/ in Banjar	Phonemic Correspondence Pair
Lebar	Lumbah	Ligar		
Malu	Mahamen	Supan		
Memeras	Mahamis	Mamarah		
Mengalir	Mahasur	Mangalir		
Menguap	Malalap	Manguap		
Meludah	Maluja	Maludah	Apocope – loss of final /d/ in Dayak; same root	One Phoneme Different Pair
Menggigit	Mamangkit	Maigut		
Memasak	Mampakasak	Bamasak		
Memukul	Mamukul	Mamukul		Identical pair
Melihat	Manampayah	Malihat		
Menangis	Manangis	Manangis		Identical pair
Berjalan	Mananjung	Bajalan		
Memanjat	Mandai’	Manaik	Metathesis – /d/ and /n/ swapped positions, vowel shift	One Phoneme Different Pair
Berburu	Mandup	Handup	Apheresis – initial /H/ lost in Dayak form	Phonemic Correspondence Pair
Menombak	Manepe’	Manumbuk		
Menggaruk	Manggayau	Manggaruk		
Mimpi	Manupi	Mimpi		
Mengunyah	Manyipa’	Manginang		
Bakar	Mapoi	Banam		
Busuk	Maram	Buruk		
Mata	Mata’	Mata		Identical pair
Mati	Matei	Mati		
Berdiri	Mendeng	Badiri		
Berburu	Mengan	Bahandup		

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 10. No. 1, January 2026
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

Berbaring	Menter	Barabah		
Ikat	meteng	Ikat		
Bintang	Metu’	Bintang		
Minum	Mihup	Minum	Metathesis – /h/ in Dayak inserted in Banjar as /n/	Phonemic Correspondence Pair
Takut	Mikeh	Takutan		
Menanam	Mimbul	Tanam		
Memilih	Mintih	Mamilih	Metathesis + Morphological change: mintih and milih share root "pilih" with inflection differences	Phonemic Correspondence Pair
Isap	Minyup	Isap		
Hitung	Mise’	Hitung		
Lihat	Mite’	Lihat		
Menjahit	Mitur (mamitur)	Manjahit		
Buka	Mukei	Singkai		
Duduk	Munduk	Duduk	Metathesis + Prothesis – du > mu, shared root duk	Phonemic Correspondence Pair
Muntah	Muta’	Muntah	Metathesis – *ta* > *nta*, shared root	Phonemic Correspondence Pair
Bernapas	Nahaseng	Bahinak		
Apa	Narai	Napa	Apocope – final /i/ in Dayak lost; phonemic shift /r/ to /p/	One Phoneme Different Pair
Potong	Netek	Tatak	Metathesis – ne ↔ ta, shared root meaning “cut/press”	Phonemic Correspondence Pair
Diatas	Ngambu	Diatas		
Gigit	Ngirut	Igut	Apheresis – loss of initial /ŋ/ in Banjar form	Phonemic Correspondence Pair
Simpan	Nyahukan	Simpan		
Mulut	Nyama	Muntung		
Nyamuk	Nyamuk	Nyamuk		Identical pair
Tahun	Nyelu	Tahun		
Belah	Nyila	Balah		
Kaki	Pai’	Batis		
Sayap	Palapas	Halar		
Pendek	Pandak	Handap	Metathesis and Apheresis – handap > pandak, same meaning “short”	Phonemic Correspondence Pair
Panjang	Panjang	Panjang		Identical pair
Kotor	Papa’	Rigat		
Bokong	Para	Burit		
Pasir	Pasir	Pasir		Identical pair
Sakit	Pehe’	Garing		
Tanah	Petak	Tanah		
Beli	Pili’	Tukar		
Jarum	Pilus	Jarum		
Telinga	Pinding	Talinga		
Undangan	Rawei	Saruan		

Sehat	Rigas	Sigar	Metatesis dan pertukaran fonem /r/ ↔ /s/ serta urutan vokal-konsonan	Phonemic Correspondence Pair
Angin	Riwut	Angin		
Laba-laba	Sabangkang	Kabibitak		
Dingin	Sadingen	Dingin	Aferesis: /sa-/ hilang di Banjar, sisanya mirip "dingen"	Phonemic Correspondence Pair
Seribu	Sakuyan	Saribu		
Kiri	Sambil	Kiwa		
Atap	Sapau	Hatap		
Sepuluh	Sapuluh	Sapuluh		Identical pair
Seratus	Saratus	Saratus		Identical pair
Kumis	Sasingut	Sasingut		Identical pair
Istri	Sawa	Bini		
Muda	Tabela	Anum		
Curi	Takau	Cuntan		
Kepala	Takuluk	Kapala		
Tali	Tali	Tali		Identical pair
Telur	Tanteluh	Hintalu	Metatesis: urutan silabel /tan-te/ ↔ /hin-ta/, dan fonem /luh/ ↔ /lu/	Phonemic Correspondence Pair
Terbang	Tarawang	Tarabang	/w/ ↔ /b/ substitution in the middle; otherwise very similar	One Phoneme Different Pair
Tertawa	Tatawe’	Tatawa	Apocope – glottal stop /ʔ/ dropped in Banjar	Phonemic Correspondence Pair
Tiga	Telu’	Tiga		
Sembilan	Jalatien	Sambilan		
Dekat	Tukep	Parak		
Tulang	Tulang	Tulang		Identical pair
Tumbuh	Tumbu’	Tumbuh	Apocope: glottal stop /ʔ/ dropped in Banjar	Phonemic Correspondence Pair
Benar	Tutu’	Banar		
Akar	Uhat	Akar		
Hujan	Ujan	Hujan	Prothesis – /h/ added in Banjar	Phonemic Correspondence Pair
Tujuh	Uju	Tujuh	Prothesis + phoneme change /t/ ↔ zero; same meaning	One Phoneme Different Pair
Orang	Uluh	Urang	Both mean "person"; similar vowel structure /u-/ and nasal ending	Phonemic Similarity Pair
Kulit	Upak	Kulit		
Semua	Uras	Barataan		
Rumput	Uru’	Kumpay		
Hidung	Urung	Hidung	Apheresis: /hi-/ dropped in Dayak; /d/ ↔ /r/ substitution	Phonemic Correspondence Pair
Garam	Uyah	Uyah		Identical pair
Leher	Uyat	Gulu		

This study applies the lexicostatistical method introduced by Swadesh (1952) to determine the percentage of lexical cognates between the Dayak Ngaju and Banjar Kuala languages. The method relies on comparing basic vocabulary items, specifically 200 core words taken from the Swadesh list, which represent fundamental, culture-free concepts likely to be preserved over time.

1. **Step-by-step Lexical Cognate Analysis:**
   a. Total words compared (n): 200
   b. Number of cognate pairs (k): 62
      1) Identical pairs: 18
      2) Phonemic correspondence pairs: 25
      3) One-phoneme difference pairs: 19
   c. Non-cognate pairs: 138
2. **Lexical Similarity Formula (Swadesh):**

   To calculate the percentage of lexical similarity (C):

$$C = \left(\frac{k}{n}\right) \times 100$$

   C = persentase kognat (lexical similarity)

   k = jumlah pasangan kata yang berkerabat (cognate pairs)

   n = jumlah total kata yang dibandingkan

   Substituting the values:

$$C = \left(\frac{62}{200}\right) \times 100 = 31\%$$

   Thus, 31% of the basic vocabulary is cognate between Dayak Ngaju and Banjar Kuala.
3. **Glottochronological Estimation (Optional)**

   If estimating the time of divergence between the two languages using Swadesh's glottochronological formula, we apply:

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

EJI (English Journal of Indragiri): Studies in Education, Literature, and Linguistics
Vol. 10. No. 1, January 2026
ISSN (Print). 2549-2144, ISSN (Online). 2589-5140
Homepage: https://ejournal-fkip.unisi.ac.id/index.php/eji/

Where:

t = time of separation (in millennia)

C = proportion of retained cognates (in decimal) → 0.31

r = rate of retention per millennium (commonly used value: 0.805)

$$t = \frac{\log(0.31)}{2\log(0.805)} \approx \frac{-0.5086}{2 \times -0.0943} \approx \frac{-0.5086}{-0.1886} \approx 2.69$$

Thus, the estimated time of separation is approximately 2.7 millennia, or 2,700 years ago.

Note: This assumes a constant retention rate and should be interpreted cautiously, especially when based solely on secondary lexical data.

4. Tools and Classification Process:

All calculations and classifications were conducted using Microsoft Excel, with word-by-word comparison across semantic meanings.:

a. Identical forms
b. Systematic phonological changes (e.g., apocope, apheresis, metathesis)
c. Phonemic correspondence following predictable patterns

Each pair was further annotated with phonological notes and categorized by type, e.g., "Identical," "Phonemic Correspondence," or "One-phoneme Difference."

Table 5. Classification Summary Table:

Category	No. Of Pairs	Percentage
Identical Cognates	18	9%
Phonemic Correspondence	25	12.5%
One-phoneme Difference	19	9.5%
Total Cognates	62	31%
Non-Cognate Pairs	138	69%

The data in this study were analyzed using a quantitative descriptive statistical method based on a lexicostatistical approach. The primary analytical technique involved calculating the percentage of cognates from the basic vocabulary pairs using Swadesh's formula:

C=kn×100C = \frac{k}{n} \times 100C=nk×100

where *k* represents the number of cognate word pairs, and *n* is the total number of compared vocabulary items (200 words). To ensure the validity of the results, each word pair was evaluated based on systematic phonological correspondence, including identical forms, phonemic variations, and recognized phonological processes such as apocope, apheresis, and metathesis. All data were processed and tabulated using Microsoft Excel, which supported the calculation and categorization of the data. Although software like SPSS or SmartPLS was not employed, data reliability and consistency were maintained by applying linguistic classification principles derived from Keraf (1984) and Swadesh (1952), with cross-referencing against two standard lexical databases: Lexirumah and the Austronesian Basic Vocabulary Database (ABVD). The results were then interpreted using Keraf's genealogical classification system to determine the degree of historical kinship between the two languages.

Based on the analysis of 200 basic vocabulary items, a total of 62 word pairs were found to be cognates, consisting of 18 identical pairs, 25 pairs with phonemic correspondence, and 19 pairs with one phoneme difference. This yields a lexical similarity rate of 31% between Dayak Ngaju and Banjar Kuala, placing the relationship in the "stock" category according to Keraf's (1991) classification. This finding confirms the research question, showing that there is indeed a historical connection between the two languages, although not close enough to be considered members of the same language family. The percentage is higher than that found in Mahriyuni et al. (2023) between Javanese and Sasak (23.8%) and comparable to Salahuddin's (2023) result between Bima and Manggarai (27%).

These results support the hypothesis that Dayak Ngaju and Banjar Kuala evolved from a shared linguistic ancestor, likely a proto-Austronesian language that underwent regular phonological divergence. Strong evidence of this kinship can be seen in similar basic vocabulary such as *"langit" (sky)*, which is identical in both languages, and *"garam" (salt)* rendered as *uyah* in both. Phonological transformations such as *"apui"* (Dayak Ngaju) to *"api"* (Banjar) through apocope and *"narai"* to *"napa"* through phonemic change further suggest a historical rather than borrowed connection. The presence of systematic processes like metathesis

and apheresis in cognate pairs aligns with Keraf's (1984) theory of regular sound change, reinforcing the reliability of the lexicostatistical method in comparative historical linguistics.

Nonetheless, several factors may have influenced the findings. First, the use of secondary data means that actual dialectal variation and phonological realization by native speakers could not be observed directly. Second, language contact and borrowing from dominant languages like Malay or Indonesian may have obscured genuine cognates. Despite these limitations, this study provides a significant contribution to the linguistic mapping of Kalimantan and lays a solid foundation for future research on regional language preservation and comparative linguistic analysis.

The novelty of this study lies in its scope, dataset, and methodological contribution. While previous research in Indonesia and Kalimantan has explored lexical kinship among other Austronesian languages, no prior study has examined the relationship between Dayak Ngaju and Banjar Kuala using a Swadesh-based lexicostatistical framework. This research is the first to provide a systematically coded cognate classification supported by verified phonological processes such as apocope, apheresis, and metathesis. Additionally, the integration of primary data from native speakers with secondary lexical databases strengthens the methodological reliability and reduces modernization bias, which has not been applied in earlier studies. The findings not only present the first quantified kinship value (31%) between the two languages but also establish empirical evidence that can support future language documentation, revitalization programs, and comparative Austronesian studies in Kalimantan.

CONCLUSION

This study concludes that the lexical kinship level between Dayak Ngaju and Banjar Kuala is 31%, placing them within the "stock" category according to Keraf's genealogical classification. This result supports the research hypothesis that the two languages share a historical relationship, thus affirming the hypothesis is accepted. The findings directly address the research question by demonstrating

systematic phonological patterns and cognate structures across 200 basic vocabulary items, validating the use of lexicostatistics as an effective method in comparative historical linguistics. The results are consistent with previous studies by Mahriyuni et al. (2023) and Salahuddin (2023), and contribute to a broader understanding of Austronesian language development, especially in under-researched regions like Kalimantan. Academically, this research reinforces the value of quantitative approaches in linguistic classification, while practically, it offers a foundation for developing language preservation strategies and local content curricula. Future research is recommended to include morphological and syntactic comparisons, supported by fieldwork with native speakers to enrich and validate lexical data.

REFERENCES

Afria, R. A., Izar, J., Prawolo, I. S., & Arezky, B. (2020). Relasi bahasa Melayu Riau, Bugis, dan Banjar: Kajian linguistik historis komparatif. *Medan Makna*, 18(1), 94–102.

Ahya, A. S., Subakti, H. S., & Surotin, S. (2022). Kekerabatan bahasa Jawa Nganjuk dengan bahasa Jawa Jombang (Kajian leksikostatistik). *Linguista*, 6(1), 51–61.

Aisyah, S., & Widayati, D. (2022). Hubungan kekerabatan bahasa Pesisir Pasar, Kampung, dan Sorkam. *Aksara*, 8(3), 2367–2376.

Darman, M. (2022). Metode leksikostatistik dalam linguistik komparatif. *Jurnal Bahasa dan Ilmu Budaya*, 8(2), 333–343.

Erni, N., Taha, M., Febriningsih, F., Wijaya, D., & Garing, J. (2022). Bahasa Lorang, Barakai, dan Dobel di Kepulauan Aru dalam kajian leksikostatistik. *Kandai*, 18(1), 61–76.

Erniati, E. (2021). Klasifikasi leksikostatistik bahasa Banggoi dan Hoti di Kabupaten Seram Bagian Timur. *Totobuang*, 9(2), 239–255.

Hakim, L. (2020). Kekerabatan bahasa Sabu dan Rote berdasarkan leksikostatistik. *Jurnal Pendidikan Bahasa dan Sastra*, 20(1), 55–63.

Humaidi, H., & Kasmilawati, K. (2023). Teknik glotokronologi untuk penentuan usia bahasa di Indonesia Timur. *Jurnal Linguistik Teoritis dan Terapan*, 12(1), 63–72.

Iqbal, M., Fauzi, F., & Satria, A. (2022). Perbandingan bahasa Kaili dan Kulawi: Kajian leksikostatistik. *Jurnal Bahasa dan Sastra*, 23(2), 130–139.

Jamzaroh, J. (2020). Pengantar ilmu bahasa dan klasifikasi bahasa. *Jurnal Ilmiah Wahana Pendidikan*, 6(3), 91–98.

Kaiping, G., & Klamer, M. (2020). *Lexirumah: A lexical database of lesser-known Austronesian languages*. Lexirumah Project. Retrieved from https://lexirumah.model-ling.eu

Kaiping, G., Greenhill, S., & Klamer, M. (2019). Austronesian lexical data in Eastern Indonesia. *Data in Linguistics*, 6(3), 41–57.

Keraf, G. (1984). *Linguistik bandingan historis*. Nusa Indah.

Mahriyuni, M., Suryadi, & Yusniar. (2023). Lexicostatistics of Javanese and Sasak languages. *Mimbar Ilmu*, 28(1), 124–130.

Muhammad, R., & Hendrokumoro, S. (2022). Analisis kekerabatan bahasa Bali dan Sasak. *Jurnal Bahasa*, 5(1), 56–64.

Salahuddin. (2023). Lexicostatistics calculation on Manggarai and Bima languages. *Parole: Journal of Linguistics and Education*, 13(1), 33–48.

Septian, R., & Rahmat, A. (2022). Lexicostatistical comparison between Banjar language and Bakumpai dialect in South Kalimantan. *Journal of Language and Culture Studies, 14*(2), 102–115.

Setiawan, E. (2020). Leksikostatistik bahasa Osing dan Jawa. *Jurnal Bahasa dan Sastra*, 25(2), 145–153.

Swadesh, M. (1952). Lexico-statistics dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society*, 96(4), 452–463.

Yuliana, A., & Hasan, H. (2021). Lexical comparison of Dayak Bakumpai and Dayak Ngaju languages using the Swadesh list. *Jurnal Bahasa dan Sastra Nusantara, 9*(3), 221–235.

Sari, M., & Arifin, T. (2023). Glottochronology and Austronesian language divergence in Borneo: A preliminary study. *Indonesian Linguistics Research Journal, 5*(1), 45–60.