

Research Article

Reconstructing the etymology of Philippine numerals' protolanguage: a socio-phonemic approach

Sherwin E. Balbuena ¹ , Princess Z. Balbuena ¹ , Michael L. Estremera ^{1,*} ¹ Dr. Emilio B. Espinosa Sr. Memorial State College of Agriculture and Technology, Philippines* Correspondence: michaelestremera@debesmscat.edu.ph<https://doi.org/10.59652/jcpas.v3i4.658>

Abstract: The Philippines is home of a diversely tongued people who were already numerate and literates even before the Spanish colonization. These people communicated leveraging protolanguage which evolved through time via linguistic change into various languages and/or dialects. This paper attempts to partially reconstruct the mother language, with specific focus on the possible sounds of counting terms used by primitive Filipinos. Using the comparative-interpretive methods, it explored twelve number concepts in nine regional languages whose sound correspondences were analyzed to deduce the phonemics of the etymological terms. Findings reveal that there were indeterminate vowel sounds which evolved from the mother language of the first four and the sixth counting terms, while there was no problem with reconstructing the etymology of the fifth and the seventh to ninth counting lexical terms. Considering some of reconstructed lexical numerals, the origin of the explored protolanguage for hundred and thousand were likewise reconstructed.

Keywords: comparative-interpretive method; counting numbers; etymology; socio-phonemic; protolanguage

1. Introduction

Historically, Filipinos were already numerate and literates even before Spanish colonizers reached the Visayan shores (Stošić et al., 2016). They could count up to millions and used units of measure for volumes and lengths (Zaide, 1999). They also used *baybayin* scripts (composed of three vowels and fourteen consonants) to write their thoughts and feelings. Due to the predominant oral traditions in prehistoric Philippines, the protolanguage was inaccurately conveyed and/or received, and eventually evolved with continued use over time (Planer, 2017; Hartmann & Pleyer, 2021). On the other hand, phonemics, a central area within linguistics, emphasizes on the intangible units of sound called phonemes, which serve to differentiate meaning in articulated language. The term “phoneme” comes from the Greek *phōnēma*, meaning sound or voice (Naqsyah & Husna, 2024). Moreover, phonemics examines how these sounds function within specific vernaculars, impelling pronunciation and semantic understanding (Kazanina et al., 2017). Deciphering phonemics is indispensable for language progress, as it helps students in identifying and producing sounds precisely. The relationship between phonemics and etymology divulges the historical evolution of sound schemes and their bearing on language change. Current studies highlight the significance of phonemic consciousness in literacy growth, particularly in early education (Duncan, 2018). Additionally, developments in computational linguistics boost our cognition of phonemic structures across various tongues.

This academic piece of work is anchored in socio-phonemic scientific field having ethno-linguistic theory as a sub-field that reconnoiters the relationship between language and cultural identity, highlighting how phonemic dissimilarities reveal ethnic associations (Alshehri, 2023). This theory is predominantly related to etymology and phonemics, as understanding the origins of words can divulge insights into how phonemes progress within explicit cultural frameworks (Fatima & Nadeem, 2025). For example, phonemic awareness is fundamental for language acquisition and can vary meaningfully across linguistic experiences, impacting literacy progress. Moreover, the historical development of phonemes often reflects social changes within societies, emphasizing the implication of etymological researches in

Received: July 27, 2025

Accepted: October 25, 2025

Published: November 6, 2025

**Copyright:** © 2022 by the authors.Submitted for open access publication
under the terms and conditions of the
Creative Commons Attribution (CC BY)
license[\(https://creativecommons.org/licenses/by/4.0/\)](https://creativecommons.org/licenses/by/4.0/).

understanding linguistic individuality (Zharkynbekova et al., 2025). By scrutinizing these associations, scientists can better comprehend how language evolves and is molded by cultural dynamic forces.

By discovering phonemics, people increase their perceptions into the perceptive procedures that trigger language use and communication (Aljasser & Ismail, 2025). Despite the all-encompassing studies on phonemics, there remains a noteworthy gap in comprehending fully how etymological advances influence phonemic constructions across diverse languages. While studies have sightseen phoneme registers and their applications, the association between historic sound changes and modern phonemic structures has not been systematically scrutinized (Wang, 2025). Therefore, this study tries to trace the ancestor language, focusing on the mathematical knowledge of primitive Filipinos, attempting to exemplify pioneering efforts for the reconstruction of Filipino protolanguage thereby offering philosophical answers to following questions: (1) What are the protolanguage numerals being used by the different regional language groups in the country? (2) What are the correspondences in the cognates of the numerals? (3) What is the probable etymological term of each numeral?

2. Materials and Methods

Since the present exploration falls under qualitative method involving etymology and phonology, researchers leveraged the use of an interpretive research design particularly appropriate (Hartman & Squires, 2024). This design permits researchers to sightsee the meanings and familiarities associated with language use, concentrating on how individuals perceive and articulate their linguistic identities. By employing methods such as interviews, focus groups, and textual analysis, researchers delved into the historical contexts of phonemes and their evolution over time, providing rich insights into cultural and social dimensions of language. To some degree, a case study approach was also employed in scrutinizing specific instances of phonological change or etymological shifts within specific communities or linguistic groups. These techniques enable an in-depth understanding of how phonological structures interact with cultural identity and historical narratives, permitting for a nuanced examination of language changing aspects.

3.1. *Anthropological Context*

Local counting terms being used by some ethno-linguistic groups in the Philippines were the subjects of this study (Mirasol, 2021). The terms used were limited only to the first ten natural numbers (1 through 10) and equivalents for the number concepts hundred and thousand. These twelve numbers will be called numerals throughout this paper. Looking up available dictionaries, related studies and textbooks, the local equivalents of these terms were identified. However, the phonetic sounds of these terms were not available in the references. But this may not pose a lot of problems in the investigation since the terms are spelled as they are pronounced (Lamorinas et al., 2024). In effect, nine regional languages across the Philippine archipelago were chosen to comprise the subjects of this study, namely: Ilokano, Pampango, Pangasinense, Tagalog, Bikolano, Masbatenyo, Waray Waray, Cebuano, and Hiligaynon. The first five languages are spoken in Luzon Region and the following four in Visayan Region. However, Cebuano is also spoken as the lingua franca in some parts of Northern Mindanao. These languages are spoken by relatively greater number of Filipinos (figure 1). Using comparative-interpretive methods of linguistic analysis, the local terms were examined by identifying regular correspondences in the cognates of the sample related languages (Konig, 2012). Correspondence means comparison of the genetically related sound systems of two or more words. Cognates are words in related languages which can be traced from the same ancestral origin.

For each correspondence, the most likely sound in the parent term is deduced. In this manner, much of the sound characteristics of the parent can be partially reconstructed. Although there are two ways by which the parent term can be determined, either using the 'majority' rule or combining the sounds of syllables with irregular correspondence (Niolaki et al., 2022), this study adopts a most appropriate rule to deduce the most probable sound structure of the parent term. It must be noted that the Philippines is home of approximately 150 languages. To use the majority rule, a study like this one must include a sizeable number of language groups to arrive at more precise results. Since only nine language groups were used, this study may consider other rules apart from the majority rule.

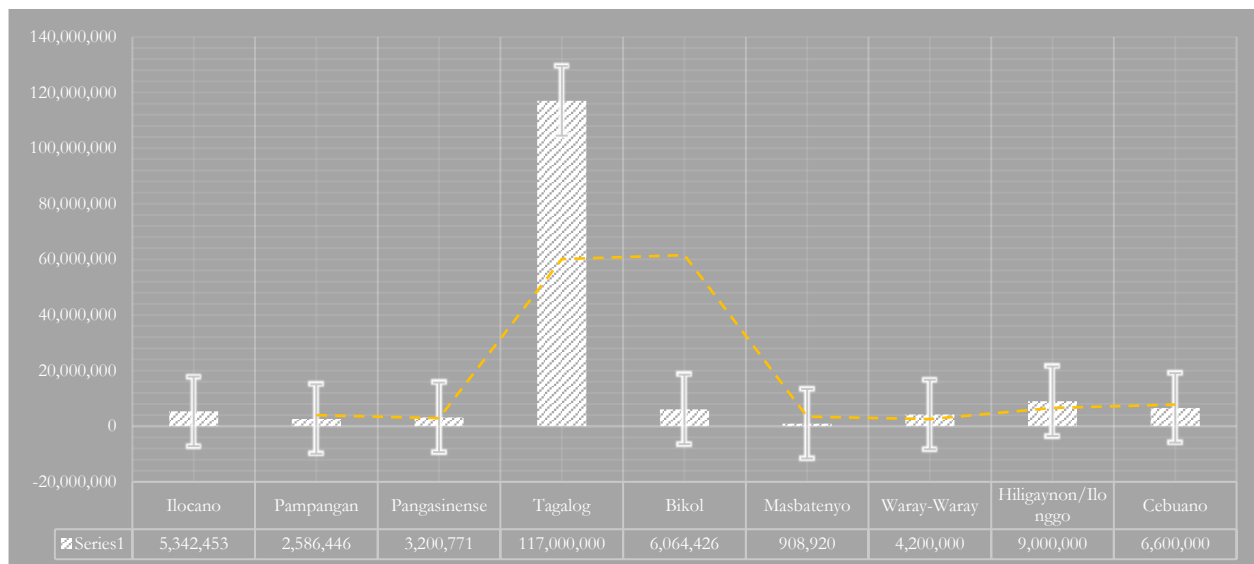


Figure 1. Number of speakers of each regional language.
Source: Ethnologue, 2024.

3.2. Data Processing

As highlighted, this study is qualitative in nature and adopted nine steps when explicating ethno-linguistic and phonemics variables (Archibald, 2025; Del Mundo & Velasco, 2025). The steps in data processing are shown in figure 2. In fact, for the current undertaking, the first step in data processing involves clearly defining the research objectives related to etymology and phonemics within ethno-linguistic studies.



Figure 2. Steps in data processing.

This includes identifying specific questions about how cultural heritage influences language evolution and phonemic variation. Establishing these objectives guides subsequent research activities (Celata & Nagy, 2022). Secondly, researchers conducted a comprehensive literature review is essential for understanding existing theories and frameworks in etymology and phonemics. This review should focus on prior studies that explore the relationship between language, culture, and identity, providing a foundation for the current research (Tamariz & Kirby, 2016). Then, the data collection involves gathering both qualitative and quantitative data relevant to the research questions. This may include historical texts, phonetic transcriptions, and interviews with speakers from specific cultural backgrounds. Utilizing diverse data sources enriches the analysis (Hepburn & Bolden, 2017). As for data Organization, researchers concerned collected data systematically for effective analysis. Researchers categorized the data based on linguistic features, such as phonemic inventories

and etymological roots, while also noting socio-cultural contexts that may influence language use (Harrington, 2023). Moreover, the qualitative analysis of the study focused on understanding the meanings behind phonemic variations and etymological changes. Researchers can employ thematic analysis to identify patterns related to cultural identity and heritage within the data (Tagg & Manford, 2025).

Similarly, utilizing case studies allowed for researchers to conduct in-depth examinations of specific instances where etymology and phonemics intersect with cultural heritage. This method provides rich contextual insights that broader analyses may overlook (Fraser, 2022). As far as interpretation of findings, it involved synthesizing qualitative results to draw conclusions about the relationship between language and cultural identity. Researchers should consider how their findings align with or challenge existing theories in ethno-linguistics (Vargas-Bianchi, 2025). The results are reported in a clear and structured manner, ensuring that both qualitative narratives and quantitative data are presented effectively. This includes discussing implications for understanding cultural heritage through language (Wood et al., 2024). In the long run, based on the findings, researchers likewise provided recommendations for future studies that could further explore the intersections of etymology, phonemics, and cultural identity. This may involve suggesting new methodologies or areas of focus that could enrich the field.

3.3. Data Analysis

In order to explicate the prototype Philippine numerals, researchers collectively utilized interpretive and case study approaches to facilitate in-depth investigation of the linguistic phenomena under study. Interpretive analysis lets scientists to comprehend how cultural circumstances influence phonemic dissimilarities and etymological shifts (Hartman & Squires, 2024). Conversely, case study for this endeavor, offered comprehensive insights into specific occurrences of phonological variation, emphasizing the relationship between historical linguistics and modern usage. By exploring individual cases, researchers uncovered the intricacies of language progress and its socio-cultural repercussions. These analyzes techniques enriched the understanding of how phonemes and word ancestries shape linguistic uniqueness within various societies.

3.4. Ethical Considerations and Protocol

The academic piece guarantees ethical accuracy by respecting native dialectal inheritance, gaining informed agreement from research informants, and maintaining social sympathy. Data gathering follows privacy etiquettes, with clear documents and acknowledgement (Sivasubramaniam et al., 2021). Results evade from distortion and acknowledge etymological groups as co-authors of data, guaranteeing ethical exchange and veracity.

3.5. Artificial Intelligence (AI) Tools Declaration

In the completion of this anthropological inquiry, AI tools specifically Perplexity AI and ChatGPT were used to improve the substance of the study and ensured that ideas reflected are in harmony with the existing body of knowledge (Gerlich, 2025). Be that as it may, authors guarantee that perspectives and key findings contained in this study are originally created and collaboratively generated by the concerned authors.

3. Results

With the aim of partially reconstructing the original terms used in counting by primitive Filipinos, this study presents nine language groups across the Philippine archipelago whose sound systems are believed to be the descendants of the ancestral Filipino language (Paz, 2008).

4.1. Number concept equivalents

Table 1 presents the local counting terms/concepts currently being used in various ethno-linguistic communities in the country (Mirasol, 2021). Only twelve number concepts or terms were identified to be the most significant subjects of this study, namely: one, two, three, four, five, six, seven, eight, nine, ten, hundred, and thousand. The rest of the counting terms, common to nearly all of these language groups, are then derived from these twelve number concepts by means of conjunction; that is, ten and two for number twelve, two tens and four for number twenty-four, and so on. For example, Cebuano speakers use *napulo ug usa* (ten and one) for eleven, *duba kanapulo ug tulo* (two tens and three) for twenty-three, and so on. But due to the verbosity of these counting terms, local speakers tend to use the Spanish numerals, which are concise and easier to pronounce. The Tagalog numerals are almost

related to other regional languages (Lamorinas et al., 2024). The first nine numerals have consistent correspondences with other numerals. The word *daan* for “hundred” is unique.

Table 2. Number concept equivalents in various regional languages in the Philippines.

Number	English	Tagalog	Bicolano	Pampango	Pangasinense	Ilokano	Hiligaynon
1	One	Isa	Saro	Metung	Sakey	Maysa	Isa
2	Two	Dalawa	Duwa	Adwa	Duara	Dua	Duha
3	Three	Tatlo	Tulo	Atlu	Talora	Tallo	Tatlo
4	Four	Apat	Apat	Apat	Apatira	Uppat	Apat
5	Five	Lima	Lima	Lima	Limara	Lima	Lima
6	Six	Anim	Anom	Anam	Anemira	Innem	Anom
7	Seven	Pito	Pito	Pitu	Pitora	Pito	Pito
8	Eight	Walo	Walo	Walu	Walora	Walo	Walo
9	Nine	Siyam	Siyam	Siam	Siyamira	Siam	Siyam
10	Ten	Sampu	Sampulo	Apulu	Samplura	Sangapulo	Pulo
x100	Hundred	Daan	---	---	Lasus	Gasut	Gatus
x1000	Thousand	Libo	Ribo	Libu	Sanlibo	Sangaribu	Libo

Number	English	Cebuano	Masbatenyo	Waray	City of Sorsogon	Sorsoganon Bacon	2nd District of Sorsogon
1	One	Usa	Usad	Usa	Isad	Saru	Sayu
2	Two	Duha	Duha	Duha	Duwa	Duwa	Duwa
3	Three	Tulo	Tulo	Tulo	Tulo	Tulo	Tulo
4	Four	Upat	Upat	Upat	Upat	Apat	Upat
5	Five	Lima	Lima	Lima	Lima	Lima	Lima
6	Six	Unum	Unom	Unom	Unom	Anum	Unom
7	Seven	Pito	Pito	Pito	Pito	Pito	Pito
8	Eight	Walo	Walo	Walo	Walu	Walu	Walu
9	Nine	Siyam	Siyam	Siyam	Siyam	Syam	Seyam
10	Ten	Napulo	Napulo	Napulo	Napulo	Sampulo	Napolo
x100	Hundred	Gatus	Gatos	Gatos	Gatus	Gatus	Gatos
x1000	Thousand	Libo	Libo, Ribo	Yukot	Sangribo	Sangribu	Sangrebu

It can be noted in table 1 that there are no problems with reconstructing the parent terms for the number concepts five, seven, eight, and nine, as their correspondences have systematic and consistent relationships (Paz, 2008). Some only differ in spelling, but essentially, they are of the same sound systems. Hence, the parent terms for these number concepts can be reconstructed as *lima*, *pitu*, *walu*, and *siyam*, respectively. We used only the vowel sounds *a*, *i*, and *u* because archaeological evidence shows that there were only three vowels in the primitive scripts *baybayin*. The later introduction of vowel sounds *o* and *e* was done to integrate foreign sounds and borrowed words to the local language. This phenomenon may be described using the Pepet Law (Conant, 1913). According to this law, in the vocalism of Indonesian languages the original indifferent vowel plays an important role. Resembling the Hebrew *shēwa*, and the obscure vowel of many Indo-European languages, it was so colorless and indefinite in pronunciation that it developed differently in different speech groups (Mirasol, 2021; Lamorinas et al., 2024). In some languages it remained practically unchanged, as in Javanese, where it is called *pepet*, while in others it evolved into various and more or less definite vowel sounds, e. g., *atēp* roof became *Jav. atēp*, *Mai. atap*, *Tag. atip*, and *Bis. atup*. The parent term for the number concept ‘one’ has evolved tremendously into various descendants *isa*, *usa*, *usad*, *maysa*, *saro*, and *metung*. But one remarkable correspondence [sa]-[sa]-[ro]-[tung] can be noted, which means that the syllabic sound [sa] has been preserved and can be considered, by majority rule, one sound component of the parent term. One concern is the identification of definite quality of the other sound component using the irregular correspondence [i]-[u]-[u]-[may]-[sa]-[me].

4.2. Partial Reconstruction Concept Equivalents

The above investigation of numerals through a socio-phonemic lens allows for a deeper understanding of how language functions as a vehicle for cultural heritage. For instance, the Tagalog counting system, which merges traditional and colonial influences, illustrates how

external factors, such as Spanish colonization, have impacted local linguistic practices (Mirasol, 2021; Lamorinas et al., 2024). The adaptation of counting methods and numeral systems signifies a broader narrative of cultural exchange and transformation. Moreover, examining the etymology of Philippine numerals unveils connections to broader Austronesian linguistic patterns. For example, the numeral *lima*, which is consistent across many Austronesian languages, serves as a reminder of shared ancestry and collective identity among these diverse cultures. This shared linguistic heritage fosters a sense of belonging and continuity among speakers, reinforcing cultural ties that transcend geographical boundaries. Table 2, conversely, highlights the probability that the first syllable for the native term for the first counting number was being pronounced as an indefinite sound.

Table 2. Partial reconstruction of Proto-Philippine numerals.

Numeral	Parent term	Correspondence [Tgl-Bik-Pam-Pag-Ilo-Hil-Ceb-Msb-War]		
		First syllable	Second syllable	Third syllable
1	*Vsa	[i]-[sa]-[me]-[sa]-[may]-[i]-[u]- [u]-[u]	[sa]-[ro]-[tung]-[key]-[sa]-[sa]- [sa]-[sad]-[sa]	
2	*dVa	[da-du-ad-du-du-du-du-du- du]	[la]-[wa]-[wa]-[a]-[a]-[ha]-[ha]- [ha]-[ha]	[wa]-[]-[]-[ra]-[]-[]-[]- []
3	*tVlo	[tat]-[tu]-[at]-[tal]-[tal]-[tat]- [tu]-[tu]-[tu]	[lo]-[lo]-[lu]-[lo]-[lo]-[lo]-[lo]- [lo]-[lo]	Pag. [ra]
4	*Vpat	[a]-[a]-[a]-[a]-[up]-[a]-[u]-[u]- [u]	(Systematic and consistent)	Pag. [ira]
5	*lima	(Systematic and consistent)	(Systematic and consistent)	Pag. [ra]
6	*VnVm	[a]-[a]-[a]-[a]-[in]-[a]-[u]-[u]- [u]	[nim]-[nom]-[nam]-[nem]- [nem]-[nom]-[num]-[num]- [num]	Pag. [ira]
7	*pitu	(Systematic and consistent)	(Systematic and consistent)	Pag. [ra]
8	*walu	(Systematic and consistent)	(Systematic and consistent)	Pag. [ra]
9	*siyam	(Systematic and consistent)	(Systematic and consistent)	Pag. [ira]
10	*ᮊapulu	[sam]-[sam]-[a]-[sam]-[sanga]- []-[na]-[na]-[na]	[pu]-[pu]-[pu]-[plu]-[pu]-[pu]- [pu]-[pu]-[pu]	[]-[lo]-[lu]-[ra]-[lo]- [lo]-[lo]-[lo]-[lo]
x100	*gatus	[da]-[]-[]-[la]-[ga]-[ga]-[ga]- [ga]-[ga]	[an]-[]-[]-[sus]-[sut]-[tus]-[tus]- [tos]-[tos]	
x1000	*libu	[li]-[ri]-[li]-[li]-[ri]-[li]-[li]-[li]- [yu]	[bo]-[bo]-[bu]-[bo]-[bu]-[bo]- [bo]-[bo]-[kot]	

Note: V is an indeterminate vowel sound, while ᮊ is an indeterminate consonant sound.

In fact, this partial reconstruction not only aids in tracing the historical development of these numerals but also emphasizes their role in contemporary Filipino identity (Mirasol, 2021; Lamorinas et al., 2024). As language evolves, it carries with it the weight of history and tradition, shaping how individuals perceive themselves within their cultural landscape. The exploration of Philippine numerals thus becomes a microcosm of larger sociolinguistic dynamics, illustrating how language serves as both a reflection and a shaper of cultural heritage (Zharkynbekova et al., 2025). Significantly, this investigation of partial reconstruction concepts in Philippine numerals provides valuable insights into the intersection of linguistics and culture. By reconstructing the etymology of these numerals, academics gains a clearer understanding of their significance in Filipino heritage formation, highlighting the enduring impact of language on identity and community cohesion.

Table 3. Reconstruction by Paz (2008).

Numeral	Proto-numeral	Numeral	Proto-numeral
one	*ʔəʔsáʔ	seven	*pituʔ
two	*daruwáʔ	eight	*waluʔ
three	*tatluʔ	nine	*siyám
four	*əmpat	ten	*puluʔ
five	*limáʔ	hundred	*gatús
six	*ʔaʔnəm	thousand	---

There are already efforts made to reconstruct the Proto-Philippine morphemes. One example is the work of Paz (2008) as captured by table 3, where she documented some of the sound systems of words in the different regional languages across the Philippine archipelago and arrived at a lexicon of reconstructed terminologies including some of the numerals in this study.

4. Discussion

Taking into account the findings, it was confirmed that the key areas explored such as etymology and phonemics are indispensable fields that illumined the intricate association between language, cultural inheritance, and distinctiveness. In Philippine context, etymology as the scientific study of the origins and historical development of words, offers insights into how language mirrors cultural influences and societal variations (Lucido et al., 2024). By tracing the roots of words, investigators were able to discover relations between language and particular cultural practices, thus strengthening group individuality (Canestrino et al. 2022). For instance, the etymological development of terms linked to traditional customs may determine how those traditions have been conserved or changed over time (Jackson et al., 2021).

Similarly, phonemics which focused on the sound systems of protolanguage numerals, further enhances the full understanding of ethnic identity. It scrutinizes how phonemic variations mirror regional tongues and social individualities (Naqsyah & Husna, 2024). The phonetic features of a language often function as indicators of ethnic belonging, influencing how people perceive themselves and others within their civic group. For example, dissimilarities in pronunciation may indicate regional associations or social class peculiarities, thus playing a central role in determining personal and shared characteristics (Planer, 2017). The relationship between etymology and phonemics highlights the significance of conserving linguistic variety as an essential aspect of cultural legacy. As languages advance, they bring with them the histories and individualities of their utterers (Kazanina et al., 2017). Comprehending phonemic variations will most likely augment appreciation for regional vernaculars and their historic origin (Duncan, 2018). This preservation is predominantly imperative in multilingual societies where prevailing languages may outshine native languages, redounding to a loss of cultural individuality (Aljasser & Ismail, 2025). Besides, the study of etymology and phonemics may make educational practices more cognizant thereby nurturing cultural mindfulness among language students. By incorporating discussions of phonetic differences and etymological experiences into curricula, teachers can help learners appreciate their dialectal heritage (Aljasser & Ismail, 2025; Wang, 2025). This method not only supplements language learning nonetheless also promoting a deeper consciousness of ethnic identity among schoolchildren from varied upbringings (Wang, 2025).

In addition, the association between etymology, phonemics, and cultural heritage has wider societal importance. As communities circumnavigate globalization and cultural interchange, comprehending the linguistic scopes of identity becomes progressively essential (Archibald, 2025; Del Mundo & Velasco, 2025; Wang, 2025). Language serves as a hallway for articulating cultural values and beliefs; thus, its conservation is indispensable for upholding cultural permanency in an ever-metamorphosing speech community. Likewise, studies that have to do with in eth-no-linguistics underscores how language mirrors folk ethnicities and shared memory within groups. Ethnolinguistic readings often divulge the influences between language use and cultural traditions such as storytelling, ceremonies, and art customs (Archibald, 2025; Duncan, 2018). By scrutinizing these associations through both etymological and phonemic perspectives, linguists, scientists, academics, and students can gain understanding into how traditional heritage has evolved across cohorts. The implication of phonemic consciousness is predominantly relevant in heritage language backgrounds.

Studies have revealed that understanding phonemic structures improves literacy development among utterers of heritage languages (Aljasser & Ismail, 2025; Archibald, 2025; Del Mundo & Velasco, 2025). These linking stresses the necessity for educational outlines that integrate both phonemic teaching and an appreciation for etymological roots to support students' linguistic individualities. Furthermore, as societies face trials related to linguistic shift and loss, understanding the phonological features of their languages becomes vital for regeneration efforts (Aljasser & Ismail, 2025; Mirasol, 2021). Phonemic scrutiny can offer insights into which sounds are most susceptible to change or loss within a society's linguistic range.

Essentially, as revealed by the data explored on protolanguage numerals, the investigation of the origins and historical progress of words, offers an introductory understanding of linguistic elements. It aids learners grasp as to how words develop over time, influenced by cultural interactions and historical occurrences (Archibald, 2025; Duncan, 2018; Del Mundo & Velasco, 2025; Wang, 2025). For example, exploring the Latin roots of English vocabulary may illumine connections between contemporary language and prehistoric civilizations. This understanding increases vocabulary acquisition and inspires learners to be engrossed with language analytically, knowing its dynamic landscape. Assimilating etymological studies into the prospectus may also stimulate learners by linking language to their cultural inheritance. For instance, scrutinizing the origins of specific language related to local civilizations can nurture a sense of individuality and belonging among schoolchildren. This connection not only augments vocabulary attainment but also expands students' understanding of numerous subjects, comprising science, history, and literature occurrences (Archibald, 2025; Duncan, 2018; Mirasol, 2021, Wang, 2025). Phonemic cognizance is fundamental for reading progress; studies reliably show that a strong understanding of phonemic constructions relates with successful reading results. By training students to distinguish and manipulate phonemes, teachers can meaningfully improve their deciphering abilities, leading to an improved reading articulation and grasp.

Moreover, phonemic teaching is chiefly significant for early literacy courses. Studies have indicated that teenagers who receive clear phonemic teaching exhibit greater success in learning to read as likened to those who do not (Aljasser & Ismail, 2025; Mirasol, 2021). This pre-supposes the need of incorporating phonemic cognizance into foundational literacy programs. The instruction of etymology and phonemics also has mathematical suggestions. Understanding arrangements in language can be equated to distinguishing mathematical orders or arrangements. For example, scrutinizing word formation processes – such as prefixes, suffixes, and root words – necessitates learners to categorize associations between dissimilar components of words (Archibald, 2025; Del Mundo & Velasco, 2025; Mirasol, 2021, Wang, 2025). This methodical approach reflects mathematical problem-solving systems where patterns must be identified and applied. Additionally, phonemic inquiry often includes arithmetical methods to evaluate sound distributions across languages or dialects. Such numerical evaluation helps learners draw links between linguistic structures and wider sociolinguistic inclinations (Planer, 2017). By leveraging mathematical concepts in linguistic lessons, teachers can nurture inter-disciplinary knowledge that boosts rational skills.

Most significantly, the educational significance of etymology and phonemics spans beyond language mechanisms; it covers cultural inheritance and individuality. Language is an image of ethnic identity, and understanding its origin nurtures a sense of fitting among learners from diverse upbringings (Naqsyah & Husna, 2024). By scrutinizing the etymology of words definite to numerous beliefs, instructors can help learners appreciate their own dialectal in-heritance as well as that of others. For instance, reviewing aboriginal languages and their phonemic constructions may enable learners to identify the significance of linguistic multiplicity in their societies (Duncan, 2018). This cognizance upholds respect for diverse cultures and inspires learners to engage in discussions about identity and tradition.

5. Conclusions

In an attempt to deduce the parent counting terms of Filipino protolanguage from the sound systems of the nine identified descendants, this study proposes the following terms as the most likely ancestral roots of Filipino numeral sounds from one through ten, hundred and thousand: *əsa, *dəwa, *təlo, *əpat, *lima, *ənəm, *pitu, *walu, *siyam, *napulu, *gatus, and *libu, respectively, with the schwa symbol “ə” pronounced as a colorless vowel. Four of these parents have not evolved tremendously and have been preserved until today (*lima, pito, walo, and siyam*), while eight of them have undergone consonant and vowel changes.

Etymology and phonemics are profoundly relevant to understanding cultural heritage and identity, as they reveal the intricate connections between language, history, and social dynamics. Etymology provides insights into the origins and transformations of words, highlighting how cultural influences shape linguistic practices over time. Phonemics, on the other hand, examines the sound systems of languages, illustrating how phonemic variations reflect regional identities and social affiliations. Together, these fields underscore the importance of preserving linguistic diversity as a vital aspect of cultural identity.

To enhance appreciation for linguistic heritage, educational institutions should integrate etymological and phonemic studies into their curricula. This approach can foster greater awareness of cultural identity among students and promote the preservation of endangered languages. Additionally, policymakers should support initiatives that document and revitalize indigenous languages to ensure that cultural narratives continue to thrive in contemporary society. By prioritizing these aspects, we can contribute to a richer understanding of our shared human experience. In conclusion, the relevance of etymology and phonemics to cultural heritage and identity cannot be overstated. These fields provide valuable frameworks for understanding how language shapes our perceptions of self and community. As we continue to explore the intricate connections between language and culture, it becomes increasingly clear that preserving linguistic diversity is essential for fostering a rich tapestry of human experience. In consideration of the findings, this study suggests the following:

Firstly, develop interdisciplinary lessons that connect etymology with subjects like history or science. For instance, exploring the origins of scientific terminology can deepen students' understanding of scientific concepts.

Secondly, utilize games that focus on phonemic manipulation to make learning engaging. Activities such as rhyming games or phoneme segmentation exercises can enhance phonemic awareness while fostering a fun learning environment.

Thirdly, encourage students to research the etymology of words from their cultural backgrounds or explore the phonemic characteristics of different languages. This project-based approach allows for personal connections to be made with language learning.

Fourthly, provide teachers with training on the importance of etymology and phonemics in literacy education. Equipping educators with this knowledge will enable them to deliver more effective instruction.

Fifthly, develop assessment tools that evaluate both phonemic awareness and etymological knowledge. These assessments can help identify areas where students may need additional support.

To the extent that this study delved on specific regional and national dialects on proto-language numerals sociolinguistically and pedagogically, this scope has failed to illuminate its foreign linguistic roots such as Malayan as Indonesian historically known as Filipino predecessors. These countries are just geographically adjacent to Philippines and could have linguistic influence and resemblance to Philippine protolanguage numerals. Nonetheless, collectively denoting, research on etymology and phonemics in relation to cultural heritage and identity faces several limitations. One significant challenge is the scarcity of comprehensive historical data, particularly for underdocumented languages, which hampers the ability to trace phonetic and etymological changes accurately. Additionally, the complexity of phonemic inventories across languages complicates comparative analyses, as variations in transcription and representation can lead to inconsistencies. Furthermore, the influence of socio-political factors on language use and preservation often remains underexplored, limiting our understanding of how these dynamics affect cultural identity.

To address these limitations, future research should prioritize interdisciplinary approaches that combine linguistic analysis with sociocultural studies. Collaborative efforts between linguists, anthropologists, and educators can enhance data collection and interpretation, ensuring a more holistic understanding of language's role in cultural heritage. Additionally, increased investment in documenting endangered languages will be crucial for preserving linguistic diversity and its associated cultural identities.

Author Contributions: Author 1: conceptualization, gathering of data, and formatting. Author 2: editing, final review, analysis. Author 3: publication.

Funding: Authors did not apply of any funds or grants in the completion of this article.

Institutional Review Board Statement: This article has undergone in-house review and had passed the graduate school's research and ethics committee on October 2024.

Acknowledgments: The authors are thankful to the participants and ethics review board of the university.

Conflicts of Interest: No potential conflicts of interest are found in this study in any form possible

References

- Aljasser, F., & Ismail, S. (2025). Perception Deception: Exploring the Gap between Self-Perception and Phonemic Perception among Arabic-speaking EFL learners. *World Journal of English Language*, 15(4), p341. <https://doi.org/10.5430/wjel.v15n4p341>
- Alshehri, S. (2023). The Relationship between Language and Identity. *International Journal of Linguistics, Literature and Translation*, 6(7), 156-161. <https://doi.org/10.32996/ijllt.2023.6.7.14>
- Archibald, J. (2025). Such sweet thunder. *Second Language Research*, 41(3), 587-604. <https://doi.org/10.1177/02676583251339901>
- Canestrino, R., Magliocca, P., & Li, Y. (2022). The Impact of Language Diversity on Knowledge Sharing Within International University Research Teams: Evidence From TED Project. *Frontiers in psychology*, 13, 879154. <https://doi.org/10.3389/fpsyg.2022.879154>
- Celata, C., & Nagy, N. (2022). Sociophonetic Variation and Change in Heritage Languages: Lexical Effects in Heritage Italian Aspiration of Voiceless Stops. *Language and Speech*, 67(2), 438-462. <https://doi.org/10.1177/00238309221126483>
- Conant, C. E. (1913). The Pepet Law in Philippine Languages. [Doctorate Dissertation], Chicago University.
- Del Mundo, D. B., & Velasco, C. Q. (2025). Phonological Processes and Oral Language Proficiency as Moderated by Ethnolinguistic Group. *International Journal of Social Science Humanity & Management Research*, 4(6), 1130-1139. <https://doi.org/10.58806/ijsshr.2025.v4i6n14>
- Duncan, L. G. (2018). Language and Reading: The Role of Morpheme and Phoneme Awareness. *Current Developmental Disorders Reports*, 5, 226-234. <https://doi.org/10.1007/s40474-018-0153-2>
- Ethnologue. (2024). Philippines: Languages. <https://www.ethnologue.com/country/PH/#languages>
- Fatima, S., & Nadeem, M. U. (2025). Family language policy and heritage language transmission in Pakistan – the intersection of family dynamics, ethnic identity and cultural practices on language proficiency and maintenance. *Frontiers in psychology*, 16, 1560755. <https://doi.org/10.3389/fpsyg.2025.1560755>
- Fraser, H. (2022). A Framework for Deciding How to Create and Evaluate Transcripts for Forensic and Other Purposes. *Frontiers in Communication*, 7, 898410. <https://doi.org/10.3389/fcomm.2022.898410>
- Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- Harrington, L. A. (2023). Incorporating automatic speech recognition methods into the transcription of police-suspect interviews: factors affecting automatic performance. *Frontiers in Communication*, 8, 116523. <https://doi.org/10.3389/fcomm.2023.116523>
- Hartman, A., & Squires, V. (2024). Bridging Perspectives: Utilizing Interpretative Phenomenological Analysis (IPA) to Inform and Enhance Social Interventions. *International Journal of Qualitative Methods*, 23. <https://doi.org/10.1177/16094069241306284>
- Hartmann, S., & Pleyer, M. (2021). Constructing a protolanguage: reconstructing prehistoric languages in a usage-based construction grammar framework. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376, 20200200. <https://doi.org/10.1098/rstb.2020.0200>
- Hepburn, A., & Bolden, G. (2017). *Transcribing for social research*. SAGE Publications Ltd, <https://doi.org/10.4135/9781473920460>
- Jackson, J. C., Watts, J., List, J.-M., Puryear, C., Drabble, R., & Lindquist, K. A. (2021). From Text to Thought: How Analyzing Language Can Advance Psychological Science. *Perspectives on Psychological Science*, 17(3), 805-826. <https://doi.org/10.1177/17456916211004899>
- Kazanina, N., Bowers, J. S., & Idsardi, W. (2017). Phonemes: Lexical access and beyond. *Psychonomic Bulletin & Review*, 25, 560-585. <https://doi.org/10.3758/s13423-017-1362-0>
- Konig, E. (2012). Contrastive linguistics and language comparison. *Languages in Contrast*, 12(1), 3-26. <https://doi.org/10.1075/lic.12.1.02kon>
- Lamorinas, D. D., Jane, A., & Jane, M. (2024). Preserving initiatives for the Filipino language: Analysis of language protection through personal efforts among Filipino students. *Environment and Social Psychology*, 9(12). <https://doi.org/10.59429/esp.v9i12.3230>
- Lucido, F., Jimenez, D., & Tang, S. (2024). Affirming culture and cultural identity in the bilingual/ESL classrooms. *Frontiers in Education*, 9, 1338671. <https://doi.org/10.3389/educ.2024.1338671>
- Mirasol, R. G. (2021). Sociolinguistic Survey of Ayta Mag-Antsi Language. *International Journal of Education and Social Science Research*, 4(2), 297-312. <https://doi.org/10.37500/ijessr.2021.4224>
- Naqsyah, I. D., & Husna, F. (2024). An Overview of Phonetics, Phonemics, and Graphemes. *Jurnal Recons*, 1(2), 58-67. <https://doi.org/10.59548/rc.v1i2.289>
- Niolaki, G., Vousden, J., Terzopoulos, A., Shepherd, D.-L., Debney, L., & Masterson, J. (2022). Spelling predictors; investigating the role of phonological ability and rapid naming in a large cross-sectional British study. *Learning and Instruction*, 80, 101635. <https://doi.org/10.1016/j.learninstruc.2022.101635>
- Paz, C. J. (2008). *Ginhawa, kapalaran, dalambati: Essays on well-being, opportunity/destiny, and anguish*. University of the Philippines Press.
- Planer, R. J. (2017). Protolanguage Might Have Evolved Before Ostensive Communication. *Biological Theory*, 12, 72-84. <https://doi.org/10.1007/s13752-017-0262-x>
- Sivasubramaniam, S., Dlabolová, D. H., Kralikova, V., & Khan, Z. R. (2021). Assisting You to Advance with Ethics in Research: an Introduction to Ethical Governance and Application Procedures. *International Journal for Educational Integrity*, 17, 14. <https://doi.org/10.1007/s40979-021-00078-6>
- Stošić, S., Marković, L., & Smith, A. (2016). Early Spanish Reign over the Philippines: Social and Spiritual Remodeling. *The Journal of International Civilization Studies*, 1(2), 187-202. <https://dergipark.org.tr/tr/download/article-file/1354555>
- Tagg, C., & Manford, P. (2025). Team transcribing: an approach to analysing and interpreting online acquaintance interviews as collaborative achievements. *Research Methods in Applied Linguistics*, 4(3), 100250. <https://doi.org/10.1016/j.rmal.2025.100250>

- Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, 8, 37-43. <https://doi.org/10.1016/j.copsyc.2015.09.003>
- Vargas-Bianchi, L. (2025). Pattern matching analysis: Overview of its rationale and application in qualitative research. *Methodological Innovations*, 18(1), 66-75. <https://doi.org/10.1177/20597991251325451>
- Wang, Y. (2025). A Study on the Efficacy of ChatGPT-4 in Enhancing Students' English Communication Skills. *Sage Open*, 15(1). <https://doi.org/10.1177/21582440241310644>
- Wood, C. V., Salusky, I., Jones, R. F., Remich, R., Caliendo, A. E., & McGee, R. (2024). Using longitudinal qualitative research to understand the experiences of minoritized people. *Methods in Psychology*, 10, 100130. <https://doi.org/10.1016/j.metip.2023.100130>
- Zaide, S. M. (1999). *The Philippines: A unique nation*. Quezon City: All Nations Publishing Co., Inc.
- Zharkynbekova, S., Ayupova, G., Galiyeva, B., Shakhputova, Z., & Zabrodsckaja, A. (2025). Ethno-Linguistic Identity of Kazakhstani Student Youth in Modern Multinational Context of Kazakhstan (Sociolinguistic Analysis of Empirical Research). *Languages*, 10(2), 33. <https://doi.org/10.3390/languages10020033>