

The Genetic Relationship Between Balinese And Balinese Malay: A Lexicostatistics And Phonemic Correspondence Study

Ida Ayu Putu Aridawati¹

¹Badan Riset dan Inovasi Nasional, Indonesia

I Gede Bagus Wisnu Bayu Temaja²

² IKIP Saraswati, Indonesia

Ida Bagus Rai Putra³

³ Universitas Udayana, Indonesia

I Wayan Tama⁴

⁴ Badan Riset dan Inovasi Nasional, Indonesia

I Made Sudiana⁵

⁵ Badan Riset dan Inovasi Nasional, Indonesia

I Wayan Sudiarta⁶

⁶ Badan Riset dan Inovasi Nasional, Indonesia

I Gusti Ayu Armini⁷

⁷ Badan Riset dan Inovasi Nasional, Indonesia

Corresponding Author: dayuarida1963@gmail.com

Abstract

This research aimed to explore the genetic relationship between Balinese and Balinese Malay quantitatively and qualitatively. The research was conducted in three steps: 1) data collection, 2) data analysis, and 3) data display. Data collection was conducted by implementing the interview method through a guided interview, recording, and note-taking techniques. The data were 200 Swadesh word lists collected from speakers of both languages as the data. Data analysis was conducted by implementing lexicostatistics and phonemic correspondence techniques. The data were displayed formally and informally, as well as in both combinations. The research results showed that both languages have a 36% genetic relationship percentage. The time between the two languages was presumed to have diverged around 2,354 – 2,157 years



ago. It is estimated that both languages started to diverge from their earlier proto-language around 2,569 years ago. In addition, six phonemic correspondences of both languages consist of /a – a /, / i – ə /, / a – ə /, / o – u /, / u – o /, and / h – r /. Concerning the levels of language classification, the genetic relationship of both languages belongs to a common language family since the results are within the percentage range (36-81%) and diverged year range (500-2,500) years. Based on the genetic language typology, it can be concluded that both languages belong to Austronesian languages. Specifically, they belong to the subfamily Malayo-Sumbawan languages.

Keywords: Genetic Relationship; Lexicostatistics; Phonemic Correspondence; Balinese; Balinese Malay.

INTRODUCTION

On the island of Bali, two mother tongues are members of the Austronesian languages, namely Balinese and Balinese Malay. Balinese is a regional language in Indonesia spoken on the island of Bali (Temaja, 2018d, 2018b), parts of Lombok, and the east coast of East Java. Balinese is one of Indonesia's languages with many speakers (Temaja, 2018a, 2021). On the island of Bali, Balinese is also the language of instruction in both traditional and government events, in addition to Bahasa Indonesia (Temaja, 2018c). The Balinese in each region on the island of Bali is different due to differences in ethnicity, region and other influences. This difference led to the development of two different dialects, Bali Aga and Bali Dataran. The Bali Aga tribe speaks the Bali Aga dialect, primarily spread across the mountainous areas of Bali Island, such as Trunyan, Tenganan, Tigawasa, Sidatapa, and others. The Bali Dataran dialect is the dialect with the most speakers spoken by the Balinese. This tribe lives in almost all areas of the island of Bali, predominantly urban areas.

Balinese Malay, or what the speakers call "Malay Loloan" or "Base Loloan," is a variant of the Malay language on the island of Bali. According to Bagus et al. (1985), Balinese Malay speakers are spread over two sub-districts, namely Negara and Melaya sub-districts in Jembrana Regency. The speakers are found in villages such as West Loloan and East Loloan, located in Negara, as well as Tegal Badeng Islam, Pengambangan, Cupel, Banyubiru, Melaya Pantai, Candikesuma, and Tuwed. Most speakers of Balinese Malay are Muslims who live in harmony with the majority of Balinese who are Hindus (Temaja, 2017). Balinese Malay is a variant or



pocket language of Malay spoken outside the origin area. Muhadjir (2000) states that in Indonesia, there are at least 30 (thirty) variants of the Malay language, which he terms as local Malay.

The interaction that took place hundreds of years ago between Balinese and Balinese Malay speakers allows them to understand each other's speech even though they still use their mother tongue when interacting. This phenomenon is interesting to research to find out how far speakers of both languages understand each other's speech. Based on this phenomenon, researchers are interested in investigating the genetic relationship between the two languages, which is a sub-field of comparative historical linguistics, one of which is by applying lexicostatistical techniques (Ersheidat & Tahir, 2020; Hendrokusumo & Temaja, 2019). In order to support lexicostatistical techniques that examine the genetic relationship quantitatively, the genetic relationship is usually also sought qualitatively, one of which is by exploring the phonemic correspondences of related words from two or more language pairs. If we also look at the language typology, this research is included in the genetic typology, namely the grouping of languages based on the lineage from the same parent language (Tsaralunga, 2019). Based on this explanation, this research aimed to determine how the Balinese relate to Balinese Malay quantitatively and qualitatively, which supports their genetic typology.

There have been quite several studies examining these two languages, including Adelaar (2005, p. 357) in his article on the Malay-Sumbawa subgroup of languages. In his research, it was explained that Balinese and Malay had a closeness of 35%. The Malay language in question is the common Malay language. Based on these results, this research is focused on proving whether the percentage of genetic relationship also applies to variants of the Malay language, such as Balinese Malay, in this case. Other research comes from Putra et al. (2015) entitled "Adaptation of Balinese Vocabulary in Balinese Loloan Malay". The research found that Balinese vocabulary was absorbed into Balinese Malay through four processes, namely: adoption and adaptation; absorption of basic vocabulary in terms of the seven elements of culture; word categories that occur in word classes; and vocabulary absorption that led to phonological dynamics. Reflecting on that, one area that has not been implemented is related to the relationship between the two languages, namely in the genetic relationship of language.

This research used Blust's theory (2013, p. 32) of genetic typological grouping regarding the Austronesian language family in which Balinese and Balinese Malay (Malay) belong to the



Austronesian language family. Furthermore, this study also intends to prove Adelaar's theory about the subgroup of Sumbawa-Malay languages.

METHOD

The research was divided into three stages: data collection, analysis, and display. Data was collected from each informant's native Balinese and Balinese Malay speakers. The Balinese language used as the data source was the standard Balinese (Bali Dataran Dialect). The method used in providing data was the interview method through a guided interview, recording, and note-taking techniques from Sudaryanto (2015). In order to support the data, additional references for both languages, such as the Balinese dictionary from Denes (1997) and the Balinese-Indonesian dictionary from Bagus et al. (1985), were used. The research instrument used Swadesh's 200 basic vocabulary, recording equipment, and notebooks. Data analysis was carried out by applying lexicostatistics and phonemic correspondence techniques from Keraf (1996). The data were presented formally and informally, as well as a combination of both (Sudaryanto, 2015). In formal display, the researcher formulates the data using linguistic signs, such as square brackets [...], which indicate phonetic spelling; long dashes (-) indicate a constituent; double cross # as a lingual unit boundary; and italic brackets // flanking phonological elements (Kridalaksana, 2008, pp. xvi–xvii). The informal display was implemented by formulating the analysis results with detailed and precise descriptions of words and a combination of these formal and informal methods.

The implementation of the lexicostatistics technique follows the procedure of Keraf (1996, pp. 126–136) with the following stages:

- 1) Collecting 200 basic Swadesh vocabulary from informants in both languages.
- 2) Counting related words or cognates.

This calculation follows the criteria including:

- a. Bringing out the gloss that does not count

It is essential to analyze the data for these two languages, considering that Balinese Malay absorbs much vocabulary from Balinese; hence these words were not included as related words.

- b. Isolating of bound morphemes

The data that contained bound morphemes must be separated between the base form and the attached bound morpheme. Therefore, only the essential words were compared.

c. Determining related word

This arrangement was applied by selecting identical pairs with phonemic correspondence, phonetic similarities, and one different phoneme. After all the related words are obtained, proceed with calculating the percentage of the relationship with the following formula:

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

Successively, the explanation of the formula is as follows: C = percentage of genetic relationship, J = related words, and n = the number of vocabularies being compared.

3) Calculating the diverging time of the two languages

After the percentage of the relationship between the two languages can be found, it is continued by calculating the diverging time using the formula:

$$W = \frac{\log.C}{2 \times \log.r}$$

The following explains the formula: W = Split time thousands of years ago, log = Logarithm of, r = Retention (constant percentage in 1000 years, commonly used 80.5% or 0.805), and C = percentage of relationship.

4) Calculating the error range

The results obtained from calculating the diverging time are in the form of a specific year. However, separating the two languages cannot be based solely on a particular year. Hence it is necessary to know the separation period. Therefore, calculating the error period is required in the following stages:

a. Calculating the standard error range

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

Successively, the explanation of the formula is as follows: S = Standard error range in percentage of genetic relationship, C = percentage of genetic relationship, and n = Number of vocabularies being compared. Then, the results of this calculation are used to obtain a new genetic relationship percentage with the formula:

$$C1 = C + S$$



The following description of the formula: C1 = Percentage of new genetic relationship, C = Percentage of genetic relationship, and S = Standard error range in percentage of related words.

- b. Calculating the new split time with the formula:

$$W1 = \frac{\log.C1}{2 \times \log.r}$$

The following is a description of the formula: W1 = Time of new split, log = Logarithm of, r = Retention (constant percentage in 1000 years, commonly used 80.5% or 0.805), and C1 = percentage of new genetic relationship.

- c. Calculating the split time of the two languages with their proto-language using the following formula:

$$\text{Split time from proto} = W1 + (W - W1)$$

Description of the formula: W1 = New split time and W = split time thousands of years ago.

The use of lexicostatistical techniques is not only used to calculate the percentage of genetic relationship and split time of a language from other languages. However, it is also used to classify these languages. Grouping is carried out by connecting the percentage of genetic relationship and split time of the two languages according to the language classification in the following table.

Table 1. language classification levels

Language Level	Split Time (in Century)	Percentage of Related Words
Language	0 - 5	100 - 81
Family	5 - 25	81 - 36
Stock	25 - 50	36 - 12
Microphylum	50 - 75	12 - 4
Mesophylum	75 - 100	4 - 1
Macrophylum	100 above	1 - less than 1%



RESULTS AND DISCUSSION

The results and discussion of the genetic relationship between the two languages were divided based on the arrangement of lexicostatistical techniques and phonemic correspondence.

Lexicostatistics

The results of calculations with this technique are explained based on the following stages.

Sorting out related words

Following Keraf's steps, the initial stage was to bring out the gloss that is not considered as borrowed words. It is mainly found in Balinese Malay, which absorbs much Balinese vocabulary. Some of these loan words include: [kambaŋ] 'float', [ambu] 'cloud', [gəde] 'big', [be] 'meat', [duwə] 'two', [gesges] 'scratch', [antəm] 'smash', [sələm] 'black', [pəsəl] 'tie', [climb] 'dirty', [bəlig] 'slick', [ləsər] 'straight', [naʔ] 'people', [kədət] 'pull', [kopin] 'ear', and [kədəʔ] 'laugh'.

Genetic Relationship Percentage

Of the 200 Swadesh basic vocabularies between Balinese and Balinese Malay, 72 related words were found with details: 20 identical pairs of words, 30 phonemic-corresponding word pairs, 15 phonetically similar word pairs, and 7 word pairs differ by one phoneme. Calculations using lexicostatistical techniques to obtain the percentage of genetic relationship were described as follows.

$$C = \frac{72}{200} \times 100\% = 36\%$$

From the calculation results, the genetic relationship between the two languages is 36%. From these figures, the two languages show a genetic relationship between languages. The genetic relationship status of the two languages shows that both belong to the language family according to the percentage term (36-81%) based on Table 1.

Split Time Of The Two Languages

Next, the calculation of the split time between the two languages is carried out:

$$W = \frac{\log. 0.36}{2 \times \log. 0.805} = 2.354 \times 1,000 \text{ years} = 2,354 \text{ years}$$



The calculation results found that the time when the two languages separated was 2,354 years ago. However, it is impossible to time a language's split year in a given year. For this reason, further calculations are carried out to determine the language split period and avoid the calculation error period. The calculation is continued in the next step.

Split Error Period And Percentage Of New Genetic Relationship

To find out the split time error period is calculated in the following way:

$$S = \frac{\sqrt{0.36(1 - 0.36)}}{200} = 0.033$$

Therefore, the resulting standard error is 0.033. This value is then used to determine the percentage of new genetic relationship or C1 with the following calculation:

$$C1 = 0.36 + 0.033 = 0.393$$

It can be seen that the percentage of new genetic relationship between the two languages is 0.363.

New time split

After the calculation above, proceed with finding out the new split time in the following way:

$$W1 = \frac{\log. 0.393}{2 \times \log. 0.805} = 2.157 \times 1,000 = 2,157 \text{ years ago}$$

Therefore, the result is the new split of the two languages, which was 2,157 years ago.

Split time of the two languages from their proto-language

Then, considering that the two languages are genetic descendants of the same proto family, it is necessary to know when they separated from the proto-language. In order to obtain this, the following calculations are used:

Split time from proto

$$= 2,360 + (2,354 - 2,157)$$

$$= 2,360 + 209$$

$$= 2,569 \text{ years ago}$$

Thus, the separation time for Balinese and Balinese Malay can be described as follows:

- 1) Balinese and Balinese Malay began to separate between 2,354-2,157 years ago, or around 337 BC if counted from 2016.



2) The two languages began to separate from their proto about 2,569 years ago.

In terms of language classification, the percentage of genetic relationship results show that the two languages are a language family. It is also supported by data on the split time of the two languages, which are in the range (of 500 - 2,500) years, according to table 1.

Phonemic Correspondence

In order to strengthen the results of the genetic relationship of languages seen quantitatively, a qualitative analysis is also needed. One of them is by looking for phonemic correspondence set. Phonemic correspondence aims to find the equivalence of the sound of a phoneme from related words. The following phonemic correspondence analysis follows the procedure of Keraf (1996, pp. 52–58). Based on the analysis results, six sets of phonemic correspondences were obtained, which had the recurrence of the appearance of more than one pair of related words. The six sets are / ə – a /, / i – ə /, / a – ə /, / o – u /, / u – o /, and / h – r /. The following explains the emergence and distribution rules for each set.

In table 2, the phonemic correspondence set / ə – a / occurs in the pre-penultimate and ultima positions. The pre-penultimate position is preceded by the onset of the consonant /k/. In contrast, the ultima position is preceded by the onset of the consonants /p/ and /n/.

Table 2. phonemic correspondence set / ə – a /

Gloss	Balinese	Balinese Malay	Phonemic Correspondence
Father	[bapě]	[bapaʔ]	ə – a / K-
Because	[kəraně]	[karəna]	ə – a / K-

The phonemic correspondence set /i – ə/ in table 3 occurs in the penultimate position and is preceded by the onset of the consonants /b/ and /d/.

Table 3. phonemic correspondence set / i – ə /

Gloss	Balinese	Balinese Malay	Phonemic Correspondence
Seed	[binih]	[bənih]	/ i – ə / K-
Listen	[dirəh]	[dəŋər]	/ i – ə / K-

In table 4, the phonemic correspondence set /a – ə/ occurs in the ultima position. These vowels are preceded by the onset of the consonants /b/, /r/, and /t/.

Table 4. phonemic correspondence set / a – ə /

Gloss	Balinese	Balinese Malay	Phonemic Correspondence
Heavy	[bahat]	[bərat]	a – ə / K-
Give	[bahaŋ]	[bəri]	a – ə / K-
Because	[kəranə]	[karəna]	a – ə / K-
Egg	[təluh]	[təlor]	a – ə / K-

Based on the data in table 5, the phonemic correspondence set / o – u / occurs in penultimate and ultima positions. The ultima position is preceded by the onset of the consonants /b/, and in the ultima position, it is preceded by the onset of the consonants /r/ and /t/.

Table 5. phonemic correspondence set / o – u /

Gloss	Balinese	Balinese Malay	Phonemic Correspondence
Hunt	[boros]	[buru]	o – u / K-
That	[ənʔo]	[tu]	o – u / K-

Table 6 shows the phonemic correspondences /u – o/ occurring in the ultima position and preceded by the onset of the consonants /n/ and /k/.

Table 6. phonemic correspondence set / u – o /

Gloss	Balinese	Balinese Malay	Phonemic Correspondence
Lake	[danu]	[dano]	u – o / K-
Tail	[ikuh]	[ekor]	u – o / K-

The phonemic correspondence set / h – r / in table 7 occurs in the ultima position and is preceded by a vowel sound. This set is a strict rule given that the number of distributions is found in several related word pairs.

Table 7. phonemic correspondence set / h – r /

Gloss	Balinese	Balinese Malay	Phonemic Correspondence
Root	[akah]	[akar]	h – r / -#
Correct	[bənəh]	[bənər]	h – r / -#
Listen	[diŋəh]	[dəŋər]	h – r / -#
Tail	[ikuh]	[ekor]	h – r / -#
Egg	[təluh]	[təlor]	h – r / -#

CONCLUSION

Based on the results, it can be concluded that the percentage level of genetic relationship between the two languages is 36%. The two languages' split period was between 2,354 - 2,157 years ago. The two languages began to split from their proto about 2,569 years ago. If related to the level of genetic relationship of language, then both are included in the level of the language family (36-81%) and (500 - 2,500 years). Six sets of phonemic correspondences for the two languages were found, including /ə – a /, /i – ə /, /a – ə /, /o – u /, /u – o /, and /h – r /. Those findings qualitatively strengthen the findings on quantitative data. The genetic relationship of the two languages supports the genetic typology theory of the Austronesian language family of Blust, where both languages are the descendants of that family. This conclusion is drawn from the results of the genetic relationship of the two languages, which are still included at the level of the same language family or the (36-81%) and (2,500 - 5,000 years) levels. The genetic relationship status of the two languages, the same language family, also supports Adelaar's theory (2005), where both are descended from the Malay-Sumbawa language subgroup. Referring to the research results from Adelaar (2005), which showed a genetic relationship between Balinese and Malay at 35%, this also applies to the Malay variant, namely Balinese Malay has an almost equal percentage of 36%. It is also proof that although Balinese Malay is separated geographically from Malay, preserving the original Malay still exists in Balinese Malay.



REFERENCES

- Adelaar, K. A. (2005). Malayo-Sumbawan. *Oceanic Linguistics*, 44(2), 357–388.
<https://doi.org/10.1353/ol.2005.0027>
- Bagus, I. G. N., Denes, I. M., Laksana, I. K. D., Putrini, N., & Ginarsa, I. K. (1985). *Kamus Melayu Bali-Indonesia*. Jakarta: Pusat Pengembangan dan Pembinaan Bahasa.
<http://repositori.kemdikbud.go.id/id/eprint/2939>
- Blust, R. (2013). *The Austronesian Languages: Revised Edition*. Canberra: Asia-Pacific Linguistics.
- Denes, I. M. (1997). *Kamus Bahasa Indonesia-Bali*. Jakarta: Pusat Pengembangan dan Pembinaan Bahasa.
- Ersheidat, G., & Tahir, H. (2020). Genetic Relationship among Languages: An Overview. *International Journal of Language Education and Applied Linguistics*, 10(1), 17–27.
<https://doi.org/10.15282/ijleal.v10.3320>
- Hendrokusumo, & Temaja, I. G. B. W. B. (2019). The Genetic Relationship Between Ma'anyan and Malagasy (Hubungan Kekerabatan Bahasa Ma'anyan dan Bahasa Malagasi). *Metalingua*, 18(2), 113–124. <https://doi.org/10.26499/metalingua.v17i2.312>
- Keraf, G. (1996). *Linguistik Bandingan Historis*. Jakarta: Gramedia Pustaka Utama.
- Kridalaksana, H. (2008). *Kamus Linguistik*. Jakarta: Gramedia Pustaka Utama.
- Muhadjir. (2000). *Bahasa Betawi: Sejarah dan Perkembangannya*. Jakarta: Yayasan Obor Indonesia.
- Putra, A. A. P., Malini, N. L. N. S., & Suparwa, I. N. (2015). Adaptasi Kosakata Bahasa Bali dalam Bahasa Melayu Loloan Bali. *Litera*, 14(1), 26–36. <https://doi.org/10.21831/ltr.v14i1.4404>
- Sudaryanto. (2015). *Metode dan Aneka Teknik Analisis Bahasa: Pengantar Penelitian Wahana Kebudayaan secara Linguistik*. Yogyakarta: Sanata Dharma University Press.
- Temaja, I. G. B. W. B. (2017). Sistem Penamaan Orang Bali. *Humanika*, 24(2), 60–72.
<https://doi.org/10.14710/humanika.v24i2.17284>
- Temaja, I. G. B. W. B. (2018a). Cecimpedan: Teka-Teki dalam Bahasa Bali. *Kongres Bahasa Indonesia XI*. <http://repositori.kemdikbud.go.id/id/eprint/10138>
- Temaja, I. G. B. W. B. (2018b). Intransitive Verb Markers in Balinese. *Jurnal Arbitrer*, 5(2), 60–66.
<https://doi.org/10.25077/ar.5.2.60-66.2018>
- Temaja, I. G. B. W. B. (2018c). Pola Reduplikasi Bahasa Bali: Perbandingannya dengan Pola Reduplikasi Bahasa-Bahasa Austronesia. *Prasasti: Journal of Linguistics*, 3(2), 190–199.
<https://doi.org/10.20961/prasasti.v3i2.17520>



- Temaja, I. G. B. W. B. (2018d). Sapaan Kekerabatan dalam Bahasa Bali (Kinship Addresses in Balinese Language). *Metalingua*, 16(2), 211–220. <https://doi.org/10.26499/metalingua.v16i2.238>
- Temaja, I. G. B. W. B. (2021). Pentingnya Makna dan Nilai Budaya Konsep Kaja Kangin (Utara Timur) di Bali pada Anak. *Widya Kumara: Jurnal Pendidikan Anak Usia Dini*, 2(2), 189–198. <https://doi.org/10.55115/widyakumara.v2i2.1566>
- Tsaralunga, I. (2019). Methodological Principles of Diachronic Study of Linguistic Variation in the Ukrainian Language. *Advanced Education*, 13, 26–35. <https://doi.org/10.20535/2410-8286.154597>

