Lexical Relationship Between Dayak Ngaju and Dayak Sampit Languages: A Lexicostatistical Study

Julie Abiya Putri[1], Imam Qalyubi[2]
[1,2]Universitas Palangka Raya, Indonesia
E-mail: julieabiya170@gmail.com

Article Info
Article History
Received: 2025-05-13
Revised: 2025-06-23
Published: 2025-07-10

Keywords:
Lexical Relationship;
Dayak Ngaju;
Dayak Sampit;
Lexicostatistics;
Cognates;
Language Family.

Abstract
This study investigates the lexical relationship between Dayak Ngaju and Dayak Sampit languages in Central Kalimantan using a lexicostatistical approach. A 200-item Swadesh list was used to compare basic vocabulary obtained from a bilingual speaker fluent in both languages. The data were analyzed by identifying cognate pairs and calculating the percentage of shared vocabulary. The findings reveal that 155 out of 200 words (77.5%) are cognates, indicating a strong lexical relationship. This percentage places the two languages within the same language family based on commonly used classification thresholds in lexicostatistics. The analysis further categorizes the cognates into four types: identical forms, regular phonemic correspondences, phonetically similar pairs, and pairs with one phoneme difference. These results provide evidence of a close historical and linguistic connection between the two languages and contribute to the understanding of language variation and development among the Dayak languages in Indonesia.

Artikel Info
Sejarah Artikel
Diterima: 2025-05-13
Direvisi: 2025-06-23
Dipublikasi: 2025-07-10

Kata kunci:
Hubungan Leksikal;
Dayak Ngaju;
Dayak Sampit,
Leksikostatistik;
Serumpun;
Rumpun Bahasa.

Abstrak
Penelitian ini menyelidiki hubungan leksikal antara bahasa Dayak Ngaju dan bahasa Dayak Sampit di Kalimantan Tengah dengan menggunakan pendekatan leksikostatistik. Daftar Swadesh yang berisi 200 item digunakan untuk membandingkan kosakata dasar yang diperoleh dari seorang penutur dwibahasa yang fasih dalam kedua bahasa tersebut. Data dianalisis dengan mengidentifikasi pasangan kata serumpun dan menghitung persentase kosakata yang sama. Temuan tersebut mengungkapkan bahwa 155 dari 200 kata (77,5%) merupakan kata serumpun, yang menunjukkan hubungan leksikal yang kuat. Persentase ini menempatkan kedua bahasa tersebut dalam keluarga bahasa yang sama berdasarkan ambang batas klasifikasi yang umum digunakan dalam leksikostatistik. Analisis tersebut selanjutnya mengkategorikan kata serumpun tersebut menjadi empat jenis: bentuk identik, korespondensi fonemik teratur, pasangan yang mirip secara fonetis, dan pasangan dengan satu perbedaan fonem. Hasil ini memberikan bukti adanya hubungan historis dan linguistik yang erat antara kedua bahasa tersebut dan berkontribusi pada pemahaman tentang variasi dan perkembangan bahasa di antara bahasa-bahasa Dayak di Indonesia.

I. INTRODUCTION

Indonesia is widely recognized as a multicultural and multilingual country (Indrariani, 2017). It comprises hundreds of ethnic groups spread across thousands of islands, resulting in a rich diversity of languages spoken throughout the archipelago. Language in Indonesia holds a significant role beyond being a medium of communication. It also serves as an essential marker of identity for various communities (Mayangsari & Inderajati, 2023). According to Fishman (1999), language plays a pivotal role in the maintenance of group identity and intergenerational cultural transmission, especially in multilingual societies. The diversity of regional languages arises from several factors such as population migration, the geographical characteristics of the Indonesian archipelago with its many islands and mountainous regions, as well as social interaction and communication patterns (Collins, 2014). These factors contribute to the complexity and richness of Indonesia's linguistic landscape. As noted by Kramsch (1998), the relationship between language and culture is deeply intertwined, and linguistic diversity reflects not only historical processes but also cultural resilience.

One of the dominant ethnic groups in Central Kalimantan is the Dayak community, which consists of various subgroups each with distinct languages and dialects, including Ngaju language. The Dayak people refer to themselves based on their geographical origins, often naming their groups after rivers, historical heroes, or natural landmarks (Darmadi, 2016). Among the languages spoken by the Dayak people is Sampit

language, predominantly used in the city of Sampit, Central Kalimantan. The existence of multiple related languages within the Dayak ethnic group illustrates the linguistic diversity within a relatively limited geographical area. As noted by Crystal (2000), such internal variation within an ethnolinguistic group is not uncommon and often serves as evidence of long-standing historical, ecological, and social separations among communities.

Languages, like their speakers, undergo continuous processes of change and development over time. Advances in transportation and communication technology have increased the intensity of contact among speakers of different languages, influencing both cultural and linguistic domains. This increased contact often results in borrowing and the adoption of certain terms, making language change a dynamic and ongoing process (Sigiro, 2015). Thomason and Kaufman (1988) argue that language contact is one of the primary drivers of linguistic change, particularly in multilingual environments like Indonesia. Such phenomena highlight the importance of studying language kinship to understand the historical relationships and development of related languages.

This study aims to examine the lexical relationship between Dayak Ngaju and Dayak Sampit languages. By investigating the degree of kinship between these two languages, this research seeks to contribute to the understanding of language variation and historical linguistics within the Indonesian context. The results are expected to provide insight into how these languages have evolved, diverged, and maintained connections throughout their histories. Moreover, it aligns with the broader efforts in comparative linguistics, which, according to Campbell (2013), help reconstruct the linguistic past and reveal patterns of human migration and cultural interaction.

II. METHOD

This study adopts a descriptive quantitative research design, utilizing a lexicostatistical approach to examine the lexical relationship between the Dayak Ngaju and Dayak Sampit languages. Lexicostatistics, a method developed by Morris Swadesh (1952), is widely recognized in historical linguistics as an effective tool for measuring degrees of linguistic relatedness based on the proportion of shared basic vocabulary. Through this approach, the study aims to quantify the level of cognate vocabulary between the two languages and thereby determine their genealogical proximity.

The data were collected using a 200-item basic vocabulary list derived from the Swadesh list, which includes universal lexical items such as pronouns, numbers, body parts, natural elements, and everyday verbs. These items are considered relatively resistant to borrowing and therefore serve as reliable indicators of genetic linguistic relationships (Crowley & Bowern, 2010). Data collection was conducted through direct interviews with a key informant who is bilingual in both Dayak Ngaju and Sampit. The selection of the informant followed the criteria proposed by Mahsun (2011), which emphasize linguistic competence, sociocultural familiarity, and psychological and physical well-being. The informant was over 40 years old, born and raised in the relevant language community, had not lived outside the region for extended periods, and demonstrated consistent use and loyalty to the local language in daily interactions.

During the interview process, the researcher prompted the informant with the vocabulary list, and the responses were recorded and transcribed. These transcriptions were then compiled into a comparative lexical list for analysis. This method ensures the reliability of the data while preserving the natural phonological and morphological characteristics of each language as spoken by native users (Hock & Joseph, 2009).

For data analysis, the study employed a systematic lexicostatistical comparison. Each lexical item from the two languages was evaluated and classified into one of several categories of cognacy. These included identical pairs (where words are the same in both pronunciation and form), pairs showing regular phonemic correspondences (systematic sound changes), phonetically similar pairs (those exhibiting articulatory resemblance), and pairs differing by a single phoneme but still recognizable as cognates due to predictable phonological variation. This classification process is grounded in the comparative method principles as outlined by Campbell (2013), who emphasizes the importance of identifying systematic correspondences rather than relying solely on surface similarity.

Following the classification, the number of cognate word pairs was tallied and analyzed using the formula proposed by Swadesh to determine the percentage of lexical similarity:

$$\mathbf{C = (k / n) \times 100\%},$$

where *C* is the percentage of shared cognates, *k* is the number of cognate pairs, and *n* is the total number of vocabulary items (200).

Based on the resulting percentage, the degree of relatedness between the two languages was classified into one of several categories, such as Language (100–81%), Family (81–36%), Stock (36–12%), and so forth, in accordance with the lexicostatistical classification scale. These thresholds provide a heuristic framework for interpreting the level of linguistic kinship (Gudschinsky, 1956), although they should be applied with caution, as linguistic evolution is influenced by both internal developments and external contact phenomena.

Through this analytical framework, the study seeks not only to measure lexical similarities quantitatively but also to contribute to the broader discourse on language classification and historical relationships among Austronesian languages in Indonesia.

III. RESULTS AND DISCUSSION

A. Research Results

The analysis of the 200 basic vocabulary items from Dayak Ngaju and Sampit languages produced a classification of word pairs based on their lexical relatedness. These word pairs were grouped into two main categories: cognate pairs and non-cognate pairs.

The results are presented in the following table:

Table 2. Word Pairs in Dayak Ngaju and Sampit Languages

No.	Glos	Dayak Ngaju	Sampit	Kognat
1	One	Ije	Eje	✓
2	Two	Dua	Due	✓
3	Three	Tilu	Telu	✓
4	Four	Epat	Epat	✓
5	Five	Lime	Leme	✓
6	Six	Jahawen	Jahawen	✓
7	Seven	Uju	Uju	✓
8	Eight	Hanya	Hanya	✓
9	Nine	Jalatien	Tiyen	✓
10	Ten	Sepulu	Sapuluh	✓
11	One Hundred	Saratus	Seratus	✓
12	One Thousand	Sekuyan	Sekoyan	✓
13	I	Aku	Aku	✓
14	You	Ikau	Ikau	✓
15	She/He	Ie	Inya	—
16	We (Inclusive)	Itah	Itah	✓
17	We (Exclusive)	Ikei	Ekei	✓
18	You (Plural)	Ketun	Ketoh	✓
19	They	Ewen	Elen	✓
20	What	Narai	Een	—
21	Who	Eweh	Jawen	—
22	When	Pea	Preya	✓
23	How much	Pire	Pere	✓
24	How	Kilenampih	Kilanampi	✓
25	Where	Kueh	Kueh	✓
26	Earlier	Endau	Tendau	✓
27	Just now	Haru	Hanyar	✓
28	Used to	Helu	Hilu	✓
29	Tomorrow	Jewu	Dapit	—
30	Later	Kareh	Kareh	✓
31	Before	Sehindai	Sehindai	✓
32	Yes	Iyuh	Ya	✓
33	No	Dia	Bare	—
34	Do not	Ela	Ela	✓
35	Can	Ulih	Oleh	✓
36	That	Jite	Ite	✓
37	There	Hekaw	Kanih	—
38	Here	Hetuh	Titun	—
39	There	Hete	Hete	✓
40	For	Akan	Akan	✓
41	Like that	Kute	Kilaute	✓
42	Such as	Kilau	Kilau	✓
43	Behind	Likut	Likul	✓
44	There is	Tege	Ada	—
45	Ther is no	Jatun	Bareda	—
46	Out of	Lepah	Lepah	✓
47	Far	Kejau	Kejau	✓
48	Near	Tukep	Tokep	✓
49	Right	Gantau	Kanan	—
50	Left	Sambil	Kiwa	—
51	All	Uras	Huras	✓
52	If	Amun	Amun	✓
53	With	Dengan	Dengan	✓
54	Is	Yete	Yate	✓
55	Drink	Mihup	Mihup	✓
56	Eat	Kuman	Kuman	✓
57	Sleep	Batiruh	Batiruh	✓
58	To hear	Mahining	Mahining	✓
59	To smell	Membewau	Manciu m	—
60	Breath	Manahaseng	Manahanseng	✓
61	Laugh	Tatawe	Tatawe	✓
62	Standing	Mendeng	Mendeng	✓
63	Lie down	Menter	Menter	✓
64	Sit	Munduk	Munduk	✓
65	Get up	Misik	Minsik	✓
66	Running	Hadari	Hadari	✓
67	To open	Mukei	Mambuka	—
68	To carry	Maimbit	Maimbit	✓
69	To push	Macung	Manunj	—

			ul	
70	To cut	Manetek	Manetek	✓
71	To dig	Mangali	Manggali	✓
72	To split	Manyila	Manyila	✓
73	To kill	Mampatei	Mampatei	✓
74	To Scratch	Manggayau	Manggaru	✓
75	To plant	Mimbul	Mananam	—
76	To choose	Mintih	Mamilih	—
77	To tie	Mameteng	Maikat	—
78	To count	Mise	Mahitung	—
79	To know	Mangatawan	Mangatawan	✓
80	To buy	Mamili	Mamili	✓
81	To Sweep	Mapas	Manyapu	—
82	To read	Membasa	Mambaca	✓
83	To fish	Mamisi	Mamisi	✓
84	Go up	Mandai	Mandai	✓
85	Go down	Muhun	Muhun	✓
86	To enter	Tame	Tame	✓
87	To burn	Mamapoi	Mangabuk	—
88	To throw away	Manjakah	Mambuang	—
89	To invite	Marawei	Mengundang	—
90	To go home	Buli	Buli	✓
91	To hit	Mahantuk	Mahantu	✓
92	To cook	Mampakasak	Merapi	—
93	To bite	Mamangkit	Mamangkil	✓
94	To steal	Manakau	Manakau	✓
95	To stay	Melai	Balihi	—
96	To apply powder	Bapupur	Bapupur	✓
97	Want	Hakun	Hakun	✓
98	Swim	Hanangui	Bakunyung	—
99	Quiet	Benyem	Suni	—
100	Will	Cagar	Cagar	✓
101	Wide-open eyes	Menceleng	Menceleng	✓
102	Come	Dumah	Dumah	✓
103	Paddle	Esei	Besei	✓
104	Day	Andau	Andau	✓
105	Year	Nyelu	Nyilu	✓
106	Late Afternoon	Halemei	Jalemei	✓
107	Night	Hamalem	Malem	✓
108	People	Uluh	Uluh	✓
109	Man	Hatue	Hatue	✓
110	Woman	Bawi	Bawi	✓
111	Name	Aran	Aran	✓
112	God	Hatala	Tuhan	—
113	Traditional Leader	Damang	Damang	✓
114	Dad	Bapa	Bapa	Yes
115	Mother	Umai	Uma	✓
116	Younger sibling	Andi	Ading	✓
117	Older sibling	Kaka	Kaka	Yes
118	Uncle	Mama	Amang	—
119	Aunt	Mina	Acil	—
120	Grandfather	Bue	Kai	—
121	Grandmother	Tambi	Nini	—
122	The youngest child	Tambusu	Busu	✓
123	The eldest child	Tambakas	Pambakas	✓
124	Child	Anak	Anak	✓
125	Relative	Pahari	Pahari	✓
126	Nephew/ Niece	Aken	Aken	✓
127	Head	Kuluk	Kolok	✓
128	Teeth	Kasinga	Kesingi	✓
129	Nose	Urung	Hidung	—
130	Neck	Uyat	Panggu w	—
131	Hand	Lenge	Lenge	✓
132	Foot	Pai	Pai	✓
133	Hair	Balau	Balau	✓
134	Ear	Pinding	Pinding	✓
135	Skin	Upak	Kulit	—
136	Heart	Atei	Atei	✓
137	Cat	Pusa	Kucing	—
138	Dog	Asu	Asu	✓
139	Chicken	Manuk	Manuk	✓
140	Frog	Bekatak	Katak	✓
141	Bee	Penyangat	Penyangat	✓
142	Rat	Balawau	Tikus	—
143	Snake	Handipe	Handepe	✓
144	Swan	Gangsa	Angsa	✓
145	Fish	Lauk	Lauk	✓
146	Spider	Sabangkang	Bangkang	✓
147	Monkey	Bakei	Bakei	✓
148	Water	Danum	Danum	✓
149	Land	Petak	Petak	✓
150	Thunder	Guntur	Guntur	✓
151	Wind	Riwut	Angin	—
152	Bamboo	Puring	Paring	✓
153	Grass	Uru	Uru	✓
154	Fire	Apui	Apui	✓
155	Root	Uhat	Uhat	✓
156	Rattan	Uei	Uei	✓
157	Leaf	Dawen	Daun	✓
158	Durian	Dahuyan	Duhiyan	✓
159	Knife	Lading	Pisau	—
160	Pants	Selawar	Selawar	✓
161	Needle	Pilus	Pilus	✓

162	Salt	Uyah	Uyah	✓
163	House	Huma	Huma	✓
164	Road	Karatak	Jalan	—
165	Message	Peteh	Peteh	✓
166	Hoe	Cangkul	Cangkul	✓
167	Water tray	Ceper	Ceper	✓
168	Kitchen	Dapur	Dapur	✓
169	Wall	Dinding	Dinding	✓
170	Boat	Jukung	Jokong	✓
171	Red	Bahandang	Bahandang	✓
172	Yellow	Bahenda	Bahenda	✓
173	Black	Babilem	Babalem	✓
174	White	Baputi	Baputi	✓
175	Green	Bahijau	Hijau	✓
176	Afraid	Mikeh	Mekeh	✓
177	Brave	Bahanyi	Bahanyi	✓
178	Shy	Mahamen	Mahamen	✓
179	Small	Kurik	Kurik	✓
180	Big	Hai	Datuh	—
181	Long	Panjang	Panjang	✓
182	Short	Pandak	Pandak	✓
183	New	Harue	Hanyar	✓
184	Old	Tahi	Tahi	✓
185	Hungry	Belawu	Belawu	✓
186	Full	Besuh	Bensuh	✓
187	Hot	Balasut	Balasut	✓
188	Cold	Sadingen	Sadingen	✓
189	Many	Are	Are	✓
190	Few	Isut	Isut	✓
191	Bright	Tarang	Tarang	✓
192	Work on	Gawi	Gawi	✓
193	Dark	Kaput	Kadap	✓
194	Sick	Pehe	Kapehe	✓
195	Clean	Rasih	Barasih	✓
196	Healthy	Barigas	Sehat	—
197	Good	Bahalap	Bengkeng	—
198	Dirty	Papa	Rigat	—
199	Great	Tamam	Jagaw	—
200	delicious	Tutu	Nyaman	—
Total Cognate Words				155

Based on the data, there are 155 pairs that are classified as cognates. These include identical word forms, words with regular phonemic correspondences, phonetically similar forms, and pairs that differ by only one phoneme. To determine the degree of lexical relatedness, a standard lexicostatistical formula was applied:

$$c = \frac{k}{n} \times 100\%$$

Where:

C = percentage of lexical relatedness

k = number of cognate pairs (155)

n = total number of vocabulary items compared (200)

Substituting the values:

$$c = \frac{155}{200} \times 100\% = 77.5\%$$

Thus, the degree of lexical relatedness between Dayak Ngaju and Sampit languages is 77.5%. Based on commonly used classification thresholds in lexicostatistical studies, this percentage falls within the *family* level of relatedness (see Table 1). This suggests that the two languages are closely related and belong to the same language family.

B. Discussion

The results of this study reveal a significant lexical similarity between Dayak Ngaju and Sampit languages, with 155 out of 200 basic vocabulary items classified as cognates, yielding a 77.5% degree of relatedness. This percentage places the two languages within the *family* level of lexical relatedness, indicating a close historical and linguistic relationship. The cognate pairs identified in this study can be categorized into four main types, reflecting different levels of phonological similarity:

1. Identical Cognates

   These are word pairs that match exactly in both pronunciation and spelling. Examples include *Mihup-Mihup* and *Sehindai-Sehindai*. Such pairs strongly indicate that these words have been retained unchanged in both languages, reflecting a very close connection and shared linguistic heritage.

2. Cognates with Phonemic Correspondences

   This category includes word pairs where systematic and regular phonemic changes occur between Dayak Ngaju and Sampit, suggesting historical sound shifts rather than coincidental resemblance. Examples are *Endau-Tendau* and *Pea-Preya*. The presence of these regular phonemic correspondences reinforces the conclusion that these languages descend from a common ancestor, with predictable phonological evolution shaping their vocabularies.

3. Phonetically Similar Pairs

   Some word pairs do not follow clear phonemic correspondences but show strong phonetic resemblance. For instance,

Manggayau and *Manggaru* share an identical initial segment "Mangga" while their endings "-yau" and "-ru" differ but exhibit similar vowel resonance and articulatory features. Although these pairs lack systematic sound correspondences, their overall phonetic similarity suggests a historical connection, possibly reflecting irregular sound changes or dialectal variation.

4. Cognates with One Phoneme Difference

This group comprises word pairs differing by only a single phoneme, likely due to dialectal variation, phonological assimilation, or environmental influence. Examples include *Pire- Pere* and *Ewen-Elen*. Despite the minor phonemic difference, the overall structure and sound patterns remain highly similar, supporting their classification as cognates. This category highlights how small phonetic shifts can occur without disrupting the recognizability of related words across closely related languages.

On the other hand, the 45 non-cognate pairs identified show either no resemblance or differences too great to suggest a common origin within the scope of basic vocabulary. This may be explained by language contact, borrowing, or independent innovations that have diversified the vocabularies. Overall, the 77.5% lexical similarity between Dayak Ngaju and Sampit reflects a substantial degree of shared vocabulary, consistent with their classification as members of the same language family. The variety in types of cognates, from identical forms to those with phonemic correspondences or slight phonetic differences, provides insight into the historical phonological processes that have shaped both languages over time.

These results also underline the importance of detailed phonetic and phonological analysis in lexicostatistical studies, as recognizing different types of cognates allows for a more nuanced understanding of language relationships beyond mere lexical counting. Future research could further investigate syntactic and morphological similarities, as well as explore dialectal variation within each language, to provide a more comprehensive picture of their linguistic kinship.

IV. CONCLUSIONS AND SUGGESTIONS

A. Conclusion

This study has demonstrated a significant lexical relationship between Dayak Ngaju and Dayak Sampit languages, with a 77.5% similarity based on a 200-word Swadesh list. This level of similarity places the two languages within the same language family, indicating a close historical connection. The identification of various types of cognates, including identical forms, regular phonemic correspondences, phonetically similar words, and those with only one phoneme difference, supports the conclusion that these languages have diverged from a common ancestral language while retaining a strong degree of lexical similarity. These findings contribute to the broader understanding of linguistic kinship within the Dayak language group and highlight the value of lexicostatistical methods in analyzing historical language development in Indonesia's multilingual context.

B. Suggestion

Based on the findings of this study, it is suggested that further research be conducted on other Dayak subgroups using similar lexicostatistical methods to map the broader linguistic relationships within the Dayak language family. Comparative studies involving more dialects and larger speaker populations could provide deeper insights into historical language development, patterns of divergence, and regional language preservation efforts.

REFERENCES

Collins, J. T. (2014). Keragaman bahasa dan kesepakatan masyarakat: Pluralitas dan komunikasi. 180–194.

Darmadi, H. (2016). Dayak asal-usul dan penyebarannya di Bumi Borneo (1). Sosial Horizon: Jurnal Pendidikan Sosial, 3(2), 322–340. https://journal.ikippgriptk.ac.id/index.php/sosial/article/view/376

Fernandes, I. Y. (1990). Telaah kualitatif dan kuantitatif linguistik komparatif (Beberapa masalah teori dan praktik). Bahan Kuliah Linguistik Historis Komparatif.

Hilmi, M. (2017). Leksikostatistik bahasa (Sasambo) Bahasa Sasak, Bahasa Sumbawa/Samawa, dan Bahasa Bima/Mbojo: Kajian Linguistik Historis

Komparatif. Fondatia, 1(1), 165–176. https://doi.org/10.36088/fondatia.v1i1.94

Indrariani, E. A. (2017). Leksikostatistik Bahasa Jawa dan Bahasa Sunda. PIBSI, 3(9), 1–10. https://core.ac.uk/download/pdf/151236694.pdf

Keraf, G. (2010). Linguistik bandingan historis. Nusa Indah.

Mahsun. (2011). Dialektologi diakronis. Yogyakarta: Gadjah Mada University Press.

Mayangsari, D., & Inderajati, A. (2023). Leksikostatistik dan glotokronologi Bahasa Berau dengan Bahasa Minangkabau. Widyaparwa, 51(2), 377–391. https://doi.org/10.26499/wdprw.v51i2.1410

Rachma, E. P. I. (2019). Analisis kekerabatan Bahasa Jawa dan Bahasa Madura: Leksikostatistik. The Relationship Between Javanese and Madura Language, 1–14.

Ruriana, P. (2018). Hubungan kekerabatan Bahasa Jawa dan Madura. Kandai, 14(1), 15–28. https://doi.org/10.26499/jk.v14i1.512

Sigiro, E. P. (2015). Kekerabatan Bahasa Tamuan, Waringin, Dayak Ngaju, Kadorih, Maanyan, dan Dusun Lawangan. Jurnal Linguistik, 1–14.