



P-ISSN 2355-2794
E-ISSN 2461-0275

Linguistic Relationship and Dialect Variation of Banten Javanese and Sundanese in Indonesia

Dase Erwin Juansah^{1*}
Akhmad Baihaqi²
Odien Rosidin¹
Erwin Salpa Riansi¹

¹Department of Indonesian Language Education, Faculty of Teacher Training and Education,
Universitas Sultan Ageng Tirtayasa, Serang 42117, INDONESIA

²Department of English Education, Faculty of Teacher Training and Education,
Universitas Sultan Ageng Tirtayasa, Serang 42117, INDONESIA

Abstract

Banten people in Indonesia speak two different local languages: Banten Javanese and Banten Sundanese. In reality, it demonstrates that both languages share similarities in phonological and morphological patterns. The purpose of this study is to investigate the language kinship and dialect variations that occur within these local languages. This research employed a case study that involved eight native speakers of Banten Javanese from Serang, as well as eight native speakers of Banten Sundanese from Pandeglang and Lebak. In this study, non-participant observation was used to investigate the dialect variation used among the respondents. Then, semi-structured interviews and documentation were used to analyze the language kinship and separation time between Banten Javanese and Sundanese. The data were analyzed using Miles et al.'s model in combination with Swadesh's lexicostatistical technique. The results revealed that (1) 51 pairs of identical words share both morphemic and phonemic similarities; (2) glottochronological estimates suggest that the separation of these local languages from the proto-language occurred between 1171 and 311 BC; and (3) regional dialects are the primary influence on the dialectal variation of these languages. Since these local languages are widely spoken in Banten, the findings reveal important linguistic relationships among local languages, emphasize the role of regional factors, and determine the value of glottochronology in tracing language development and variation.

Keywords: Dialect variation, glottochronology, kinship, local language.

* Corresponding author, email: daseerwin77@untirta.ac.id

Citation in APA style: Juansah, D. E., Baihaqi, A., Rosidin, O., Riansi, E. S. (2025). Linguistic relationship and dialect variation of Banten Javanese and Sundanese in Indonesia. *Studies in English Language and Education*, 12(3), 1106-1118.

Received April 27, 2025; Revised July 14, 2025; Accepted August 6, 2025; Published Online September 3, 2025

<https://doi.org/10.24815/siele.v12i3.45734>

Copyright © 2025 by Authors, published by *Studies in English Language and Education*. This is an open-access article distributed under the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>)

1. INTRODUCTION

Talking about language and society is always interesting because language is utilized and encountered in all aspects of human life. According to [Sudaryat \(2015\)](#), human culture and civilization cannot exist without language. Previously, [Koentjaraningrat \(2009\)](#) defined language as a symbol that describes the social groups of humans, both orally and in writing. These views believe that language is a basic requirement for people, and it plays a vital role in the countries and communities where they live.

Indonesia is a vast country with diverse inhabitants and tribes or social ethnicities, including Sundanese, Javanese, Batak, Sasak, among others. This diversity is reflected in the languages they use. [Bucholtz and Hall \(2005\)](#) observed that language is connected simultaneously with both the identification of the individuals talked to and the identity of the people speaking. For example, the Sundanese are identified by their use of Sundanese, whereas the Javanese use Javanese. Although each tribe or ethnic group inherits its native language, the Indonesian language serves as an official or national language that unifies all Indonesian people.

It is common for a language spoken by a single tribe or ethnic group to have distinct speech variants. These speech variances are commonly known as dialects. [Meyerhoff \(2018\)](#) described dialects as sub-varieties of a single language. Because speech communities are varied, dialects in a single language will always exist. The Sundanese, who primarily live in Banten and West Java Provinces in Indonesia, speak Sundanese as their mother tongue or first language. In addition to Sundanese language, Banten people also use Banten Javanese as their mother tongue. As a result, Banten Sundanese is predominantly spoken in Pandeglang and Lebak Regencies, while Banten Javanese is primarily used in Serang Regency, Serang City, and Cilegon City.

Most Banten people speak these two local languages as their mother tongue or first language. Although these are two separate languages, they have similarities in phonological and morphological patterns, as well as semantics ([Maemunah, 2017](#)). Since Banten Javanese and Sundanese share a common origin for some phonological and morphological patterns, they were most likely considered to have evolved from a similar proto-language. They are presumed to be genetically connected, otherwise known as language kinship.

Language kinship refers to the relationship between languages that emerged from a shared ancestor language. It demonstrates how two or more languages are linked through historical development. According to [Swadesh \(2017\)](#), linguistics can shed light on prehistory by: (1) establishing facts concerning the origin (proto) and subsequent divergence of language, (2) discovering diffused features (phonetics and vocabulary) among languages of prehistoric culture contacts, and (3) reconstructing the vocabulary of old stages of languages. In 1955, Swadesh presented lexicostatistics as a technique for uncovering prehistoric language in an effort to clarify these concerns ([Swadesh, 2017](#)). This technique estimates when kinship languages diverged by comparing their 200-item basic vocabulary lists to determine linguistic relatedness ([Koentjaraningrat, 2009](#); [Swadesh, 2017](#)).

Some former studies had identified the issue regarding the linguistic relationship, kinship, and dialect variations among local languages in Indonesia. These can be investigated from the study by [Aziz et al. \(2020\)](#), who studied the language kinship of Tamiang languages in Nanggroe Aceh Darussalam Province; [Rahim and Agus \(2022\)](#), who focused on language kinship in Makassar, Sulawesi Selatan Province; [Sudirman and Dacholfany \(2019\)](#), who studied the kinship of local languages of Komering and Lampung Menggala in Lampung Province; and [Setiawan \(2017\)](#), who investigated the linguistic relationship of Sasak and Samawa Languages in Nusa Tenggara Barat Province. Meanwhile, the present study aims to examine the language kinship and dialectical variations between Banten Javanese and Sundanese spoken in Banten Province. This investigation addresses a notable gap, as limited research has been conducted focusing specifically on these local languages. Specifically, this study seeks to answer the following questions:

1. How is the language kinship between Banten Javanese and Sundanese?
2. When did Banten Javanese and Sundanese separate from their proto-language?
3. What dialect variations occur between Banten Javanese and Sundanese?

This study holds significance as it provides new perspectives on the linguistic kinship and dialectal variations between Banten Javanese and Sundanese, an area that has been underexplored in previous research. By employing glottochronological methods, it enhances the understanding of historical language divergence and the sociolinguistic patterns of Banten Province. Moreover, the findings contribute to the broader field of Indonesian linguistics while supporting ongoing efforts to preserve and document local languages as essential components of Indonesia's cultural heritage.

2. LITERATURE REVIEW

This section reviewed the relevant literature on language kinship, glottochronology, and dialect variation. It covers key concepts related to language kinship and historical language relationships, glottochronology and the timing of language divergence, as well as dialect variation and regional linguistic diversity across different regions.

2.1 Language Kinship and Historical Language Relationships

Language kinship denotes the genetic relationship among languages, reflecting their descent from a common proto-language (Wardhaugh & Fuller, 2021). Such kinship typically emerges as a result of historical contact, migration, or divergence among distinct speech communities. As emphasized by Gapur et al. (2018), the existence of language kinship is empirically substantiated through the presence of inherited linguistic features, basic vocabulary, and lexical elements that trace back to a shared linguistic origin. These situations allow for the identification of basic vocabulary and lexical elements across languages, which are termed cognates. Cognates serve as crucial linguistic evidence in the reconstruction of proto-language and the classification of languages into families and groups (Roy, 2024).

Yule (2020) described cognates as the sets of words of kinship languages that have a similar form and meaning. If language kinship is about related languages, cognates are the evidence within these languages that shows that relationship. So, cognates are the evidence of language kinship. Yule (2020) inserted the examples of cognates for the English words father and brother.

Tabel 1. The example of cognates.

Sanskrit	Latin	Greek	German	English
<i>pitar</i>	<i>pater</i>	<i>pater</i>	<i>vater</i>	father
<i>bhratar</i>	<i>frater</i>	<i>phrater</i>	<i>bruder</i>	brother

The study of language kinship is systematically conducted through the method of comparative reconstruction, a foundational method in historical and comparative linguistics. This method was initially introduced by Franz Bopp in 1816 and later formalized by August Schleicher in 1852 as a scientific tool for reconstructing unattested linguistic forms of proto-languages (Campbell, 2017). Comparative reconstruction operates by identifying systematic correspondence among cognates, lexical, and grammatical elements across related languages to infer the features of their common origin. As Campbell (2017) asserts, this method is central to the classification of languages into origins based on shared inherited features.

Durie and Ross (1995) further emphasized that comparative reconstruction relies on the comparative analysis of morphological paradigms and core lexicon to uncover evidence of genetic relationship and to reconstruct the linguistic history of language families. The method involves a rigorous examination of regular sound correspondence, morphosyntactic structures, and semantic relations that enable linguists to propose the phonological, morphological, and lexical characteristics of a proto-language. Through this systematic comparison, the historical development and the estimated time of divergence between Banten Javanese and Sundanese can be scientifically determined.

2.2 Glottochronology and Language Divergence Timing

A practical instrument in the empirical analysis of language kinship is the Swadesh wordlist, a standardized compilation of 200 core lexical items designed to represent the most stable and culturally neutral elements of human vocabulary (Swadesh, 2017). These items encompass fundamental semantic domains that exhibit a high degree of resistance to borrowing and semantic shift, such as body parts, natural phenomena, kinship terms, basic actions, colors, and numerals (Koentjaraningrat, 2009). The Swadesh wordlist is widely used in historical and comparative linguistics, especially in reconstruction and glottochronology, to quantify relationships among languages.

In this present study, the 200 Swadesh wordlist is employed to systematically evaluate the linguistic kinship and estimate the separation time, or glottochronology, between Banten Javanese and Sundanese. Glottochronology, as formerly developed by Swadesh in 1955, is predicated on the principle that core vocabulary items are changing, replacing, or remaining at a lexicostatistic rate over time (Swadesh, 2017). By calculating the percentage of core vocabularies across languages, the lexicostatistic technique enables the estimation of the time at which these two languages diverged from a proto-language, thereby contributing to a more rigorous reconstruction of linguistic relationships.

2.3 Dialect Variation and Regional Linguistic Diversity across Regions

Dialect is defined as a systematically distinct variety of a language associated with a specific region or social group. As Wolfram and Schilling (2016) assert, a dialect represents an identifiable language variety that reflects the linguistic norms of a particular speech community. The study of dialect, known as dialectology, constitutes a critical discipline for analyzing language variation and regional differentiation in linguistic systems. Since language kinship is predicated on the diachronic process of lexical retention, invention, and replacement within speech communities, dialectology inquiry becomes indispensable for tracing the language change (Trudgill, 2012; Trudgill & Chambers, 2017). As Jendra (2010) notes, dialectology explains spatial and social variation while aiding the reconstruction of historical language relations through cognates and morphophonemic patterns, providing a basis for understanding how varieties evolve, interact, and diverge.

Further discussion of dialectology developed into three major areas: regional or geographical dialect, temporal dialect, and social dialect (Chaer & Agustina, 2010; Coulmas, 2013; Holmes, 2013). The term regional or geographical dialect describes the linguistic variations that members of a speech community employ in the location where they live. Due to the disparities in their geographic locations, Banten and West Java speakers use different dialects of Sundanese language. The term temporal dialect describes linguistic variations that are exclusive to a given historical period. The variations of Old English, Middle English, and Modern English constitute such an example. The last is social dialect, which is produced by the social class or status of the speakers. Building on the work of Errington (2014) and Sukarno (2018), the Javanese as spoken in Yogyakarta exhibits a hierarchical sociolinguistic structure, wherein speakers strategically employ distinct speech levels, *krama inggil* (highly polite), *krama madya* (moderately polite), and *ngoko* (colloquial), to encode and negotiate social relationships. This study builds on the framework to examine kinship, separation time, and dialectal variation between Banten Javanese and Sundanese, offering insights into local diversity and language contact.

3. METHOD

3.1 Research Design

The method used in this research was a case study. As an exploration of a bounded system over time through detailed and in-depth data collection (Creswell, 2013), a case study also seeks

to discover meaning, process, and in-depth understanding of an individual, process, situation, or phenomenon (Emzir, 2010; Yin, 2009). In the present study, the case study serves as an in-depth understanding of language kinship, separation time, and dialect variations that occurred between Banten Javanese and Sundanese in Banten Province, Indonesia.

3.2 Research Participants

The sources of data were selected purposively from four different loci. Lemah Abang and Kragilan Village in Serang Regency were selected for Banten Javanese natives, while Cihideung Village in Pandeglang Regency and Kadu Ketug Village in Lebak Regency were selected for Banten Sundanese natives. These are based on demographic analysis, which revealed that 100% of the population in these villages are Banten Javanese and Sundanese natives. A total of 16 respondents aged between 30 and 60 years participated in this study. Eight respondents were native speakers of Banten Javanese (four males and four females), and eight respondents were native speakers of Banten Sundanese (four males and four females). The selection criteria for the respondents were as follows: (1) they were graduates of elementary and/or junior high school, (2) they resided in the villages where they lived, and (3) they actively used and maintained their languages without interference from other languages.

3.3 Data Collection Techniques

This study employed a triangulated data collection from non-participant observation, semi-structured interviews, and documentation of 200 Swadesh wordlists. Non-participant observation, adapted from Swadesh (2017) and Khalilia et al. (2023), was utilized to capture dialect variation among respondents. This facilitated phonological, lexical, and morphosyntactic features that characterize the spoken varieties under investigation. Semi-structured interviews and documentation, meanwhile, were adapted from Swadesh (2017), Gapur et al. (2018), and Sudirman and Dacholfany (2019), serving as primary techniques for investigating language kinship and separation time between Banten Javanese and Sundanese. The use of the 200 Swadesh wordlist within these techniques enabled a systematic investigation of cognate sets. Therefore, these techniques provided robust data for analyzing the kinship and separation time between these two languages.

3.4 Data Analysis

Data analysis used Miles et al. (2018) model, encompassing data reduction, data display, and drawing/verification of conclusions. This analysis was combined with the lexicostatistic formulas of 200 Swadesh wordlists. The lexicostatistic formulas for separation time are explained as follows (Greenhill & Gray, 2009; Lees, 1953):

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

Note:

w = separation time in thousands of years ago

r = retention or constant percentage in 1000 years (index), which 80.5% is completed to 81%

$\log$ = logarithm

C = percentage of kinship

2 = separation time divided between two languages

The formula for standard error is defined as follows (Gudschinsky, 1956; Lees, 1953):

$$S = \sqrt{\frac{C(I - C)}{n}}$$

Note:

S = standard error in the percentage of identic vocabulary

C = percentage of identical vocabulary

n = number of words compared

4. RESULTS

4.1 Lexical Similarities between Banten Javanese and Sundanese

Lemah Abang Village has a population exceeding one thousand individuals, with a demographic distribution of approximately 40% males and 60% females. The entire population identifies as native speakers of Banten Javanese. A comparable demographic pattern is observed in Kragilan Village, which also comprises 40% male and 60% female residents, all of whom are Banten Javanese natives. In contrast, Cihideung Village, home to approximately two thousand residents, maintains the same gender ratio of 40% males and 60% females, yet is exclusively populated by speakers of Banten Sundanese. Similarly, Kadu Ketug Village, with an estimated population of one thousand, consists of 45% males and 55% females, and is entirely inhabited by Banten Sundanese speakers.

Data displayed through semi-structured interviews and documentation reveal that respondents across the studied villages collectively identified 38 word pairs exhibiting identical vocabularies from the 200 Swadesh wordlist as summarized in Table 2. These shared forms, covering basic concepts such as body parts (*ati*, *jantung*), natural elements (*gunung*, *laut*), and daily objects (*ibu*, *susu*), show strong phonological and morphological similarities. Such findings provide empirical support for lexical convergence and potential language kinship between these two languages. These shared identical vocabularies demonstrate consistent phonological and morphological similarities between Banten Javanese and Sundanese.

Table 2. Identical vocabularies between Banten Javanese and Sundanese.

Banten Javanese	Banten Sundanese	English	Banten Javanese	Banten Sundanese	English
<i>abu /abu/</i>	<i>abu /abu/</i>	ash	<i>gunung /gunun/</i>	<i>gunung /gunun/</i>	mount
<i>angin /anin/</i>	<i>angin /anin/</i>	wind	<i>ibu /ibu/</i>	<i>ibu /ibu/</i>	mother
<i>ambekan /ambakan/</i>	<i>ambekan /ambakan/</i>	breath	<i>jantung /jantun/</i>	<i>jantung /jantun/</i>	heart
<i>ati /ati/</i>	<i>ati /ati/</i>	liver	<i>kabeh /kabeh/</i>	<i>kabeh /kabeh/</i>	all
<i>awan /awan/</i>	<i>awan /awan/</i>	cloud	<i>kali /kali/</i>	<i>kali /kali/</i>	river
<i>balik /balik/</i>	<i>balik /balik/</i>	turn back	<i>kandel /kandəl/</i>	<i>kandel /kandəl/</i>	thick
<i>belok /bəlok/</i>	<i>belok /bəlok/</i>	dirty	<i>kembang /kəmban/</i>	<i>kembang /kəmban/</i>	flower
<i>bintang /bintan/</i>	<i>bintang /bintan/</i>	star	<i>kuku /kuku/</i>	<i>kuku /kuku/</i>	nail
<i>buah /buah/</i>	<i>buah /buah/</i>	fruit	<i>kulit /kulit/</i>	<i>kulit /kulit/</i>	skin
<i>bulan /bulan/</i>	<i>bulan /bulan/</i>	moon	<i>langit /lanit/</i>	<i>langit /lanit/</i>	sky
<i>bulu /bulu/</i>	<i>bulu /bulu/</i>	feather	<i>laut /laut/</i>	<i>laut /laut/</i>	sea
<i>buntut /buntut/</i>	<i>buntut /buntut/</i>	tail	<i>lima /lima/</i>	<i>lima /lima/</i>	five
<i>cekel /cəkəl/</i>	<i>cekel /cəkəl/</i>	hold	<i>manuk /manuk/</i>	<i>manuk /manuk/</i>	bird
<i>cacing /cacin/</i>	<i>cacing /cacin/</i>	worm	<i>panas /panas/</i>	<i>panas /panas/</i>	hot
<i>daging /dagin/</i>	<i>daging /dagin/</i>	meat	<i>pendek /pendek/</i>	<i>pendek /pendek/</i>	short
<i>dorong /doron/</i>	<i>dorong /doron/</i>	push	<i>tali /tali/</i>	<i>tali /tali/</i>	rope
<i>endog /əndog/</i>	<i>endog /əndog/</i>	egg	<i>tarik /tarik/</i>	<i>tarik /tarik/</i>	pull
<i>garing /garin/</i>	<i>garing /garin/</i>	dry	<i>susu /susu/</i>	<i>susu /susu/</i>	milk
<i>getih /gətih/</i>	<i>getih /gətih/</i>	blood	<i>uyah /uyah/</i>	<i>uyah /uyah/</i>	salt

The lexical comparison between Banten Javanese and Sundanese, based on the 200 Swadesh wordlist, revealed 38 identical vocabularies or word pairs, as presented in Table 2. These identical vocabularies represent 19% of the total lexical items examined, indicating a notable degree of vocabulary overlap between the two languages.

Following this comparison, the analysis proceeded to examine phonemic similarities, as presented in Table 3.

Table 3. Phonemic similarity between Banten Javanese and Sundanese

Banten Javanese	Banten Sundanese	English	Banten Javanese	Banten Sundanese	English
<i>aran</i> / <i>aran</i> /	<i>ngaran</i> / <i>ŋaran</i> /	name	<i>siji</i> / <i>siji</i> /	<i>hiji</i> / <i>hiji</i> /	one
<i>bengi</i> / <i>bəŋi</i> /	<i>wengi</i> / <i>wəŋi</i> /	night	<i>taun</i> / <i>taun</i> /	<i>tahun</i> / <i>tahun</i> /	year
<i>kuning</i> / <i>kuniŋ</i> /	<i>koneng</i> / <i>koneŋ</i> /	yellow	<i>telu</i> / <i>təlu</i> /	<i>tilu</i> / <i>tilu</i> /	three
<i>mutah</i> / <i>mutah</i> /	<i>utah</i> / <i>utah</i> /	vomit	<i>tipis</i> / <i>tipis</i> /	<i>ipis</i> / <i>ipis</i> /	thin
<i>ngitung</i> / <i>ŋitun</i> /	<i>itung</i> / <i>itun</i> /	count	<i>watu</i> / <i>watu</i> /	<i>batu</i> / <i>batu</i> /	stone
<i>peres</i> / <i>pəɾəs</i> /	<i>meres</i> / <i>məɾəs</i> /	press	<i>weteng</i> / <i>wətəŋ</i> /	<i>beteng</i> / <i>bətəŋ</i> /	stomach
<i>sapa</i> / <i>sapa</i> /	<i>saha</i> / <i>saha</i> /	who			

Table 3 presents 13 lexical items, or 6.5% of the 200-word Swadesh list, that exhibit phonemic similarity between Banten Javanese and Sundanese. These similarities are characterized by systematic phoneme correspondences, such as /s/ in *siji* versus /h/ in *hiji*, and the presence or absence of nasal consonants as seen in *aran* versus *ngaran*. Such patterns suggest a degree of historical relatedness and potential shared origins. The comparative reconstruction of phonemic similarity can be displayed in Table 4.

Table 4. Comparative reconstruction between Banten Javanese and Sundanese.

Banten Javanese	Phonetic symbol	Banten Sundanese	Phonetic symbol	English
<i>aran</i>	(-) nasal	<i>ngaran</i>	(+) nasal	name
<i>bengi</i>	stop	<i>wengi</i>	semi-vowel	night
<i>kuning</i>	closed-vowel	<i>koneng</i>	opened-vowel	yellow
<i>mutah</i>	(+) nasal	<i>utah</i>	(-) nasal	vomit
<i>ngitung</i>	(+) nasal	<i>itung</i>	(-) nasal	count
<i>peres</i>	stop	<i>meres</i>	nasal	press
<i>sapa</i>	stop	<i>saha</i>	glottal	who
<i>siji</i>	fricative	<i>hiji</i>	glottal	one
<i>taun</i>	(-) glottal	<i>tahun</i>	(+) glottal	year
<i>telu</i>	opened-vowel	<i>tilu</i>	closed-vowel	three
<i>tipis</i>	(+) stop	<i>ipis</i>	(-) stop	thin
<i>watu</i>	semi-vowel	<i>batu</i>	stop	stone
<i>weteng</i>	semi-vowel	<i>beteng</i>	stop	stomach

Table 4 illustrates systematic phonemic correspondences between Banten Javanese and Sundanese, highlighting both similarities and shifts that indicate historical relatedness. For instance, several cognates demonstrate variation in nasality (e.g., *aran* vs. *ngaran*, *mutah* vs. *utah*), vowel quality (e.g., *kuning* vs. *koneng*, *telu* vs. *tilu*), and consonantal changes involving stops, fricatives, semi-vowels, or glottal (e.g., *bengi* vs. *wengi*, *siji* vs. *hiji*, *sapa* vs. *saha*). These consistent phonological shifts suggest shared origins and provide evidence of a close linguistic kinship between the two languages, while also reflecting dialectal variation shaped by historical development.

4.2 Estimating the Time of Language Separation

Based on the displayed data comprising 38 pairs of identical vocabularies and 13 pairs of phonemic similarities, a total of 51 word pairs, representing 25.5% (rounded to 26%), exhibit linguistic similarity between Banten Javanese and Sundanese. This suggests a notable lexical and

phonological correspondence between the two languages. To estimate the separation time between these languages, it is calculated using the following lexicostatistic formula:

$$\begin{aligned}
 r &= 81\% = 0.81 \text{ (log } 0.81 = -0.091) \\
 C &= 26\% = 0.26 \text{ (log } 0.26 = -0.585) \\
 w &= \frac{\log C}{2 \log r} \\
 &= \frac{\log 0.26}{2 \log 0.81} \\
 &= \frac{-0.585}{2 (-0.091)} \\
 &= \frac{-0.585}{-0.183} \\
 &= 3196
 \end{aligned}$$

The results indicate that the estimated separation time between Banten Javanese and Sundanese is approximately 3196. To ensure the accuracy and minimize potential errors in calculation, the data were further distributed to a standard error formula as presented below:

$$\begin{aligned}
 S &= \sqrt{\frac{0.26 (1 - 0.26)}{200}} \\
 &= \sqrt{\frac{0.26 (0.74)}{200}} \\
 &= \sqrt{\frac{0.1924}{200}} \\
 &= 0.098
 \end{aligned}$$

The result of new C is $0.26 + 0.098 = 0.358$ (rounded to 0.36). Thus, the recalculation of the new separation time between Banten Javanese and Sundanese can be explained as follows.

$$\begin{aligned}
 w &= \frac{\log C}{2 \log r} \\
 &= \frac{\log 0.36}{2 \log 0.81} \\
 &= \frac{-0.443}{2 (-0.091)} \\
 &= \frac{-0.443}{-0.183} \\
 &= 2336
 \end{aligned}$$

The abovementioned result verified that the new separation time between Banten Javanese and Sundanese is **2336**. To calculate the error period, the previously estimated separation time is reduced by the latest time ($3196 - 2336 = 860$). Thus, the error period was 860. Given the current year is 2025, the separation time between Banten Javanese and Sundanese from the proto language is estimated to have occurred throughout an extended period, about 1171 to 311 BC. This period highlights the historical and complex development of these two local languages.

4.3 Dialect Variations Based on Region

The observation results display that dialect variations between Banten Javanese and Sundanese are primarily shaped by regional or geographical dialect. This observation is supported by glottochronological data, which suggest that the separation of Banten Javanese and Sundanese

from the proto-language was significantly influenced by the regional division of their respective speech communities. Lemah Abang and Kragilan Villages, both located within Serang Regency, are known for their use of Banten Javanese. In contrast, Cihideung Village in Pandeglang Regency and Kadu Ketug Village in Lebak Regency predominantly use Banten Sundanese. These linguistic patterns clearly illustrate that regional or geographical variations have developed in alignment with the territorial origins of the respective communities. More broadly, Banten Javanese dialects are primarily spoken in Serang Regency, Serang City, and Cilegon City, while Banten Sundanese dialects are widely spoken across Pandeglang and Lebak Regencies, as well as in some regions of Tangerang Regency.

In contrast to the Sundanese dialect spoken in West Java, the Banten Sundanese dialect has retained several conservative features that distinguish it from its neighboring varieties. This distinctiveness reflects a conscious effort by local communities to preserve older linguistic forms as a marker of regional identity. Over time, such features have evolved alongside social and cultural changes, with some communities maintaining traditional patterns while others adapt to new sociocultural settings. These patterns show that local language functions as both communication and a vital expression of cultural identity and community cohesion.

5. DISCUSSION

The findings address the first question, identifying 51 word pairs that account for 25.5%, or approximately 26%, of the lexical items analyzed from the Swadesh wordlist. These pairs exhibit systematic phonemic correspondences and lexical equivalence, suggesting a substantial degree of linguistic relationship between Banten Javanese and Sundanese. The presence of these results supports that both language varieties share a common linguistic substrate, particularly within the core of the vocabulary domain. These results confirm the previous studies by [Perwitasari \(2019\)](#) and [Mutmainah et al. \(2024\)](#), which similarly reported recurrent phonological and lexical similarities between Javanese and Sundanese. This evidence reinforces the notion of linguistic relationship, aligning with broader theories of Malayo-Polynesian language as a proto-language for Banten Javanese and Sundanese ([Bellwood, 2014](#); [Sneddon, 2003](#)). Consequently, this provides further data for reconstructing the local language development over time.

The findings from the second research question reveal that the separation time, or glottochronology, between Banten Javanese and Sundanese is estimated to have occurred in 1171 to 311 BC. These results are along with some historical manuscripts and evidence, as [Collins \(2005\)](#) and [Vlekke \(2008\)](#) confirmed that the migrations of Deutero Malay, or Young Malay, into the archipelago were estimated from 1500 to 500 BC. These new Malays, then, developed into several tribes, such as Sundanese, Javanese, Balinese, and others. Through this historical evidence, it can be determined that the Banten Javanese and Sundanese were originally from a similar proto-language. Furthermore, [Sneddon \(2003\)](#) and [Bellwood \(2014\)](#) stated that both Banten Javanese and Sundanese are sister languages and come from Malayo-Polynesian as a proto-language, and they diverged thousands of years ago. As proved in this study, around 1171 to 311 BC, the Malayo-Polynesian language developed into Banten Javanese and Sundanese.

The findings for the last question prove that regional or geographical dialects predominantly shape the dialectical variations of these languages. Banten Javanese is widely used by people concentrated in Serang and Cilegon, while Banten Sundanese is mostly spoken by speech communities concentrated in Pandeglang and Lebak. Banten Javanese was mainly brought through migration from Central and East Java, especially during Mataram Sultanate, to the northern coastal area of Banten ([Michrob & Chudari, 2003](#)). Over time, Banten Javanese developed with simplified grammar and speech levels, while Banten Sundanese historically served as the official language of the Sunda Kingdom across Western Java and Southern Banten ([Michrob & Chudari, 2003](#)).

In contrast to the West Java dialect, Banten Sundanese dialect evolved its own dialectal 'conservative' features throughout time. Sometimes, Banten people and also other speech communities often adjust their speech to reflect their local identities by developing unique

linguistic features as a form of regional or geographical distinction from others. Over time, regional linguistic landscapes naturally evolve; some communities preserve older linguistic forms, while others adjust and adapt to new sociocultural settings. The findings of this study underline the role of local language as an essential tool for expressing regional identity, particularly in informal and everyday contexts. As living local languages, Banten Javanese, Sundanese, and other local languages not only embody the nation's rich cultural heritage and diversity (Altinyelken et al., 2014; Asari et al., 2018; Chen et al., 2021; Forman, 2014; Indarti & Manara, 2025; Mudana, 2023), but also serve as vital instruments for maintaining linguistic diversity and strengthening cultural pride (Akhtar & Menjivar, 2012; Festman, 2021; Liberman et al., 2017; Putri et al., 2019; Utami et al., 2017). The sustained use of these languages in media, education, social life, and governance affirms their role in preserving cultural diversity and community cohesion.

These three findings form an integrated picture of linguistic kinship, separation time, and dialect variation between Banten Javanese and Sundanese. The identical vocabularies and phonemic similarities confirm a linguistic relationship (Alimahmudi & Siagian, 2022), while the glottochronological estimates contextualize the historical depth of their divergence. The geographical patterns of dialect distribution further demonstrate how historical migration, political contexts, and cultural identity have contributed to the linguistic landscape of the region.

Finally, the broader implications of this study highlight the dynamic nature of language development in a multilingual and multicultural region such as Banten Province. It emphasizes the role of historical migration and regional identity in shaping linguistic diversity and contributes to a deeper understanding of how local languages adapt, persist, and evolve over time. In doing so, this research not only advances the study of Malayo-Polynesian languages but also supports ongoing efforts to document, preserve, and revitalize local languages as vital components of Indonesia's cultural heritage (Endardi & Rahayu, 2023).

6. CONCLUSION

This study concludes that there are 51 pairs of identical words between Banten Javanese and Sundanese that exhibit both morphemic and phonemic similarities, consisting of 38 identical vocabularies and 13 phonemic similarities. Based on glottochronological estimates, these similarities suggest that the separation time of these two local languages from a proto-language occurred between 1171 to 311 BC. The findings also reveal that regional or geographical factors primarily influence dialect variations between Banten Javanese and Sundanese. Specifically, Banten Javanese dialects are predominantly spoken in the northern area of Banten, including Serang Regency and Cilegon City, while Banten Sundanese dialects are more commonly used in the southern region of Banten, including Pandeglang and Lebak Regencies.

These findings reveal linguistic relationships between these local languages and show how regional factors have shaped their development. This study also offers meaningful contributions to the field of linguistics by highlighting the role of glottochronology in understanding the linguistic divergence and the emergence of dialect variations over time. However, the limited sample size and duration of this present study may constrain the transferability of its findings. Therefore, further longitudinal research focusing on language kinship and dialect diversity in other Indonesian local languages is highly recommended. This importance is not only viewed as a way of investigating the origin of local languages but also as a means of maintaining and revitalizing the Indonesian local languages.

REFERENCES

- Akhtar, N., & Menjivar, J. A. (2012). Cognitive and linguistic correlates of early exposure to more than one language. *Advances in Child Development and Behavior*, 42, 41-78. <https://doi.org/10.1016/B978-0-12-394388-0.00002-2>

- Alimahmudi, H., & Siagian, I. (2022). Phonetic studies: Similarities and differences in the meaning of Javanese and Sundanese words in Indonesian. *Literatus*, 4(1), 135-138. <https://doi.org/10.37010/lit.v4i1.695>
- Altinyelken, H. K., Moorcroft, S., & van der Draai, H. (2014). The dilemmas and complexities of implementing language-in-education policies: Perspectives from urban and rural contexts in Uganda. *International Journal of Educational Development*, 36, 90-99. <https://doi.org/10.1016/j.ijedudev.2013.11.001>
- Asari, S., Fauziyah, N., & Uchtiawati, S. (2018). Improving teacher pedagogic competences in remote areas through lesson study activity. *International Journal of Education and Literacy Studies*, 6(2), 53-62. <https://doi.org/10.7575/aiac.ijels.v.6n.2p.53>
- Aziz, Z. A., Safhida, M., & Mahmud, M. (2020). Kinship terminology in the Tamiang language: A Malay variety spoken in Eastern Aceh, Indonesia. *Studies in English language and Education*, 7(2), 642-656. <https://doi.org/10.24815/siele.v7i2.16802>
- Bellwood, P. (Ed.). (2014). *The global prehistory of human migration*. Wiley Blackwell.
- Bucholtz, M., & Hall, K. (2005). Language and identity. In A. Duranti (Ed.), *A companion to linguistic anthropology* (pp. 369-394). Blackwell Publishing Ltd. <https://doi.org/10.1002/9780470996522.ch16>
- Campbell, L. (2017). The history of linguistics: Approaches to linguistics. In M. Aronoff & J. R. Miller (Eds.), *The handbook of linguistics* (2nd ed., pp. 95-117). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119072256.ch6>
- Chaer, A., & Agustina, L. (2010). *Sosiolinguistik: Perkenalan awal* [Sociolinguistics: An introduction]. Rineka Cipta.
- Chen, X., Dewaele, J. M., & Zhang, T. (2021). Sustainable development of EFL/ESL learners' willingness to communicate: The effects of teachers and teaching styles. *Sustainability*, 14(1), 396. <https://doi.org/10.3390/su14010396>
- Collins, J. T. (2005). *Bahasa Melayu bahasa dunia: Sejarah singkat* [Malay language as a world language: A brief history]. Yayasan Obor Indonesia.
- Coulmas, F. (2013). *Sociolinguistics: The study of speakers' choices* (2nd ed.). Cambridge University Press.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five traditions*. Sage Publications Inc.
- Durie, M., & Ross, M. (Eds.). (1995). *The comparative method reviewed: Regularity and irregularity in language change*. Oxford University Press. <https://doi.org/10.1093/oso/9780195066074.001.0001>
- Emzir. (2010). *Metodologi penelitian kualitatif: Data analisis* [Qualitative research methodology: Data analysis]. PT Rajagrafindo Persada.
- Endardi, J., & Rahayu, M. (2023). The diversity of multilingual communities: A case study in eight regions in Indonesia. In R. N. Indah, M. Huda, Irham, M. Afifuddin, Masrokhin, & D. E. N. Rakhmawati (Eds.), *Proceedings of the 4th Annual International Conference on Language, Literature and Media (AICOLLIM 2022)* (pp. 277-287). Atlantis Press. https://doi.org/10.2991/978-2-38476-002-2_27
- Errington, J. (2014). *Indonesian among Indonesia's languages*. Cornell University Press.
- Festman, J. (2021). Learning and processing multiple languages: The more the easier? *Language Learning*, 71(S1), 121-162. <https://doi.org/10.1111/lang.12437>
- Forman, R. (2014). How local teachers respond to the culture and language of a global English as a foreign language textbook. *Language, Culture and Curriculum*, 27(1), 72-88. <https://doi.org/10.1080/07908318.2013.868473>
- Gapur, A., Siregar, D. S. P., & Pujiono, M. (2018). Language kinship between Mandarin, Hokkien Chinese and Japanese (lexicostatistics review). *Aksara*, 30(2), 301-318. <https://doi.org/10.29255/aksara.v30i2.267.301-318>
- Greenhill, S. J., & Gray, R. D. (2009). Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods. In K. A. Adelaar & A. Pawley (Eds.), *Austronesian historical linguistics and culture history: A festschrift for Robert Blust* (pp. 375-397). Pacific Linguistics.

- Gudschinsky, S. C. (1956). The ABC's of lexicostatistics (glottochronology). *Word*, 12(2), 175-210. <https://doi.org/10.1080/00437956.1956.11659599>
- Holmes, J. (2013). *An introduction to sociolinguistics* (4th ed.). Pearson Education Ltd. <https://doi.org/10.4324/9781315833057>
- Indarti, D., & Manara, C. (2025). Attitudes toward language and language choice in regional intermarriage families: A case study in Tangerang, Banten. *Studies in English Language and Education*, 12(1), 426-441. <https://doi.org/10.24815/siele.v12i1.36436>
- Jendra, M. I. I. (2010). *Sociolinguistics: The study of society's languages*. Graha Ilmu.
- Khalilia, H., Bella, G., Freihat, A. A., Darma, S., & Giunchiglia, F. (2023). Lexical diversity in kinship across languages and dialects. *Frontiers in Psychology*, 14, 1229697. <https://doi.org/10.3389/fpsyg.2023.1229697>
- Koentjaraningrat. (2009). *Pengantar ilmu antropologi* [An introduction to anthropology]. Rineka Cipta.
- Lees, R. B. (1953). The basis of glottochronology. *Language*, 29(2), 113-127. <https://doi.org/10.2307/410164>
- Liberman, Z., Woodward, A. L., Keysar, B., & Kinzler, K. D. (2017). Exposure to multiple languages enhances communication skills in infancy. *Developmental Science*, 20(1), 12420. <https://doi.org/10.1111/desc.12420>
- Maemunah, E. (2017). Makna kosakata “jatuh” dalam bahasa Sunda dan bahasa Jawa [The meaning of “fall” in Sundanese and Javanese languages]. *Aksara*, 29(2), 239-252. <https://doi.org/10.29255/aksara.v29i2.38.239-252>
- Meyerhoff, M. (2018). *Introducing sociolinguistics* (3rd ed.). Routledge. <https://doi.org/10.4324/9780429507922>
- Michrob, H., & Chudari, M. (2003). *Proses Islamisasi di Banten* [The Islamization process in Banten]. Dinas Pendidikan Provinsi Banten.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). *Qualitative data analysis: A methods sourcebook* (4th ed.). Sage Publications.
- Mudana, W. (2023). The effect of ethnoscience-based course review horay learning towards cultural concept understanding and science process skills of the elementary school students. *Nurture*, 17(2), 137-148. <https://doi.org/10.55951/nurture.v17i2.253>
- Mutmainah, S., Haryanto, F. A. R., Astuti, D., Kismanto, S. A., Qonitulhaq, E. F. R., & Muharudin, E. (2024). Variations of Sundanese in agrarian societies in Central Java. *Athena: Journal of Social, Culture and Society*, 2(4), 408-417. <https://doi.org/10.58905/athena.v2i4.302>
- Perwitasari, A. (2019). *The acquisition of English vowels by Javanese and Sundanese native speakers*. Leiden University Press.
- Putri, I., Daulay, S., & Lubis, M. (2019). Phonological interference and morphology of BMT in student narrative essays, Indonesia. *Budapest International Research and Critics in Linguistics and Education (BirLE) Journal*, 2(1), 61-71. <https://doi.org/10.33258/birle.v2i1.186>
- Rahim, A. R., & Agus, M. (2022). Kekerabatan Bahasa Makassar dan Bahasa Selayar: Analisis leksikostatistik dan glotokronologi [The relationship between the Makassarese and the Selayar Language: A lexicostatistical and glottochronological analysis]. *Gema Wiralodra*, 13(1), 215-232. <https://doi.org/10.31943/gw.v13i1.215>
- Roy, C. (2024). Cognates in linguistic analysis: Examining the interconnections of language similarities. *Mathematica Eterna*, 14(2), 225.
- Setiawan, I. (2017). Relation of Sasak and Samawa Language: Diachronic study in the language kinship of an ethnic group in Indonesia. *International Journal of Linguistics, Literature and Culture*, 3(5), 83-90.
- Sneddon, J. (2003). *The Indonesian Language: Its history and role in modern society*. University of New South Wales Press Ltd.
- Sudaryat, Y. (2015). The mindset of Sundanese people in their language expressions. *Indonesian Journal of Applied Linguistics*, 5(1), 126-133. <https://doi.org/10.17509/ijal.v5i1.839>

- Sudirman, A. M., & Dacholfany, I. (2019). Language kinship between Komerling variation and Lampung Menggala. *Language*, 3(1), 1-13. <https://doi.org/10.22146/sasdayajournal.43883>
- Sukarno. (2018). Politeness strategies, linguistic markers and social contexts in delivering requests in Javanese. *Indonesian Journal of Applied Linguistics*, 7(3), 659-667. <https://doi.org/10.17509/ijal.v7i3.9816>
- Swadesh, M. (2017). *The origin and diversification of language*. Routledge. <https://doi.org/10.4324/9781315133621>
- Trudgill, P. (2012). *Dialects* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203314609>
- Trudgill, P., & Chambers, J. K. (2017). *Dialects of English: Studies in grammatical variation*. Routledge. <https://doi.org/10.4324/9781315505459>
- Utami, D. H., Wello, M. B., & Atmowardoyo, H. (2017). The phonological interference of students' first language in pronouncing English sounds (A case study on Buginese and Makassarese students). *ELT Worldwide*, 4(2), 205-212. <https://doi.org/10.26858/eltww.v4i2.4414>
- Vlekke, B. H. M. (2008). *Nusantara: Sejarah Indonesia* [Archipelago: The history of Indonesia]. Kepustakaan Populer Gramedia.
- Wardhaugh, R., & Fuller, J. M. (2021). *An introduction to sociolinguistics* (8th ed.). Blackwell Publishing Ltd.
- Wolfram, W., & Schilling, N. (2016). *American English: dialects and variation*. John Wiley & Sons.
- Yin, R. K. (2009). *Case study research: Design and methods*. Sage Publications Inc.
- Yule, G. (2020). *The study of language* (7th ed.). Cambridge University Press.