

PHONEMIC CORRESPONDENCES AND LEXICAL RELATEDNESS BETWEEN BAHASA INDONESIA AND CEBUANO

Roxanne Marie Flores

Universitas Gadjah Mada

roxannemarieflores@mail.ugm.ac.id

ABSTRACT

Using a historical-comparative linguistic approach, this study examines lexical relatedness and systematic phonemic correspondences between Bahasa Indonesia, the national language of Indonesia, and Cebuano, one of the most widely spoken Philippine regional languages, with particular attention to their relevance for *Bahasa Indonesia bagi Penutur Asing* (BIPA) in the Philippines. The study used an expanded Swadesh-based dataset of 221 core vocabulary items. Bahasa Indonesia forms were taken from the *Kamus Besar Bahasa Indonesia* (KBBI VI), while Cebuano forms were taken from Wolff's *A Dictionary of Cebuano Visayan* and supplemented by native-speaker consultation. Cognates were identified by cross-referencing the forms with Proto-Austronesian reconstructions and excluding confirmed loanwords. The analysis combined lexicostatistical calculation and qualitative phonemic correspondence analysis. A total of 81 cognate pairs were identified, producing a lexical similarity score of 36.65%. Seven systematic correspondences were found, including initial and final /h/ elision, vowel raising patterns, and correspondences involving /r/ and /j/. These findings support the shared Austronesian heritage of Bahasa Indonesia and Cebuano while also showing distinct phonological developments in each language. The study enriches Indonesian-Philippine comparative linguistics and provides an empirical basis for developing linguistically informed BIPA materials for Cebuano-speaking learners.

Keywords: Austronesian, BIPA, comparative linguistics, phonemic correspondence

INTRODUCTION

The *Bahasa Indonesia bagi Penutur Asing* (BIPA), 'Indonesian for Speakers of Other Languages' program, administered by the *Badan Pengembangan dan Pembinaan Bahasa* 'Agency for Language Development and Cultivation', has been developing both rapidly and intentionally in the Philippines over the past few years. Since its first deployment in 2016, the BIPA program in the Philippines has expanded to 14 institutions with 898 officially enrolled students by 2018 (Sukma, 2020). In 2023, that number has grown even further, reaching up to 35 institutions, with a total of 95 BIPA instructors officially assigned to the country, making the Philippines one of the largest recipient nations of BIPA instructors worldwide (Yonatan, 2024).

The relevance of Cebuano to BIPA is also supported by the geography of Indonesian-Philippine language contact. Indonesia maintains an embassy in Manila and a

Consulate General in Davao City, with the latter serving areas where Cebuano-speaking communities are prominent. This institutional presence makes Cebuano-speaking regions important sites for considering how Indonesian language teaching may be adapted to learners' regional linguistic backgrounds.

The Philippines is a highly multilingual country, with around 180 distinct native languages (Blust, 2013). Among these is Cebuano, pronounced /seb'wano/, the etymology of which is widely accepted to be derived from either *sibu* meaning 'port' (Irmano et al., 2020) or *sugbu* meaning 'dive into water' (Benitez, 2017) with the addition of the Spanish suffix *-ano* denoting a people group. Taken together, Cebuano from either *sibu-ano* or *sugbu-ano* roughly translates to 'people of the port' or 'people who dive into waters' respectively.

Cebuano is particularly significant in the Philippine linguistic landscape. The 2020 Census of Population and Housing indicates that Tagalog is the most widely used language

at home, while the Bisaya/Binisaya language group, with Cebuano as its dominant variety, is also used by millions of households. These figures place Cebuano among the largest ethnolinguistic communities in the Philippines and justify its inclusion in Indonesian-Philippine comparative linguistic research.

Despite this demographic significance, as Wolff (1972) notes, Cebuano falls behind Tagalog in national and institutional prestige. Cebuano is limited in formal education beyond the early primary years. When it comes to BIPA research and instructional materials, what is currently available, such as *Bahasa Filipino*, *Bahasa Sahabatku* (Ramadhani & Treyes, 2019) are primarily developed around Filipino/Tagalog/ This leaves the rest of the Philippine regional languages almost completely unexplored in educational materials that cater to BIPA constituents. Despite this demographic significance, Cebuano has received less national and institutional visibility than Tagalog (Wolff, 1972). Its limited formal use beyond the early primary years is reflected in available BIPA materials, which have primarily been developed around Filipino or Tagalog, such as *Bahasa Filipino*, *Bahasa Sahabatku* (Ramadhani & Treyes, 2019). This creates a practical gap for Cebuano-speaking BIPA learners, whose first-language resources remain underrepresented in instructional materials.

Existing scholarship may be grouped into three related strands: studies of Indonesian languages, studies of Philippine languages, and interlinguistic comparisons between Malay or Bahasa Indonesia and Tagalog/Filipino. The first two strands have produced valuable descriptions of separate linguistic contexts (Candria, 2022; Hendrokumoro, 2023; Prasetyo & Hendrokumoro, 2023; Ramadhan et al., 2023; Nelson, 2001; Reid & Liao, 2004; Robinson & Lobel, 2013; Jubilado, 2021; Ingilan & Jubilado, 2021). The third strand has shown that Bahasa Indonesia and Tagalog/Filipino share lexical and structural similarities (Arizo et al., 2020; Javier, 2021; Azizah et al., 2022; Raymunde et al., 2023; Astorga et al., 2024; Safitri et al., 2024). However, these studies do not yet provide a direct account of Bahasa Indonesia and Cebuano correspondences.

Only a limited number of studies have moved beyond Tagalog. Wardana et al. (2020) examined Malay, Tagalog, and Ilokano through lexicostatistics and glottochronology, while Jubilado and Manuli (2009) analysed Malay and Cebuano ditransitives. However, the Malay in Jubilado and Manuli's study refers to Bahasa Melayu, the national language of Malaysia, rather than Bahasa Indonesia. Consequently, a direct historical-comparative account of lexical relatedness and systematic phonemic correspondences between Bahasa Indonesia and Cebuano remains unavailable.

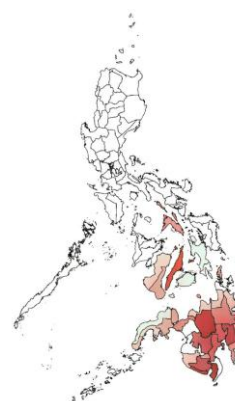


Figure 1. Map of the Cebuano-speaking regions in the Philippines with the higher concentration of speakers represented in darker red colour (retrieved from Thinking Machines)

Cebuano is spoken primarily on the island of Cebu and also across a wide geographic range of islands including Bohol and Siquijor, parts of Negros Oriental and Leyte, and much of northern and eastern Mindanao (Tramp, 1997). It belongs to the Central Philippine grouping within Malayo- Polynesian, similar to Tagalog. Pereltsvaig (2012) notes that Cebuano and Tagalog are arguably closer to each other than any other major languages in the Philippines.

Cebuano traces its origins from the Austronesian language family, whose speakers are believed to have migrated from Formosa (modern-day Taiwan) down to the Philippine archipelago, further into maritime Southeast Asia and further all the way to Madagascar in the west and Easter Island in Polynesia going east (Adelaar & Himmelmann, 2011; Pereltsvaig, 2021; Blust,

2019). The practical potential of this study is directly tied to this shared linguistic ancestry.

The gap matters pedagogically because lexical similarity has already been identified as one factor supporting Filipino learners' interest in BIPA (Sukma, 2020). Since Bahasa Indonesia and Filipino share Austronesian roots and display lexical and structural parallels (Raymunde et al., 2023; Astorga et al., 2024; Arizo et al., 2020; Safitri et al., 2024), extending the comparison to Cebuano can reveal whether similar linguistic bridges can be used for Cebuano-speaking learners.

Accordingly, this study addresses a specific research gap: the absence of an empirical comparison of Bahasa Indonesia and Cebuano based on cognate identification, lexical relatedness, and recurring phonemic correspondences. Addressing this gap is important not only for Austronesian comparative linguistics but also for BIPA instruction in Cebuano-speaking areas such as Cebu, Davao, and Cagayan de Oro, where teachers may benefit from understanding learners' first-language phonological resources.

This study has three main objectives: (1) to identify cognate pairs between Bahasa Indonesia and Cebuano; (2) to measure the degree of lexical similarity between the two languages using lexicostatistics, and (3) to analyse the recurring and systematic phonemic correspondences among the cognate pairs. This study has three main objectives: (1) to identify cognate pairs between Bahasa Indonesia and Cebuano; (2) to measure the degree of lexical relatedness between the two languages using lexicostatistics; and (3) to analyse the recurring and systematic phonemic correspondences among the cognate pairs.

METHOD

This study used a mixed-methods historical-comparative linguistic design, combining quantitative lexicostatistical computation with qualitative phonemic correspondence analysis. The quantitative component measured the degree of lexical relatedness between Bahasa Indonesia and Cebuano, while the qualitative component examined recurring sound correspondences among the identified cognate pairs. These two

components were integrated sequentially: cognate identification first generated the lexical dataset, lexicostatistical computation established the degree of relatedness, and phonemic correspondence analysis then explained the recurring sound patterns underlying that relatedness.

The primary dataset was an expanded Swadesh-based list of core vocabulary items. Although the Swadesh list is commonly referred to as a 200-item list, the working list initially contained 224 lexical prompts because several accepted variants and locally relevant core meanings were checked during data compilation. Three items, ice, snow, and freeze, were excluded because they refer to environmental phenomena not present in the tropical climates of Indonesia and the Philippines and therefore tend to yield loanwords rather than native lexical items. The final working dataset therefore consisted of 221 items. This procedure ensured that the dataset remained focused on stable core vocabulary while also being transparent about the difference between the conventional 200-item label and the actual working list used in the analysis.

Data for Bahasa Indonesia were collected from the *Kamus Besar Bahasa Indonesia* (KBBI VI), the official dictionary of the Indonesian language published by the *Badan Pengembangan dan Pembinaan Bahasa* 'Agency for Language Development and Cultivation'. Data for Cebuano were sourced from *A Dictionary of Cebuano Visayan*, compiled by John U. Wolff and published by Cornell University's Southeast Asia Program and the Linguistic Society of the Philippines. Each lexical item was checked against the relevant dictionary entry, and the basic lexical form most closely corresponding to the Swadesh meaning was selected for analysis.

To strengthen data credibility, the dictionary-based forms were verified through consultation with native speakers of Bahasa Indonesia and Cebuano. The consultation focused on confirming whether the selected forms were recognisable, semantically appropriate, and commonly understood in the intended meaning. When dictionary entries provided multiple possible forms, native-speaker feedback was used to confirm the most suitable form for the comparative dataset. This step was intended to reduce the

risk of selecting rare, archaic, or semantically mismatched lexical items.

Cognate identification was conducted by cross-referencing the collected Bahasa Indonesia and Cebuano forms with Proto-Austronesian (PAN) reconstructions in Wurm and Wilson's *English Finderlist of Reconstructions in Austronesian Languages* (1975). Word pairs were classified as cognates when both forms could plausibly be traced to a shared PAN form and when the relationship was supported by recurring sound patterns rather than superficial resemblance alone. Confirmed loanwords and forms without sufficient historical support were excluded from the analysis.

After the cognate identification process, lexicostatistical percentage was calculated by dividing the number of confirmed cognate pairs by the total number of items in the working dataset (221 items) and multiplying the result by 100.

Lastly, systematic phonemic correspondence analysis was conducted to identify recurring sound patterns between the two languages. A correspondence was considered systematic when the same sound relationship appeared across multiple cognate pairs and, where available, was supported by additional word pairs outside the Swadesh-based list. To enhance reliability, the analysis used triangulation across dictionary evidence, PAN reconstruction, native-speaker verification, loanword exclusion, and repeated pattern checking. These procedures were used to strengthen the validity of the cognate classification and the credibility of the identified phonemic correspondences.

FINDINGS AND DISCUSSION

The analysis addressed the three objectives of the study. First, 81 cognate pairs were identified from the 221-item working dataset. Second, the lexicostatistical computation produced a lexical similarity score of 36.65%.

Following Crowley and Bower's (2010) lexicostatistical classification, this percentage places Bahasa Indonesia and Cebuano within the level of language family subgrouping. This result should not be interpreted as evidence of mutual intelligibility; rather, it indicates that the two languages preserve a substantial set of historically related core vocabulary while also showing distinct phonological developments.

Third, the cognate pairs revealed seven systematic phonemic correspondences: (1) elision of /h/ in word-final position; (2) elision of /h/ in word-initial position; (3) vowel raising of /o/ to /u/; (4) vowel raising of /a/ to /u/ in a final syllable preceding /m/; (5) correspondence of /e/ to /i/, /u/, and /a/; (6) correspondence of /r/ to /g/, /l/, /d/, and /n/; and (7) correspondence of /j/ to /d/, /l/, /w/, and /Ø/. These patterns show that the relationship between the two languages is not limited to isolated cognate pairs but is supported by recurring and interpretable sound correspondences.

Taken together, the lexical and phonemic findings provide a clearer response to the research problem raised in the introduction. The 36.65% lexical similarity establishes the degree of relatedness, while the seven systematic correspondences explain how that relatedness is manifested at the sound level. For BIPA instruction, these findings suggest that Cebuano-speaking learners may encounter both points of familiarity and predictable areas of phonological difference when learning Bahasa Indonesia.

Phoneme elision: /h/ → /Ø/ - word-final

The most frequent phonemic correspondence between Bahasa Indonesia and Cebuano is the elision of the voiceless glottal fricative /h/ when it comes to word-final positions. There are eight (8) cognate pairs from the Swadesh list that demonstrate this correspondence.

Table 1. Phoneme elision: /h/ → /Ø/ - # word-final position

Gloss	PAN form	Indonesian	Cebuano
fruit	*/bu'ah/	/buah/	/bunga/
tongue	*/dilah/	/lidah/	/dila/
to spit	*/luvah/	/ludah/	/luwa/
white	*/putih/	/putih/	/puti/
full	*/pənuh/	/penuh/	/punu/

wet	*/bat'ah/	/basah/	/basa/
far	*/d'a'uh/	/jauh/	/layu/
ten	*/puluh/	/sepuluh/	/napulu/

All identified cognate pairs in this correspondence have PAN forms that have /h/ on word-final positions. This suggests that Bahasa Indonesia strongly retains PAN /h/ on word-final positions while Cebuano has innovated and developed without a final /h/. Other word pairs outside the Swadesh list that also exhibit this phonemic correspondence are *pilih~pili* ‘choose’,

panah~pana ‘arrow’, and *sembah~simba* ‘to worship’.

Phoneme elision: /h/ → /∅/ - word-initial

Another recurring phonemic correspondence between Bahasa Indonesia and Cebuano is also the elision of /h/ in word-initial positions. This is observed in six (6) items from the Swadesh list.

Table 2. Phoneme elision: /h/ → /∅/ - word-initial position

Gloss	PAN form	Indonesian	Cebuano
head	*/'ulu/	/hulu/	/ulu/
nose	*/ig'uŋ/	/hidung/	/ilung/
liver	*/'ataj/	/hati/	/atay/
to suck	*/t'əpt'ep/	/hisap/	/supsup/
rain	*/'udan/	/hujan/	/ulan/
black	*/'i(n)təm/	/hitam/	/itum/

In this correspondence, the PAN forms of the cognate pairs do not contain /h/ in word-initial positions. This suggests Bahasa Indonesia does not only retain /h/ in word-final positions, it has also innovated the addition of /h/ in word-initial position.

Vowel raising: /o/ → /u/

An instance of vowel correspondence between the two languages was identified wherein regardless of its position in the word, the mid-back vowel /o/ in Bahasa Indonesia raises to the high back vowel /u/ in Cebuano.

Wolff (1972) in his discussion of Cebuano's phonology and transcription states that Cebuano-Visayan operates with a three-vowel system: *a*, *i*, and *u*, and while some Cebuano publications occasionally write *i* as *e* and *u* as *o*, these alternations occur without consistency. Pesirla's (2003) paper on pedagogic grammar for Cebuano-Visayan also identifies a three-vowel nuclei: *a*, *i*, and *u*, and this three-vowel system is consistent with the orthographic and phonological conventions promoted by *Lubas sa Dagang Bisaya* (LUDABI), a

group dedicated to the preservation of traditional Cebuano language usage (Deiparine, 2025). However, as Deiparine (2025) also argues, /e/ and /o/ cannot just be reduced to mere variants of /i/ and /u/ especially in the cases of borrowed vocabulary like *edyukasyon* ‘education’ and *kompyuter* ‘computer’.

A five-vowel analysis on Cebuano indeed accommodates loanword phonology quite well, but since the data for this study is based on the Swadesh list and targets core native vocabulary instead of newly-borrowed words, the three-vowel system used in Wolff's 1972 dictionary is deemed more applicable for this study. It is eventually this three-vowel system that explains the correspondence between Bahasa Indonesia's mid-back vowel /o/ to Cebuano's high-back vowel /u/.

This vowel raising correspondence is observed in four (4) cognate pairs from the Swadesh list. It is also observed in other word pairs outside the Swadesh list like *ketok~tuktuk* ‘knock’, *bodoh~bugu/bubu* ‘stupid’, *tolak~tulak* ‘push’, *otak~utuk/utak* ‘brain’, and *tonjok~suntuk* ‘punch, hit’.

Table 3. Vowel raising: /o/ → /u/

Gloss	PAN form	Indonesian	Cebuano
tree	*/puqun/	/pohon/	/punuan/
tail	*/'ikuɣ/	/ekor/	/ikug/
to cut	*/pu[n]tuŋ/	/potong/	/putul/
to turn	*/balik/	/belok/	/liku/

Vowel raising: /a/ → /u/ - #m final syllable

Another instance of vowel raising was identified in which the low-central vowel /a/ in Bahasa Indonesia rises to the high-back vowel /u/ in Cebuano in final-syllable positions preceding the bilabial nasal /m/. This correspondence is observed in two (2) words from the Swadesh list.

Because only two cognate pairs from the Swadesh list exhibit this pattern, the evidence from the list alone is not sufficient to establish a systematic correspondence. However, further inspection of Bahasa

Indonesia-Cebuano cognate pairs outside the Swadesh list shows the same pattern in forms such as *asam~aslum* ‘sour’, *dalam~lalum* ‘deep’, and *tanam~tanum* ‘plant’.

The consistency of these word pairs outside the Swadesh list showing the same correspondence supports the Swadesh data as evidence of a genuine systematic correspondence, and not just mere coincidence, of vowel raising between Bahasa Indonesia /a/ to Cebuano /u/ before /m/.

Table 4. Vowel raising: /a/ → /u/ - __#m final syllable position

Gloss	PAN form	Indonesian	Cebuano
black	*/'i(n)təm/	/hitam/	/itum/
six	*/'ənəm/	/enam/	/unum/

Vowel correspondence: /e/ → /i/, /u/, /a/

Since native Cebuano operates with a three-vowel system, *a*, *i*, and *u*, it is unsurprising that in all the cognate pairs examined, none of the Cebuano forms retain the phoneme /e/. Several Cebuano words containing /e/ were initially identified, but were later on confirmed as loanwords and excluded from data analysis such as *ate* (from Chinese *aci* ‘sister’), *berde* (from Spanish *verde* ‘green’), and *karne* (from Spanish *carne* ‘meat’).

This is consistent with Deiparine’s (2025) observation that the addition of the vowels *e* and *o* can be properly acknowledged in Cebuano when it comes to loanwords. However, since this study looks at core and native vocabulary, all confirmed loanwords are excluded from analysis.

Regardless of its position within the word, /e/ in Bahasa Indonesia consistently corresponds to any of the native three-vowel system in Cebuano: either /i/, /u/, or /a/, with /e/ to /i/ correspondence occurring most frequently in eight (8) cognate pairs.

Table 5. Vowel raising: /e/ → /i/

Gloss	PAN form	Indonesian	Cebuano
you (singular)	*/kav/	/engkau/	/ikaw/
narrow	*/t'ə(m)pit/	/sempit/	/sigpit/
seed	*/bənih/	/benih/	/binhi/
egg	*/[t]əluy/	/telur/	/itlug/
tail	*/'ikuɣ/	/ekor/	/ikug/
neck	*/lihiɣ/	/leher/	/li-ug/
to turn	*/balik/	/belok/	/liku/
to squeeze	*/piRis/	/peras/	/pisil/

Bahasa Indonesia /e/ sometimes also goes through a process of vowel backing and rounding wherein it eventually corresponds

to /u/ in Cebuano. This is observed in six cognate (6) pairs from the Swadesh list.

Table 6. Vowel backing and rounding: /e/ → /u/

Gloss	PAN form	Indonesian	Cebuano
four	*'ə(m)pat/	/empat/	/upat/
six	*'ənəm/	/enam/	/unum/
heavy	*/bəɣat/	/berat/	/bug-at/
to swell	*/beNkeg/	/bengkak/	/but-uk/
to hear	*/dəŋəɣ/	/dengar/	/dungug/
full	*/pənuh/	/penuh/	/punu/

Lastly, /e/ in Bahasa Indonesia at times also lowers to the low-front vowel /a/ in

Cebuano. This is observed in four (4) pairs from the Swadesh list.

Table 7. Vowel lowering: /e/ → /a/

Gloss	PAN form	Indonesian	Cebuano
wide	*/lapa/	/lebar/	/lapad/
dust	*/ləbu'/	/debu/	/abug/
straight	*/tusug/	/terus/	/tanus/
ten	*/puluh/	/sepuluh/	/napulu/

Consonant correspondence: /r/ → /g/, /l/, /d/, /n/

The phoneme /r/ can be found in a number of words from the Cebuano Swadesh list but interestingly, it is absent from all identified cognate pairs between Bahasa Indonesia and Cebuano. Regardless of its position within the word, the voiced alveolar approximant /r/ in Bahasa Indonesia

consistently corresponds to a different sound in Cebuano, either /g/, /l/, /d/, or /n/. Among these correspondences, /r/ corresponding to the voiced velar stop /g/ happens the most frequently, occurring in eight (8) cognate pairs from the Swadesh list. Overall, twelve (12) pairs containing /r/ in Bahasa Indonesia have a different corresponding sound in Cebuano.

Table 8. Consonant correspondence: /r/ → /g/, /l/, /d/, /n/

Gloss	PAN form	Indonesian	Cebuano
heavy	*/bəɣat/	/berat/	/bug-at/
egg	*/[t]əluy/	/telur/	/itlug/
tail	*'ikuy/	/ekor/	/ikug/
neck	*/lihiɣ/	/leher/	/li-ug/
to hear	*/dəŋəɣ/	/dengar/	/dungug/
to sleep	*/tiduy/	/tidur/	/tulug/
new	*/bayu/	/baru/	/bag-u/
hundred	*/yatut'/	/ratus/	/gatus/
taste	*/rasa/	/rasa/	/lasa/
to squeeze	*/piRis/	/peras/	/pisil/
wide	*/lapa/	/lebar/	/lapad/
straight	*/tusug/	/terus/	/tanus/

Consonant correspondence: /j/ → /d/, /l/, /w/, /ø/

There are five cognate pairs containing the voiced postalveolar affricate [dʒ]

realized as /j/ in Bahasa Indonesia. In all their corresponding Cebuano forms, it is replaced by a different sound, either by /d/, /l/, /w/, or it is deleted altogether. This correspondence is consistent with the non-native phonological status of /j/ in Cebuano.

Wolff (1972) notes that the occurrence of /j/ in Cebuano is limited to loanwords, primarily from English. It surfaces in items like *dyip* ‘jeep’ and *dyus* ‘juice’, and even then, the phoneme /j/ is transcribed initially and medially as *dy*, and in final position as *ds*, such as *dyurds* /jurj/ for the name ‘George’. This non-native status of /j/ is

supported by the Cebuano Swadesh list data, wherein across all 221 items, no instance of /j/ was observed in the Cebuano items.

In comparison with the PAN reconstructions, there are additional insights that can be gleaned from this pattern. In all five cognate pairs, the corresponding PAN forms contain /d/. This suggests that the two languages may have diverged in their development of PAN /d/. Cebuano retained or innovated PAN /d/ reflexes without developing /j/. Meanwhile Bahasa Indonesia more frequently reflects PAN /d/ as /j/.

Table 9. Consonant correspondence: /j/ → /d/, /l/, /w/, /Ø/

Gloss	PAN form	Indonesian	Cebuano
road	*/d'alan/	/jalan/	/dalan/
bad	*/d'ahat/	/jahat/	/daut/
far	*/d'a'uh/	/jauh/	/layu/
rain	*/'udan/	/hujan/	/ulan/
to sew	*/d'ahit/	/jahit/	/uwan/
			/tahi/

Completely identical lexical forms

Lastly, the following cognate pairs between Bahasa Indonesia and Cebuano are forms with no identified phonemic variations. There are fourteen (14) items of these identical cognate pairs from the

The identical forms are pedagogically significant because they provide immediate lexical bridges for Cebuano-speaking BIPA learners. In classroom practice, such items may be used to introduce learners to the idea that Bahasa Indonesia and Cebuano share historically related vocabulary, while the

Swadesh list. Other completely identical word pairs between Bahasa Indonesia and Cebuano include: *angkat~angkat* ‘to lift’, *buta~buta* ‘blind’, *kilat~kilat* ‘thunder’, *sesuai~sesuai* ‘in accordance’, and *sama~sama* ‘same, like’.

non-identical cognates can be used to demonstrate predictable sound correspondences. This supports a more linguistically informed approach to BIPA instruction rather than treating Indonesian vocabulary as entirely unfamiliar to Cebuano-speaking learners.

Table 10. Identical lexical forms between Bahasa Indonesia and Cebuano

Gloss	PAN form	Indonesian	Cebuano
I	*/'aku/	/aku/	/aku/
other	*/laqin/	/lain/	/lain/
five	*/lima/	/lima/	/lima/
child	*/'anak/	/anak/	/anak/
lice	*/kutu/	/kutu/	/kutu/
eye	*/mata/	/tidur/	/tulug/
fingernail	*/kuku/	/kuku/	/kuku/
breast	*/t'ut'u/	/susu/	/susu/
to tie	*/d'alin/	/tali/	/tali/
moon	*/bulan/	/bulan/	/bulan/
stone	*/batu/	/batu/	/batu/
lake	*/danav/	/danau/	/danaw/

sky	*/laŋit/	/laŋit/	/laŋit/
ash	*/'abu/	/abu/	/abu/

CONCLUSION

This study identified 81 cognate pairs between Bahasa Indonesia and Cebuano from a 221-item Swadesh-based dataset, producing a lexicostatistical similarity score of 36.65%. This result places the two languages within the level of language family subgrouping and demonstrates that their relationship is reflected not only in shared Austronesian ancestry but also in a measurable set of historically related core vocabulary.

The study also identified seven systematic phonemic correspondences, including the elision of /h/ in word-final and word-initial positions, vowel raising patterns, and consonant correspondences involving /r/ and /j/. These findings show that the relationship between Bahasa Indonesia and Cebuano is not merely based on isolated lexical similarity but is supported by recurring sound patterns. The results

therefore enrich Austronesian comparative linguistics while providing a more specific account of Indonesian-Philippine interlinguistic relatedness beyond the more frequently studied Tagalog/Filipino context.

Pedagogically, the findings provide a useful basis for developing BIPA materials for Cebuano-speaking learners in the Philippines. Awareness of these sound correspondences can help BIPA instructors anticipate possible pronunciation patterns, highlight familiar cognates, and design vocabulary activities that build on learners’ existing linguistic knowledge. Future research may expand the dataset beyond core vocabulary, examine morphological and syntactic parallels, and include other Philippine regional languages, such as Ilokano, Bikol, and Maguindanaon, whose speech communities may also be relevant to BIPA development.

REFERENCES

Adelaar, A., & Himmelmann, N. (Eds.). (2011). *The Austronesian languages of Asia and Madagascar*. Routledge.

Arizo, C. D., Palayon, A. L., Tornito, A. P. V., & Sukma, B. P. (2020). Comparative analysis of Filipino and Indonesian monophthongs. *Lexeme: Journal of Linguistics and Applied Linguistics*, 2(1). <http://openjournal.unpam.ac.id/index.php/LJLAL>

Astorga, J. M. S., Cortes, C. J. S., Angkico, C. J. P., Alvarez, S. R., Paragua, K. G., Recto, T. F., & Mancio, J. M. Y. (2024). Phonological similarities between Tagalog and Bahasa Indonesia using phonology as human behavior. *Ignatian International Journal for Multidisciplinary Research*, 2(12). <https://doi.org/10.5281/zenodo.14305001>

Azizah, P. N., Rosidin, O., Anggraini, A., & Devi, K. (2022). Bentuk dan makna kosakata Bahasa Indonesia dan Bahasa Tagalog dalam lirik lagu Zack Tabudlo “Asan Ka Na Ba.” *Pena Literasi*, 5(2), 161–170. <https://doi.org/10.24853/PL.5.2.161-170>

Benitez, J. (2017). Historicity of city and town names in the Province of Cebu. *CNU Journal of Higher Education*, 11(1), 1–12. <https://www.researchgate.net/publication/354545132>

Blust, R. (2013). *The Austronesian languages*. Asia-Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

Blust, R. (2019). The Austronesian homeland and dispersal. *Annual Review of Linguistics*, 5(1), 417–434. <https://doi.org/10.1146/annurev-linguistics-011718-012440>

Candria, M. (2022). Telaah linguistik historis komparatif terhadap bahasa Indonesia, Jawa, Madura, dan Bali. *Endogami: Jurnal Ilmiah Kajian Antropologi*, 6(1), 67–75.

- <https://doi.org/10.14710/endogami.6.1.67-75>
- Crowley, T., & Bower, C. (2010). *An introduction to historical linguistics* (4th ed.). Oxford University Press.
- Deiparine, M. S. (2025). Cebuano-Visayan interference and word articulation of English phonology learners. *World Journal of English Language*, 16(1), 112–119.
<https://doi.org/10.5430/wjel.v16n1p112>
- Hendrokumoro. (2023). Kekerabatan bahasa Jawa dan Bali. *Deskripsi Bahasa*, 6(1), 19–27.
<https://doi.org/10.22146/db.v6i1.5772>
- Ingilan, S., & Jubilado, R. (2021). The state of Tausug and Sama-Bajau linguistics. *Hawai'i Journal of the Humanities*, 2(1), 99–115.
<https://hilo.hawaii.edu/humanities/journal/issues/documents/humanities/journal/issues/TheStateofTausugandSama-BajauLinguistics.pdf>
- Irmano, K. J. V., Sitoy, E. D., Delgado, A. R., Lozano, E. B., & Inocian, R. B. (2020). Cebu's places named after endemic flora: Basis for innovative teaching models. *The International Journal of Social Sciences and Humanities Invention*, 7(10), 6242–6255.
<https://doi.org/10.18535/ijsshi/v7i010.06>
- Javier, J. R. (2021). Komparatibong pagsusuri ng applicative/causative marking ng Bahasa Indonesia at morphological focus marking ng Tagalog [Comparative analysis of applicative/causative marking in Bahasa Indonesia and morphological focus marking in Tagalog]. *Daluyan: Journal ng Wikang Filipino*, 27(1), 28–44.
<https://journals.upd.edu.ph/index.php/djwf/article/view/8589>
- Jubilado, R. (2021). Comparative ergative and accusative structures in three Philippine languages. *Southeastern Philippines Journal of Research and Development*, 26(1), 1–18.
<https://doi.org/10.53899/spjrd.v26i1.121>
- Jubilado, R., & Manuli, M. (2009). Malay and Cebuano ditransitives: A minimalist perspective. *Journal of Modern Languages*, 19(1), 85–98.
<https://ejournal.um.edu.my/index.php/JML/article/view/3617>
- Nelson, H. (2001). Lexicostatistics applied to the historical development of three languages of the Philippines. *Deseret Language and Linguistic Society Symposium*, 27(1).
<https://scholarsarchive.byu.edu/dlls/vol27/iss1/7>
- Pereltsvaig, A. (2021). *Languages of the world: An introduction* (3rd ed.). Cambridge University Press.
<https://doi.org/10.1017/9781108783071>
- Philippine Statistics Authority. (2020). *Technical notes, 2020 census of population and housing: Tagalog is the most widely spoken language at home*.
<https://psa.gov.ph/content/tagalog-most-widely-spoken-language-home-2020-census-population-and-housing>
- Prasetyo, G. T., & Hendrokumoro (2023). Kekerabatan bahasa Jawa dan bahasa Batak Simalungun: Kajian linguistik historis komparatif. *Deiksis*, 15(3), 322–334.
<https://doi.org/10.30998/deiksis.v15i3.18362>
- Ramadhan, J. A., Anu, C., Mulyadi, & Nurhidayati, S. A. (2023). Linguistik historis komparatif dalam perbandingan rekonstruksi bahasa Mandailing dan bahasa Melayu. *Jurnal Pendidikan Multidisipliner*, 6(12), 307–313.
<https://edu.ojs.co.id/index.php/jpm/article/view/73>
- Ramadhani, R. P., & Treyes, R. A. (2019). *Bahasa Filipino, Bahasa Sahabatku*:

- Bahan belajar bahasa asing pendukung diplomasi kebahasaan.* Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan. <https://repositori.kemendikdasmen.go.id/20906/1/Bahasa%20Filipino%20Bahasa%20Sahabatku.pdf>
- Raymunde, P., Caangay, R., Mamonong, V., & Ponce, M. (2023). Tagalog, Bahasa Indonesia, and Bahasa Melayu language: A comparative analysis. *Psychology and Education: A Multidisciplinary Journal*, 14(5), 586–605. <https://ejournals.ph/article.php?id=21821>
- Reid, L., & Liao, H. C. (2004). A brief syntactic typology of Philippine languages. *Language and Linguistics*, 5(2), 433–490. <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/55ad1381-5025-418b-bd1e-03e9703ca6a9/content>
- Robinson, L. C., & Lobel, J. W. (2013). The Northeastern Luzon subgroup of Philippine languages. *Oceanic Linguistics*, 52(1), 125–168. <https://research.ebsco.com/c/wel7c4/viewer/pdf/2b23usbavr?route=details>
- Safitri, H., Ariani, N., & Pratiwi, D. I. (2024). Reduplication in Tagalog and Indonesian language (Bahasa): A comparative study. *Jurnal Onoma: Pendidikan, Bahasa, dan Sastra*, 10(2), 1829–1843. <https://ouci.dntb.gov.ua/en/works/4Eey1gz4/>
- Sukma, B. P. (2020). Potensi dan tantangan penginternasionalan Bahasa Indonesia di Filipina. *Widyaparwa*, 48(1), 1–13. <https://widyaparwa.kemendikdasmen.go.id/index.php/widyaparwa/article/view/435>
- Thinking Machines. (n.d.). *The language landscape of the Philippines in 4 maps*. Retrieved February 5, 2026, from <https://stories.thinkingmachines.philippine-languages>
- Wardana, M. K., Widayati, D., Taib, R., & Iqbal, M. (2020). Lexicostatistics of Malay, Tagalog, and Ilocano languages: A comparative historical linguistic study. *Jurnal Education and Development*, 10(3), 475–479. <https://journal.ipts.ac.id/index.php/ED/article/view/4259>
- Wolff, J. U. (1972). *A dictionary of Cebuano Visayan*. Cornell University Southeast Asia Program & Linguistic Society of the Philippines. <https://www.gutenberg.org/files/40074/40074-h/40074-h.htm>
- Wurm, S. A., & Wilson, B. (1975). *English finderlist of reconstructions in Austronesian languages*. Pacific Linguistics.
- Yonatan, A. Z. (2024, August 31). Tujuh negara dengan jumlah lembaga BIPA terbanyak di dunia. GoodStats. Retrieved January 2, 2026, from <https://www.goodstats.id/article/7-negara-dengan-jumlah-lembaga-bipa-terbanyak-di-dunia-k0E1W>