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/

Comparing Basic Words and Using Lexicostatistics to Identify Language Connections among Indonesian, Dayak Ngaju, and Javanese

Tri Aprilia Suciati[1], Imam Qalyubi[2], Misrita[3]
[1,3] Universitas Palangka Raya, [2] Universitas Islam Negeri Palangka Raya
Central Kalimantan- Indonesia
Email: [1] *suciatitriaprilia@gmail.com* , [2]*imamqalyubi@gmail.com*,
[3]*ritakalang3@gmail.com*

Article History:
Received: 04/06/2025
Accepted: 22/11/2025
Published: 08/01/2026

Keywords:
lexicostatistics, Indonesian, Dayak Ngaju, Javanese, linguistic kinship

Abstract

This study discusses the degree of kinship between Indonesian, Dayak Ngaju, and Javanese languages using a lexicostatistical approach. This method is employed to determine the linguistic relationships based on the number of basic vocabulary items that share similarities in form and meaning. Based on data from 15 glosses, the results show a 100% cognate rate among the three languages, indicating a strong historical connection as part of the Austronesian language family.

Kata Kunci:
leksikostatistik, bahasa indonesia, dayak ngaju, bahasa jawa, kekerabatan linguistik

Abstrak

Penelitian ini membahas tentang derajat kekerabatan antara bahasa Indonesia, bahasa Dayak Ngaju, dan bahasa Jawa dengan menggunakan pendekatan leksikostatistik. Metode ini digunakan untuk menentukan hubungan kebahasaan berdasarkan jumlah kosakata dasar yang memiliki kesamaan bentuk dan makna. Berdasarkan data dari 15 glosarium, hasilnya menunjukkan tingkat kekerabatan 100% di antara ketiga bahasa tersebut, yang menunjukkan adanya hubungan historis yang kuat sebagai bagian dari rumpun bahasa Austronesia.

INTRODUCTION

Language is a fundamental tool of human communication that not only conveys ideas but also records the history and culture of societies. In historical comparative linguistics, the relationships among languages can be analyzed through lexicostatistics—a quantitative method comparing basic vocabulary items between languages to determine their linguistic kinship.

The Indonesian, Dayak Ngaju, and Javanese languages are all part of the Austronesian language family, though they originate from different branches and represent diverse cultural and historical trajectories. Indonesian, derived from Malay, has undergone significant lexical modernization and standardization since becoming the national language. Javanese, with its stratified speech levels and ancient literary tradition, reflects deep historical roots. Dayak Ngaju, spoken by indigenous communities in Central Kalimantan, retains many archaic features and is less influenced by national standardization policies.

The importance of studying linguistic kinship lies in uncovering the historical development and diffusion of languages. Recent linguistic studies have emphasized that similarities in lexicon and phonological structure can be statistically measured to determine common ancestry. Such measurements also reflect cultural exchanges, migration patterns, and linguistic convergence over time.

METHOD

This research employed a descriptive quantitative method using a modified version of the Swadesh list consisting of 15 basic vocabulary items. The data were collected from three language sources: standard Indonesian, Dayak Ngaju (a language spoken in Central Kalimantan), and Javanese (primarily spoken in Java).

Each pair of words was compared based on phonological and morphological similarity to determine whether the items were cognates. Criteria for identifying cognates included:

1. Identical forms
2. Regular phonemic correspondences
3. Significant phonetic resemblance

The lexicostatistical formula used for calculating the percentage of kinship is as follows:

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

Where:

C = percentage of kinship
k = number of cognate pairs
n = total number of glosses compared

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/

This formula helps identify the closeness of languages based on quantitative measures, allowing researchers to map the historical connections between languages across time and space

This part explains the rationale for the application of specific procedures, which includes research design, participants (population and sample), data collection technique, research instrument(s), data analysis technique.

FINDINGS AND DISCUSSION

Findings

Based on the analysis of 15 basic vocabulary items, all word pairs across the three languages were identified as cognates. Thus, the lexicostatistical calculation is as follows:

1. Cognate pairs (k) = 15
2. Total glosses (n) = 15

$$C = (15/15) \times 100\% = 100\%$$

This indicates a high degree of kinship among the three languages, at least within this limited dataset. The data support the hypothesis that these languages descended from a common ancestral language within the Austronesian family.

Table 1. Examples of Cognate Pairs

No	Gloss	Indonesian	Dayak Ngaju	Javanese	Cognate
1	One	Satu	Ije	Siji	✓
2	Two	Dua	Dua	Loro	✓
3	Three	Tiga	Telu	Telu	✓
4	Four	Empat	Epat	Papat	✓
...	...	...	...	...	...

Based on comparative observations, several linguistic relationships can be identified. Indonesian and Javanese display relatively regular sound correspondences, such as the shift from /b/ to /w/, along with shared morphological features that indicate a close historical and structural relationship. In contrast, a comparison between Indonesian and Dayak Ngaju shows that Dayak Ngaju tends to preserve more conservative linguistic forms, while also exhibiting some unique innovations. Meanwhile, the relationship between Dayak Ngaju and Javanese

reveals fewer direct parallels; however, certain similarities in root forms can still be identified, suggesting a genealogical connection among the three languages.

Several notable phonological phenomena emerge from the analysis, including prothesis, vowel shifting, and phonemic simplification. Prothesis is evident in the addition of initial sounds to words, which functions to ease articulation and maintain syllable structure across related forms. Vowel shifting reflects systematic changes in vowel quality over time, often influenced by stress patterns, surrounding consonants, or broader phonotactic constraints. Phonemic simplification, meanwhile, involves the reduction or merger of certain sounds, leading to more streamlined phonological systems. Collectively, these phenomena are consistent with well-documented patterns in Austronesian diachronic linguistics, where sound change operates in regular and predictable ways, reinforcing the historical and genetic relationships among the languages under comparison.

Discussion

Apheresis refers to a phonological process in which one or more sounds at the beginning of a word are removed or reduced, resulting in a shorter or altered phonetic form. This process can be observed in several cognate words across Indonesian, Dayak Ngaju, and Javanese, indicating systematic sound change as well as historical linguistic relationships among these languages. For instance, the Indonesian word *satu* corresponds to *ije* in Dayak Ngaju and *siji* in Javanese. In this case, the initial consonant /s/ present in Indonesian and Javanese is absent in the Dayak Ngaju form, reflecting a process of initial sound loss. A similar pattern can be seen in the word *empat* in Indonesian, which appears as *epat* in Dayak Ngaju and *papat* in Javanese. Here, the initial consonant /m/ in Indonesian is reduced or removed in the Dayak Ngaju form, while Javanese shows a different development with the addition or strengthening of the initial /p/. Overall, these examples demonstrate how apheresis contributes to phonetic variation and word shortening, while still preserving recognizable semantic and etymological connections across the three languages.

Then, apocope refers to a phonological process in which one or more sounds at the end of a word are omitted, resulting in a shorter and more economical form of pronunciation. This phenomenon can be clearly observed in the comparative data presented in the table. For example, the Indonesian word *berapa* undergoes apocope in Dayak Ngaju, becoming *pire*, where the final vowel and syllable are omitted, while in Javanese it appears as *pira*, showing a similar reduction with slight phonological adjustment. Likewise, the Indonesian word *kecil* experiences the loss of the final consonant /l/ in Dayak Ngaju, resulting in *kece*, whereas in Javanese it becomes *cil*, reflecting an even more reduced form. These examples demonstrate that the omission of final sounds, such as /ba/ in *berapa* or /l/ in *kecil*, leads to shorter word forms and illustrates how apocope functions as a common phonological process across these languages, albeit with different outcomes depending on each language's phonological system and historical development.

Moreover, syncope refers to a phonological process in which one or more sounds located in the middle of a word are omitted, resulting in a shorter and more economical pronunciation. This type of sound change commonly occurs in the natural development of languages as speakers tend to simplify articulation without significantly affecting meaning. In the present comparison, syncope can be clearly observed across Indonesian, Dayak Ngaju, and Javanese. For instance, the Indonesian word *bekerja* undergoes syncope in Dayak Ngaju, becoming *gawi*, and in Javanese, appearing as *gawe*. In this case, the medial sequence */ker/* is omitted, producing a more concise phonetic form while retaining the original semantic content related to "working." Similarly, the Indonesian word *tertentu* corresponds to *tertento* in Javanese, where the medial sound */an/* is reduced or omitted, resulting in a simplified pronunciation. These examples demonstrate that syncope plays a significant role in shaping phonetic variation among related languages, highlighting both shared historical roots and language-specific phonological developments.

The phonetic change process in Indonesian, Dayak Ngaju, and Javanese can be observed through systematic modifications in word forms, particularly involving the loss or alteration of sounds at the beginning (apheresis), in the middle (syncope), or at the end of words (apocope). These phonetic changes are not random but reflect

broader patterns of language evolution, in which speakers tend to simplify articulation to achieve greater efficiency and ease of pronunciation. In addition, such changes are influenced by local dialects, speech rhythms, and phonological constraints specific to each language. Over time, these adaptations become conventionalized within speech communities, resulting in stable phonetic variants that distinguish one language or dialect from another while still preserving traces of their shared historical origins.

These phonetic changes show how regional languages influence each other and adapt to the needs of everyday communication. The omission of sounds in local languages can shorten words that were originally longer, speed up pronunciation, and create phonetic variations between languages. Thus, from the explanation of the phonetic changes that occur, we can see that Aphesis, Apocrypha, and Syncope play an important role in the formation of word variations in Indonesian, Dayak Ngaju, and Javanese, and provide insight into the evolution of phonetics in linguistic societies.

Prothesis refers to a phonological process in which an additional sound is inserted at the beginning of a word, often as a result of historical sound change or adaptation to the phonotactic rules of a particular language. Based on the data presented in the table, several word comparisons can be observed to illustrate this phenomenon. One example is the word *bamboo,* which appears as bamboo in Dayak Ngaju and *pring* in Javanese. In this case, the Javanese form does not demonstrate prothesis in the strict sense, as there is no added initial sound. Instead, the difference reflects a broader phonological and lexical change that results in a distinct word form. This indicates that not all variations at the word-initial position can be classified as prothesis; some represent independent lexical developments or historical sound shifts. Therefore, careful analysis is required to distinguish true instances of prothesis from other types of phonological or morphological change in comparative linguistic studies.

Paragogue refers to a type of phonetic change in which an additional sound is attached to the end of a word. This phenomenon often occurs as a result of phonological adaptation, ease of pronunciation, or historical language development. In the comparative data presented in the table, several examples from Dayak Ngaju and Javanese illustrate the occurrence of paragogue, where an extra sound appears at the word-final position, signaling a phonetic modification between the two languages. One example can be observed in the word *lime* in Dayak Ngaju, which corresponds to *limo* in Javanese. This change reflects the addition of a final vowel sound, resulting in a slightly altered word form while maintaining the same lexical meaning. Another example is *sepulu* in Dayak Ngaju and *sedasa* in Javanese. In this case, the presence of the final vowel *–u* in the Dayak Ngaju form indicates a word-final sound addition when compared to its Javanese counterpart. Despite these examples, the occurrence of paragogue in the data set appears to be relatively limited. Most of the compared lexical items do not exhibit the addition of sounds at the end of words, suggesting that paragogue is not a dominant phonetic process in the relationship between Dayak Ngaju and Javanese. Instead, other types of sound changes, such as substitution or deletion, seem to occur more frequently, making paragogue a relatively minor pattern within the overall phonological comparison.

Epenthesis refers to a phonetic process in which an additional sound is inserted into the middle of a word, often to ease pronunciation or to conform to the phonological patterns of a particular language. This phenomenon can be observed in the comparison between Dayak Ngaju and Javanese, as illustrated by the example *head* in Dayak Ngaju and its corresponding form in Javanese. Although the basic lexical form remains the same in both languages, the Javanese pronunciation exhibits a subtle phonetic modification. Specifically, an additional sound appears in the medial position, resulting in a pronunciation that is slightly longer or more diphthongized than in Dayak Ngaju. Importantly, this change does not alter the meaning of the word, nor does it involve a significant shift in its overall phonetic structure. In a broader perspective, epenthesis tends to occur more frequently in Javanese, where it commonly manifests as sound lengthening or the insertion of extra vocalic elements within a word. This reflects the phonological tendency of

Javanese to favor smoother and more sonorous sound patterns. In contrast, Dayak Ngaju generally retains shorter and more conservative phonetic forms, with fewer instances of medial sound insertion. The data also suggest the presence of an added /u/ sound at the end of certain Dayak Ngaju words, which might initially resemble paragoge (sound addition at the end of a word). However, based on the available table, clear and consistent examples of paragoge are limited. Most lexical items do not demonstrate systematic sound addition in word-final position. Therefore, while epenthesis is a noticeable and recurrent phonological process—particularly in Javanese—paragoge appears to be relatively rare or marginal in the dataset under analysis.

Metathesis is a type of phonetic change that involves the reordering or exchange of sound positions within a word, often without altering its core meaning. This phenomenon commonly occurs in languages that share historical or genealogical relationships, where phonological variation develops through time and regional usage. An example of metathesis can be observed in the comparison between *lime* in Dayak Ngaju and *limo* in Javanese. Although the phonetic realization of the word differs, particularly in the arrangement and articulation of vowel sounds, the lexical item remains clearly recognizable in both languages. This similarity indicates that the two forms originate from the same linguistic root. The variation in pronunciation reflects natural phonological adaptation influenced by regional speech patterns rather than a semantic shift, demonstrating how metathesis contributes to phonetic diversity while maintaining lexical continuity across related languages.

CONCLUSION

The lexicostatistical analysis of Indonesian, Dayak Ngaju, and Javanese languages using 15 basic vocabulary items reveals a 100% cognate rate. The data suggest a strong kinship and shared Austronesian origin. Broader research will enrich our understanding of historical language development in Indonesia.

The findings reveal that even with a small sample of glosses, core lexical items remain highly stable and provide clear insights into historical relationships among languages. Further implications involve understanding how language preservation and cultural transmission in isolated regions such as Kalimantan help maintain archaic structures.

REFERENCES

Eka, Ni Luh, I Wayan Suryasa, and I Wayan Budi Utama. "The Adaptation Strategy of Central Kalimantan's Dayak Ngaju Religious System to the State Official Religions in Palangka Raya." International Journal of Linguistics, Literature and Culture 7, no. 3 (2021): 111–119.

Farashwanti, N., and R. Anisa. "Potential Application of Batang Garing Cultural Values of Dayak Tribe in Orangutan Conservation." Journal of Sustainability, Society, and Eco-Welfare 2, no. 1 (2024): 49–61.

Karyani, Lida, Mangatur Rudolf Nababan, and Sri Marmanto. "Translation Analysis on Dayak Cultural Terms from Dayak Ngaju to Indonesian and English." Langkawi: Journal of The Association for Arabic and English 6, no. 1 (2020): 1–14.

Mayangsari, Dwi. "Lexicostatistics of Javanese and Sasak Languages: Comparative Historical Linguistic Studies." Mimbar Ilmu 25, no. 2 (2023): 45–58.

Putri, Dina Lutfiana, and Agus Subiyanto. "Konstruksi Verba Serial Direksional pada Bahasa Dayak Ngaju Kalimantan Tengah." Diglosia: Jurnal Kajian Bahasa, Sastra, dan Pengajarannya 6, no. 2 (2023): 413–424.

Wardana, Muhammad Kiki, Dwi Widayati, Rostina Taib, and Muhammad Iqbal. "Lexicostatistics of Malay, Tagalog and Ilocano Languages: A Comparisonal Historical Linguistic Study." Jurnal Education and Development 10, no. 3 (2022): 475–479.

