LiNGUA Jurnal Ilmu Bahasa dan Sastra

LiNGUA Vol. 20, No. 1, June 2025 • ISSN 1693-4725 • e-ISSN 2442-3823

The Taiwanesians Emigrated from The Southeast Asian Islands: A Phonic and Historical Approach

Chen Hurng Yu

tanhongru@gmail.com
Tamkang University, New Taipei City, Taiwan

First Received: 03 February 2025; Final Proof Received: 08 June 2025

Abstract

The origins and migration routes of Taiwanesian populations have long been a topic of scholarly debate, with various theories suggesting connections to regions in Southeast Asia, mainland China, and beyond. Clarifying these connections is essential for understanding the ethnic and linguistic development of Taiwan's indigenous peoples and their ties to other Austronesian-speaking communities. This study seeks to explore the phonetic and historical relationships between Taiwanesians and Southeast Asian ethnic groups, especially those from the Philippines, Indonesia, Malaysia, New Guinea, and Vietnam, addressing gaps in earlier migration theories that often overlooked oceanographic conditions and linguistic evidence. To accomplish this, the research applies a comparative phonetic analysis of basic vocabulary across 67 ethnic languages, using data from four historically documented Taiwanesian languages alongside sixteen contemporary ones. The analysis draws on digitized dictionaries and historical sources and incorporates insights from regional linguists and anthropologists to strengthen the comparative approach. The results show notable phonetic similarities between Taiwanesian languages and several languages spoken in the Philippines—such as Hiligaynon, Ivatan, Ilocano, and Tagalog—as well as connections with Indonesian languages like Bugis, Sangir, Toba Batak, and Minangkabau. In contrast, the phonetic overlap with languages from New Guinea is limited, indicating weaker linguistic ties. Historical records, including evidence of ocean currents and oral histories, support the hypothesis of a south-to-north migration pathway that enabled the movement of populations from the Indonesian archipelago and central Philippine Islands toward Taiwan. This research questions previous assumptions that Taiwanesians once spoke early forms of Bahasa Malay or Bahasa Indonesia, instead revealing complex linguistic diversity and migration patterns within the broader Austronesian world. By combining phonetic comparisons with historical context, the study offers new perspectives on the ethnic and linguistic affiliations of Taiwanesian populations. These findings provide a foundation for further interdisciplinary research on Austronesian migration and highlight the value of phonetic evidence as a tool for uncovering ancient population movements and fostering greater cross-regional collaboration in historical linguistics and anthropology.

Keywords: Taiwanesians; Indonesia; Philippine; ethnic language; Southeast Asia

INTRODUCTION

There are various viewpoints on the historical relationship between Taiwanesian populations (The Taiwanesian refers to Taiwanese aborigines, and the Taiwanese refers to the Han people in Taiwan) and the ethnic groups of Southeast Asia. One theory posits that populations migrated from the Southern Islands to Taiwan, an idea endorsed by scholars like John Dodd (1882) and Japanese researcher Inō Kanori (1993). On the other hand, an opposing perspective, supported by Peter Bellwood (1991; 1999), aligns with archaeological and linguistic evidence suggesting that

Taiwanesian populations crossed the Bashee Strait to settle in Luzon in the Philippines around 3,500 B.C. Linguists such as Paul K. Benedict (1975) and Robert Blust (1985) hold views that align with Bellwood's argument. However, it is crucial to acknowledge that these hypotheses are theoretical at this stage.

Regardless of the theories proposed by linguists, an essential factor has often been overlooked to validate their arguments: ocean currents, which heavily influence population movement. In ancient times, when navigation tools were rudimentary, crossing the ocean was far from an easy feat. One of the most significant currents is the Kuroshio in the western Pacific. Seasonal monsoons affect ocean currents, altering their directions in winter and summer. The Kuroshio originates from northern waters near the New Guinea Islands and flows northwest along the eastern coasts of the Philippines, Taiwan, and the Ryukyu Islands. This current consistently moves from south to north, irrespective of the season. Similarly, the current in the South China Sea along the Philippines' western coast also flows from south to north. By contrast, the ocean currents off the eastern coastline of mainland China move from north to south during winter and reverse in summer (Britainnica, no date). A particularly treacherous underwater current in the Taiwan Strait, known historically as the "black water ditch", posed great challenges to cross the area. As a result, in ancient times, navigation by primitive tools made it easier for populations to move from New Guinea, Borneo, or the Philippines towards Taiwan, rather than the other way around, moving from Taiwan to the Philippines or the Indonesian archipelago.

This research approach begins with a comparative analysis of the fundamental vocabularies of various ethnic languages to explore potential connections between early Taiwanesians and Southeast Asian ethnic groups. The period before Dutch and Spanish colonization in Taiwan serves as an important reference point since it is likely that the native Taiwanesian languages remained relatively untouched by European influences. The Dutch left behind four language dictionaries when they ruled Taiwan, which became the subject of the current research.

Secondly, since the northern islands of the Indonesian archipelago, such as Borneo, the Philippines, New Guinea, and Taiwan, did not establish a unified dynasty (empire-like state) or adopt a common national language in ancient times, the tribes inhabiting these islands remained scattered and linguistically distinct. Each tribe's language retained its original features. If linguistic similarities exist between the languages of ethnic groups spread across these islands, it indicates the possibility of ethnic connections among them.

The first researcher to study the comparison between Taiwanesian languages and Southeast Asian languages was Albrecht Wirth. In 1897, he examined five ethnic groups from Taiwan, Favorlang, Sideis, Mt. Morrison, Tchihuan, and Tayal, and conducted a comparison with the Tagal (modern-day Tagalog) ethnic group from the Philippines using 120 words. He concluded that these languages differed significantly in their phonetic characteristics. Moreover, the languages spoken in central Taiwan were found to be highly diverse and mixed, comprising elements from four to five different ethnic groups. Some of these linguistic elements have origins in the surrounding ethnic groups, while others are contributions from foreign invaders, leading to a great variety of speech forms. As a result, it is challenging to classify these languages into defined categories, such as Malay, Polynesian, or Tagal, among others. Furthermore, linguistic traces of Melanesian speech have been noted among the frizzle-haired inhabitants of Tobel Tobago (Wirth, 1897). However, it remains regrettable that Wirth did not document the specific methods he used for his phonetic comparisons.

In traditional tribal societies, individuals of the same ethnic group typically live together. These groups tend to exhibit a high degree of racial and linguistic uniformity. Such homogeneity can endure over time when members of the tribe intermarry within their community. However, intermarriage between separate tribes introduces influences from other races and languages, reflecting the background of the spouse. If these marriages occur on a small scale, the minority languages and races brought in are likely absorbed into the dominant tribe. Despite this, certain words from the non-native language might be adopted by the dominant tribe, causing the language to evolve or blend. Particularly for tribes without a written form of their language, the spoken vocabulary may gradually change with each new generation, leading to linguistic shifts and, eventually, the development of an entirely new language due to differences in expressions and usage.

Secondly, when tribes live in close proximity to one another, interactions such as intermarriage and trade can facilitate racial and linguistic blending. This often arises due to influences like new introductions, mispronunciations, sound change, borrowing, analogical change (Campbell, 1999), and intentional distinctions. Over time, this process leads to the gradual development of new languages or phonetic systems.

In island environments, an isolated island that lacks interaction with tribes from surrounding islands can develop a uniquely pure race and language, preserving its linguistic traits for an extended period with minimal alteration. On the other hand, if the island's tribes maintain frequent contact with those on other islands, the language's purity tends to diminish.

Suppose islands located far apart show similarities in language and phonetics, despite lacking political connections or a unified national language, it is reasonable to conclude that there is a racial and linguistic link between them. While, as previously mentioned, languages may adopt borrow, imitate, or incorporate phrases through trade and communication, making them appear more alike than they are in terms of ethnic significance, this may hide their ethnic difference. For island populations separated by vast seas, but have strikingly similar or nearly identical phonetics, it would be a theoretically blind spot to say that there is no racial correlation between them.

Building on the theoretical explanation above, this discussion examines the phonetic similarities between the languages of the Taiwanesians and Southeast Asian ethnic groups. Despite living far apart, these groups were connected in ancient times through travel by small boats between islands, spanning from the Indonesian archipelago to the Philippines and eventually to Taiwan. Linguistic analysis reveals a slightly higher degree of phonetic similarity among their languages, suggesting a possible ancestral or racial connection among these peoples. Over time, their migration and intermarriage with other tribes gave rise to a more intricate blend of ethnicities and linguistic traits. By comparing phonetics, this article aims to develop an initial framework linking racial background and language phonics.

Throughout history, word comparisons have been employed as evidence of family relationships (Campbell, 1999). Campbell (1999) also emphasized that most scholars insist that basic vocabulary should be part of the supporting evidence presented in favor of any distant genetic relationship. Basic vocabulary is usually not defined rigorously but is understood generally to include body parts, close kinship, frequently encountered aspects of the natural world (mountain, river, cloud, and the like), and low numbers. Basic vocabulary is in general resistant to borrowing, and so, similarities found in comparisons involving basic vocabulary items are unlikely to be due to diffusion and hence stand a better chance of being evidence of distant genetic relationships, of being inherited from a common ancestor, than other kinds of vocabulary (Campbell, 1999).

This study does not examine the origins or development of language families, nor does it investigate the identity of the earliest language creators. It avoids delving into grammatical or syntactical comparisons and does not address the causes behind language evolution. Instead, its primary objective is to compare the phonetics of basic vocabulary in dictionaries, as highlighted by Lyle Campbell (1999), to assess the depth of their relationships through phonetic similarity. By doing so, it indirectly illustrates their group associations. The article also argues that such phonetic analyses and comparisons can effectively uncover unique connections among groups across numerous islands spread over a vast region.

METHOD

Four historic Taiwanesian languages are selected in this study (including Tilaossen-Sinkan, Formosan, Favorlang, and Lower Tamsui (Fèng Shan)), along with sixteen contemporary Taiwanesian languages, such as Amis, Paiwan, Puyuma, Bunun, Thao, Rukai, Yami, Saisiyat, Atayal, Tai Luge, Seediq, Tsou, Kavalan, Sakilaya, La Aluwa, and Kanakanavu. Additionally, the study includes eleven Philippine languages, such as Hiligaynon (Ilonggo), Tagalog, Tuwali Ifugao, Igorot, Cebuano, Pampango, Ilokano, Bikol, Waray, and Ivatan (Ibatan). For the Indonesian region, eighteen languages are covered, including Sangir, Bugis, Toba Batak, Minangkabau, Sundanese, Bahasa Indonesia, Ternatanese, Acehnese, Javanese, Madurese, Balinese, Banjar, Makasar, Manggarai, Lamaholot, Alor, Bima, and Tetun. The analysis further incorporates seven Malaysian languages, such as Bahasa Malay, Sea Dayak, Kadazan Dusun, Bisaya, Timugon Murut, Bau Bidayuh, and Central Melanau.

Moreover, seven New Guinea languages have also been included, namely Gapa Paiwa, Hiri Motu, Koiari, Tok Pisin, Sentani, Kwoma, and Bariai. Other selected languages are the Gia Rai language (from Vietnam), Kinh, Hmong (Miao), as well as Central Okinawan, summing up to a total of 67 ethnic languages for comparative analysis. The chosen languages predominantly represent larger ethnic groups and are documented in English-language dictionaries. However, languages spoken by smaller ethnic populations are excluded due to the unavailability of English-language dictionary references.

This research is conducted in two phases. The first phase involves analyzing four ancient Taiwanesian languages documented in old Taiwanese records. These include the Tilaossen-Sinkan indigenous dialect (also known as the Sinkan language), the Formosan language (likewise referred to as the Sinkan language), the Favorlang language (or Huweilong language), spoken in Favorlang village, which, in 1636, was situated 6-7 miles north of Wankan and home to approximately 344 households, or around 1,000 people (Blussé, 2010), and the Lower Tamsui (Fèngshan) language. These languages will be compared with the languages of regions such as the Philippines, Indonesia, Malaysia, New Guinea, and Vietnam. The second phase involves examining the sixteen Taiwanesian languages in comparison to the aforementioned Austronesian languages to identify potential correlations between them.

In phase 1, the phonetic similarities among four ancient Taiwanesian languages and the ethnic languages of Indonesia, Malaysia, the Philippines, New Guinea, and Vietnam are analyzed. The researcher examined four Taiwanesian languages that were documented and spoken over three centuries ago. As these languages were recorded alongside Chinese, Dutch, or English translations, they provide valuable materials for this analysis.

Currently, the Philippines is home to 187 ethnic languages. This study focuses on 11 of the most widely used Filipino languages, namely Hiligaynon, Tagalog, Ifugao, Igorot, Cebuano, Pampanga, Ilocano, Bikol, Waray, Ivatan, and Ibatan. Tagalog, spoken by the residents of Manila and its surrounding areas, was designated as the national language of the Philippines in 1938. Ifugao is primarily spoken by an ethnic group based in the mountainous Ifugao province of Central Luzon. The Igorot language is used by the Indigenous communities residing in the Cordillera Mountains of Northern Luzon, also known collectively as the Cordillera people (Cordilleran Culture, no date). Both the Ifugao and Igorot peoples traditionally exhibit similar cultural traits to early Taiwanesians, such as facial tattoos and headhunting. Cebuano, on the other hand, is the primary language spoken in Cebu Island and nearby areas in the Visayan region of central Philippines. Meanwhile, Ilocano is predominantly spoken by communities in Northern Luzon.

The Hiligaynon language is primarily spoken in the western section of the Visayas region and the Soccsksargen area, located in Cotabato in the southern-central part of Mindanao. It is widely regarded as the second major language in the Visayas region. Approximately 9.3 million people speak Hiligaynon, with its reach spanning the provinces of Panay, Negros, and Guimaras (Ager, 2024). On the other hand, the Bikol language is used by inhabitants of the Bikol Peninsula situated at the southeastern edge of Luzon Island. The Waray language is predominantly spoken in the eastern Visayas, encompassing parts of northern Samar, southern and western Leyte, as well as Biliran and Pintuyan Island (Taborete, no date). Similarly, the Pampanga language is spoken in the provinces of Pampanga and southern Tarlac in central Luzon, as well as in localized areas in northeastern Bataan, western Bulacan, southeastern Zambales, and southwestern Nueva Ecija. The Ivatan language is indigenous to the population of the Batanes Islands, while the residents of the Babuyan Islands use the Ibatan language as their communication.

Indonesia is an archipelagic nation rich in ethnic diversity, home to 1,340 linguistic groups (Statistics Indonesia, no date). For the purposes of this study, specific languages have been chosen, including the national language, Bahasa Indonesia, Sundanese from West Java, Minangkabau from the Padang region in Sumatra, Toba Batak and Acehnese from northern Sumatra, Bugis from southern Sulawesi, and Ternatenes and Sangir from the northern Moluccas. In addition, Javanese, Balinese, Bima, and Tetun from East Java will also be examined.

Taking into account that Bahasa Malay is widely spoken by the Malays, the current research also explores the shared cultural practice of facial tattooing between the Sea-Dayak people of Sarawak and the Taiwanesians. Additionally, it includes an investigation into the Kadazan Dusun, Bisaya, Timugon Murut, Bau Bidayuh, and Central Melanau languages.

Given that the northern sea of the New Guinea archipelago serves as the origin of the Kuroshio in the western Pacific Ocean, the research also encompasses the languages of Hiri Motu, Tok Pisin, Koiari, Santani, Kwoma, Bariai, and Gapa Paiwa, which are widely spoken across the New Guinea Islands.

To explore the extent of Austronesian languages, the researcher chose Central Okinawan, Hmong, Kinh, and Gia Rai for this analysis. In phase 2, the phonics of the 16 Taiwanesian languages were compared with those of various ethnic languages across Southeast Asia and New Guinea. For this analysis, the online Dictionary of Aboriginal Languages in Taiwan served as a reference tool, providing Romanized spellings of words in the 16 Taiwanesian languages.

The book The Ethnic Minorities in Vietnam, authored by Vietnamese scholars Dang Nghiem Van, Chu Thai Son, and Luu Hung, highlights that the Gia Rai people historically inhabited the central highlands of southern Vietnam. The Gia Rai, alongside ethnic groups such as the Ede, Cham, Raglai, and Chu Ru, are categorized under the Malay-Polynesian language family (Dang, 1984). Philippine anthropologist E. Arsenio Manuel theorized a connection between the ethnic groups of the Red River Basin and those in the Philippines. He also suggested that the Gia Rai language of southern Vietnam shares similarities with Tagalog. According to his findings, these groups migrated directly to Luzon, contributing to the formation of ethnic communities near Manila (Manuel, 1994).

To validate Manuel's hypothesis (1994), the researcher chose the Gia Rai language to conduct an in-depth comparison between it and the languages of Taiwan and Southeast Asia, while also demonstrating a south-to-north migration pattern among these ethnic groups.

ANALYSIS

Phase 1

A comparison was made among four old Taiwanesian languages, revealing a low proportion of phonetically similar words. Out of 113 words, only five share a similar pronunciation across all four languages. Excluding the Favorlang language, the other three languages exhibit 10 phoneticallysimilar words, while 25 similar words are found between the Tilaossen-Sinkan language and the Formosan language. Although these old Taiwanesian languages belong to the same linguistic family, their similarities are insufficient to allow mutual intelligibility.

Secondly, a comparison is made between these four old Taiwanesian languages and 11 Filipino languages. The Tilaossen-Sinkan language shares 12 phonetically similar words with Hiligaynon, 11 words with both Ivatan and Ibatan, and 10 words each with Tagalog, Igorot, and Ilocano. The Formosan and Ilocano languages share 8 phonetically similar words, while they also have 7 similar words with the Hiligaynon, Igorot, and Tagalog respectively. The Tilaossen-Sinkan language and the Formosan languages exhibit a strong linguistic connection with the Filipino languages. Numerous words, like "moon," "rain," "river," "fire," "father," "eye," "tooth," "dog," "pig," "three," "five," "six," "seven," "female," and "hair," share strikingly similar pronunciation. The Favorlang and Lower Tamsui Fengshan languages exhibit a greater inclination toward the Hiligaynon and Cebuano languages, although this tendency is less pronounced compared to the Tilaossen-Sinkan language and the Formosan language.

Table 1. Comparison of the Phonic Similarity of the Four Old Taiwanesian Languages with Filipino Languages

		Filipino Languages									
Four Old Taiwanese Aboriginal Languages		Tagalog	Ifugao	Igorot	Cebuano	Pampanga	Ilocano	Bikol	Waray	Hiligaynon	Ibatan/ Ivatan
	Tilaossen-sinkan	10	8	10	9	2	10	9	7	12	11
	Formosan	7	3	7	6	4	8	4	4	7	5
	Favorlang	3	2	3	3	0	3	0	1	3	2
	Lower Tamsui Fèng-shan	4	4	5	5	2	5	5	4	6	4

Note: The numbers in the table represent the number of similar words.

Table 2. Comparison of the Phonic Similarity of the Four Old Taiwanesian Languages with Malaysia Languages

		Malaysia Languages						
		Bahasa Malay	Sea Dayak	Kadazan Dusun	Bisaya	Timugon Murut	Bau Bidayuh	Central Melanau
Four Old Taiwanesian languages	Tilaossen-sinkan	6	6	5	8	5	2	4
	Formosan	4	4	5	8	5	3	3
	Favorlang	0	0	2	5	2	2	3
	Lower Tamsui Fèng-shan	3	4	5	7	5	2	3

Table 3. Comparison of the Phonic Similarity of the Four Old Taiwanesian Languages with Indonesian Languages

		Indonesian Languages																	
		Bahasa Indonesia	Sundanese	Minangkabau	Toba Batak	Ternatese	Bugis	Sangir	Acehnese	Javanese	Madurese	Balinese	Banjar	Makasar	Mangarai	Lamaholot	Alor	Bima	Tetun
Four Old Taiwanese Aboriginal Languages	Tilaossen-Sinkan	6	5	5	6	1	8	7	4	6	4	6	4	5	7	6	6	0	7
	Formosan	4	1	3	4	1	6	6	2	5	3	7	3	5	7	7	8	1	10
	Favorlang	0	0	0	0	0	1	1	0	1	0	2	1	2	1	1	2	3	2
	Lower Tamsui Fèng-shan	4	1	3	3	1	5	5	2	3	1	4	3	4	5	4	4	0	4

Note: The numbers in the table represent the number of similar words.

Thirdly, to examine the similarities between the four old Taiwanesian languages and the eight languages of Indonesia. Among these comparisons, the Tilaossen-Sinkan language shares 8 similar words with the Bugis language, 7 similar words with the Sangir language, and 6 similar words each with the Sea Dayak language, Toba Batak language, Bahasa Indonesia, and Bahasa Malay. The Formosan language shares six similar words with both the Bugis and Sangir languages. In contrast, the Favorlang language exhibits minimal connection to the Indonesian language. The Bugis and Sangir languages show the strongest resemblance to the Lower Tamsui Fengshan language, with each sharing five similar words. It is interesting to point out that the words for "month," "father," "eye," "dog," "three," "four," "five," "six," and "hair" share noticeably similar phonetics across both old Taiwanesian and Indonesian languages. The Bugis, Sangir, Sea Dayak, and Toba Batak languages, spoken in Indonesia and Malaysia, share a strong linguistic connection with the Tilaossen-Sinkan and Formosan languages found in Taiwan. Renowned as expert seafarers, the Bugis people excelled in maritime trade and piracy across Indonesian and Malay waters in ancient times. Their possible migration to the Visayan areas or southern Taiwan may have been driven by established trade pathways or unexpected sea storms.

The four old Taiwanesian languages bear a significant resemblance to the Hiligaynon language of the Philippines and the Bugis language of Indonesia. Regarding the relationship between Hiligaynon and Bugis, a comparison shows that both share 19 phonetically similar words. According to The Formosan Encounter, edited by Leonard Blussé, Natalie Everts, and E. Frech, it is noted that the Soulangh (modern-day Jiali) people used a blend of strange and mixed words brought from Malaya and nearby regions. Examples include terms such as “babi” (pig), “takut” (scared), “busuk” (rotten), “makan" (eat), "ikan" (fish). This speech bears similarities to the language spoken on the

Sangir Islands (Blussé, 2010), a group of islands located off Northern Sulawesi's coast. Specifically, this language is used by inhabitants of Big Sangir Island in North Sulawesi and Davao in the southern part of Mindanao, Philippines. In 2010, around 110,000 people in Indonesia, as well as 55,000 individuals in the Philippines, spoke the Sangir language (Omniglot, no date). Meanwhile, the Bugis language is primarily used in the southern region of Sulawesi.

The researcher compared the Sangir language with the Bugis language and identified 32 phonetically similar words. This suggests the possibility that the Bugis people might have migrated northward to northern Sulawesi, passing through the Sangir Islands before reaching Davao, Mindanao. Alternatively, they may have traveled westward to Cotabato, an area associated with the Hiligaynon language group. Another scenario is that they went directly north from Sulawesi to Cotabato without passing through the Sangir Islands, and from there migrated to Panay Island. They journeyed from Panay Island via the northern Luzon coast, passing through the Babuyan and Batanes Islands, continuing to Siaoliuqiu Island, and finally reaching the Tainan region in Taiwan.

In detail, Tilaossen-Sinkan shares 8 similar words with Bugis, 7 with each of Manggarai, Sanghir, and Tetun, and 6 words with Toba Batak, Bahasa Indonesia, Javanese, Balinese, Lamaholot, and Alor respectively. Formosan shares 10 similar words with Tetun, 8 with Alor, and 7 with Balinese, Lamaholot, and Manggarai. Additionally, it has 6 similar words with Bugis and Sangir, respectively. On the contrary, Favorlang shows minimal connection with Indonesian languages. Among the Lower Tamsui Fèng-shan language's closest links to Indonesian languages are Bugis, Sangir, and Manggarai, each with 5 similar words

Fourthly, to compare the four old Taiwanesian languages with the seven languages of New Guinea. The Tilaossen-Sinkan language shares 8 words with any New Guinea language, while the Formosan language has 5 shared words. The Lower Tamsui Fengshan language possesses 4 words similar to those found in New Guinea languages, and the Favorlang language has just 1 word in common with the languages of New Guinea. From the perspective of New Guinea, the Bariai language features 7 words that bear similarities to the four old Taiwanesian languages. This reveals that the linguistic link between the four old Taiwanesian languages and New Guinea is quite weak. Among the New Guinea languages, Japa Paiwa, Hiri Motu, and Bariai each have just one shared word—such as the term for “eye”—in common with the four old Taiwanesian languages.

Based on the analysis above, strong linguistic similarities are observed between the four old Taiwanesian languages and the Filipino language, with the Indonesian language following closely, while the relationship with New Guinea languages appears limited. Within the four old Taiwanesian languages, the Tilaossen-Sinkan language demonstrates a stronger connection to the Filipino language compared to the Favorlang language. Additionally, the Tilaossen-Sinkan language shows a greater resemblance to New Guinea languages than the other three old Taiwanesian languages. This suggests that the Tilaossen-Sinkan language likely absorbed elements from Papuan languages.

Phase 2

Among the current sixteen Taiwanesian languages, the Taroko and Seediq languages share a high degree of similarity, with 63 words in common. Likewise, the Taroko and Atayal languages closely resemble the Seediq language, sharing 27 similar-sounding words. Additionally, the La Aluwa and Kanakanavu languages have 25 words that are alike, while the Bunun and Thao languages share 13 similar words. Overall, apart from the Taroko and Seediq languages, the remaining Taiwanesian languages exhibit very few phonetically similar words, making mutual communication nearly impossible.

The comparative analysis between the sixteen Taiwanesian languages and other ethnic languages includes the following:

(1) A comparison of the four old Taiwanesian languages with the present-day sixteen Taiwanesian languages.

In the 17th century, the Dutch used a Roman script to document the sounds of four old Taiwanesian languages. Today, however, there are sixteen Taiwanesian languages still in use. When compared, these four old Taiwanesian languages are not closely linked to the current sixteen languages. Of the four, the Tilaossen-Sinkan language and the Rukai language show stronger connections, sharing 10 phonetically similar words, and display a slight relation to the other 14

languages, though they are more distant from the Atayal language. The Formosan language shows closer ties to 14 of the current languages, with the exception of Atayal and Tsou. In contrast, Favorlang exhibits no connection to Tsou or Kanakanavu and reveals only a weak relationship with the other 14 languages. Similarly, the Lower Tamsui Fengshan language lacks links to the Atayal and Taroko languages and is only minimally related to the remaining 14 current Taiwanesian languages.

(2) A narrative description of the connections between the sixteen Taiwanesian languages and the Filipino languages.

The sixteen Taiwanesian languages share the most similarity with the Ivatan (and Ibatan) language, which is geographically nearest to Taiwan. Notably, the Yami language has 30 words in common with Ivatan, while the Puyuma language shares 14 similar words. The Amis language, on the other hand, has 11 words similar to Ivatan.

Secondly, the close relation to the Hiligaynon language is the languages of Yami, Rukai, Kavalan, and Sakilaya, for example, the Yami language has 18 similar words; the Puyuma language has 12 similar words; the Rukai language, the Kavalan language, and the Sakilaya language has 11 similar words respectively; the Amis language has 10 similar words.

Thirdly, the Hiligaynon language is closely related to the following languages: Yami, Rukai, Kavalan, and Sakilaya. For instance, the Yami language has 18 similar words, while the Puyuma language shares 12 similar words. Additionally, the Rukai, Kavalan, and Sakilaya languages each have 11 similar words, and the Amis language shares 10.

Among the sixteen Taiwanesian languages, the Yami language shows the strongest connection to Filipino languages, sharing 15 words with Igorot, Cebuano, Ilocano, and Bikol, 14 words with both Tagalog and Ifugao, and 13 words with Waray. The Puyuma language also demonstrates notable links, with 13 similar words to Ilocano, Bikol, and Waray, 12 words in common with Cebuano, 10 with Ifugao, and 9 words each with Tagalog and Igorot. Similarly, the Sakilaya language shares 10 words with the Igorot, Cebuano, Ilocano, and Bikol languages, and has 9 words in common with Tagalog, Ifugao, and Waray. In contrast, the sixteen Taiwanesian languages show the least connection with the Pampanga language, likely because the Pampanga people live in the mountainous regions of Central Luzon, where limited access to the sea has historically restricted opportunities for migration.

The sixteen Taiwanesian languages share the least connection with the Pampanga language, primarily due to the fact that the Pampanga people reside in the mountainous regions of Central Luzon, where limited access to the sea restricts opportunities for migration.

		16 Taiwanesian Languages															
		Amis	Paiwan	Puyuma	Bunun	Thao	Rukai	Yami	Saisiya	Atayal	Taroko	Seediq	Tsou	Kavalan	Sakilaya	La Aluwa	Kanakanavu
Four Old Taiwanesian Languages	Tilaossen-Sinkan Language	5	9	8	6	8	10	8	6	1	4	6	4	4	5	4	4
	Formosan language	3	5	7	5	6	5	4	3	1	2	2	1	3	4	3	4
	Favorlang language	1	2	2	1	4	3	1	2	2	2	1	0	3	3	1	0
	Lower Tamsui (Fèng Shan)	2	2	4	3	2	3	3	2	0	0	1	2	3	3	1	1

Table 4. Phonic Similarity among the 16 Taiwanesian Languages, Southeast Asia, New Guinea, and Other Ethnic Languages

Philippine ethnic languages	Ibatan/Ivatan	11	5	14	10	9	7	30	6	2	3	3	3	9	11	8	4
	Hiligaynon	10	9	12	9	8	11	18	6	1	6	7	3	11	11	9	5
	Igorot	8	8	9	5	3	8	15	0	2	3	5	6	10	10	8	5
	Cebuano	9	7	12	5	5	9	15	4	4	5	5	2	9	10	5	5
	Ilocano	10	7	13	8	5	8	15	2	3	4	5	4	9	10	8	4
	Bikol	8	7	13	4	4	6	15	3	3	1	2	3	6	10	5	2
	Ifugao	10	7	10	4	5	9	14	1	1	5	5	5	7	9	6	4
	Tagalog	8	7	9	6	5	8	14	3	3	4	5	4	8	9	7	5
	Waray	8	7	13	4	5	8	13	1	1	4	4	3	8	9	7	4
	Pampanga	3	1	4	4	1	2	4	2	0	1	2	1	3	3	2	1
Malaysia ethnic languages	Sea Dayak	7	4	12	3	5	4	12	3	3	3	3	1	7	5	4	1
	Bahasa Malay	4	2	5	2	1	3	7	2	2	4	4	1	6	6	3	1
	Kadazan Dusun	11	8	16	9	7	8	15	9	5	8	10	4	15	14	9	5
	Bisaya	11	9	16	10	9	8	19	10	4	9	10	3	15	14	9	4
	Timugon Murut	6	5	9	6	7	4	10	6	2	5	5	2	10	9	4	1
	Bau Bidayuh	2	4	4	1	3	3	6	2	1	3	4	1	7	5	3	1
	Central Melanau	1	4	4	5	4	4	5	6	1	3	5	2	6	4	1	1
Indonesia ethnic languages	Sangir	11	9	16	7	**7**	10	16	3	3	4	5	5	14	13	9	4
	Bugis	7	5	11	6	8	8	12	1	1	5	7	6	10	9	7	4
	Toba Batak	8	6	13	4	5	6	11	4	2	3	4	3	10	11	3	4
	Minangkabau	5	5	10	2	2	4	9	5	5	2	3	1	8	7	3	0
	Bahasa Indonesia	4	4	7	2	2	4	9	3	3	6	7	3	8	6	4	2
	Sundanese	4	3	7	1	2	3	5	3	1	2	2	1	6	5	4	1
	Ternatenes	1	1	1	9	0	0	1	3	1	0	2	2	2	2	2	2
	Acehnese	3	1	8	2	2	3	10	3	3	2	2	0	7	6	3	0
	Javanese	9	8	12	9	4	7	15	5	4	6	8	3	13	10	8	6
	Madurese	8	5	7	2	3	5	6	2	2	4	5	3	8	6	5	5
	Balinese	7	5	14	8	7	10	13	7	3	4	6	5	15	13	9	6
	Banjar	8	6	13	5	5	7	16	5	4	7	7	1	12	8	6	4
	Makasar	7	6	11	5	5	7	11	7	1	5	5	2	13	7	7	4
	Manggarai	9	7	12	3	5	7	12	2	3	4	5	3	13	12	9	6
	Lamaholot	7	5	11	2	5	6	9	3	4	5	5	6	9	9	8	7
	Alor	6	6	12	5	5	7	9	3	2	3	4	6	10	10	9	8
	Bima	3	0	2	0	1	0	1	1	0	0	0	2	1	2	2	0
	Tetun	7	9	8	5	5	7	10	7	2	4	5	4	14	11	9	7
New Guinea ethnic languages	Gapa Paiwa	5	2	6	6	5	4	6	1	2	2	2	1	3	3	1	1
	Hiri Motu	3	3	3	4	2	4	5	1	2	1	1	1	3	3	1	1
	Koiari	2	1	1	0	1	1	2	0	1	0	0	1	0	1	1	0
	Tok Pisin	0	0	2	1	1	1	1	0	0	0	0	0	0	0	0	1
	Sentani	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
	Kwoma	0	0	0	1	1	1	0	1	1	0	0	1	0	0	0	0
	Bariai,	4	4	4	4	2	5	5	1	2	1	1	2	2	2	1	0
Other ethnic languages	Central Okinawa	3	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	Kinh (Vietnamese national language)	1	0	1	1	1	0	1	1	2	1	1	1	1	1	0	0

Note: The numbers represent the number of similar words.

(3) The narration of the linguistic connections between sixteen Taiwanesian languages and Malaysian languages.

Sixteen Taiwanesian languages share notable similarities with the Bisaya, Kadazan-Dusun, and Sea Dayak languages. For instance, the Bisaya language and Yami share 19 similar words, while Bisaya shows 16 shared words with Puyuma, 15 with Kavalan, and 14 with Sakilaya. Similarly, Kadazan-Dusun has 16 similar words with Puyuma, 15 with Yami and Kavalan, 14 with Sakilaya, and 11 with Amis. Sea Dayak also shares 12 similar words with each of Puyuma and Yami. However, relations among other languages are relatively weaker. These findings illustrate that the Yami, Puyuma, Amis, Kavalan, and Sakilaya languages in Taiwan demonstrate profound similarities with the Bisaya, Kadazan-Dusun, and Sea Dayak languages of Malaysia.

(4) A comparison of sixteen Taiwanesian languages with the Indonesian languages.

The study revealed a particularly strong connection between the Puyuma and Yami languages and the Indonesian languages. Puyuma shares 16 similar words with Sangir, 14 with Balinese, 13 with Toba Batak and Banjar, 12 with Javanese, Manggarai, and Alor, 11 with Bugis, Makassar, and Lamaholot, 10 with Minangkabau, 8 with Acehnese, and 7 with Sundanese. Yami has 16 words that are similar to those in both Sanghir and Banjar, 15 words to Javanese, 13 words to Balinese, 12 words to Bugis and Manggarai respectively, and 11 words to both Toba Batak and Makassar. In Indonesia, several languages share a close connection with Taiwanesian languages, such as Sanghir, Banjar, Balinese, Tetun, Toba Batak, Javanese, Makassar, Bugis, and Manggarai. On the other hand, Ternatenese and Bima exhibit the least resemblance.

When comparing the four old Taiwanesian languages and sixteen Taiwanesian languages with the languages of the New Guinea Islands, the correlation proves to be minimal, displaying a limited presence of Papuan linguistic elements. Similarly, it is evident that Southeast Asian ethnic languages share little connection with the languages of the New Guinea Islands. On average, only about 3-3.5% of phonetic similarities can be identified.

Table 5. Phonic Similarity among the New Guinea, Filipino, Indonesia, and Malaysian languages

N. Guinea SE. A	Gapa Paiwa	Hiri Motu	Koiari	Tok Pisin	Sentani	Kwoma	Bariai
Philippine ethnic languages							
Ibatan/Ivatan	4	3	1	2	1	0	3
Hiligaynon	5	3	1	3	0	0	5
Igorot	5	3	3	2	1	0	3
Cebuano	4	3	2	1	0	0	4
Ilocano	3	4	3	2	1	0	4
Bikol	3	3	1	2	2	0	6
Ifugao	4	3	3	0	2	0	6
Tagalog	3	3	3	1	1	0	4
Waray	5	3	2	1	0	0	3
Pampanga	2	1	0	1	0	0	4
Indonesian ethnic languages							
Sangir	7	4	1	2	0	3	6
Bugis	4	7	2	1	2	3	6
Toba Batak	4	3	0	1	0	2	4
Minangkabau	3	4	0	1	1	2	5
Bahasa Indonesia	3	4	2	1	0	0	2
Sundanese	2	2	1	0	1	0	3
Ternatenes	3	0	2	1	0	1	0
Acehnese	3	4	2	0	2	1	5
Javanese	4	5	1	3	0	2	4
Madurese	3	5	0	1	0	0	5
Balinese	5	6	1	1	1	0	5
Banjar	6	5	0	2	0	1	5
Makasar	5	5	0	2	0	1	4
Manggarai	2	4	1	1	0	0	4
Lamaholot	3	4	2	1	1	0	3
Alor	6	8	1	1	1	0	6
Bima	0	2	1	1	3	1	3
Tetun	5	8	1	3	2	1	8

Malaysian ethnic languages							
Bahasa Malay	4	3	0	2	0	1	5
Sea Dayak	4	4	0	2	1	3	6
Kadazan Dusun	6	4	2	3	1	1	6
Bisaya	4	5	1	1	1	0	6
Timugon Murut	2	3	1	0	2	2	4
Bau Bidayuh	1	2	0	0	0	0	2
Central Melanau	1	1	0	1	0	1	1

(6) To conduct a comparison of the Gia Rai language with the languages of Taiwan, the Philippines, Indonesia, and Malaysia, as illustrated in Table 6. The findings of the study are summarized as follows. The Gia Rai language shares a close relationship not only with Tagalog but also with nearly all Filipino languages studied. It has 20 words similar to those in both Ilocano and Hiligaynon, as well as 19 words similar to Tagalog.

Table 6. The Gia Rai language compares the similarity with the Taiwanesian, Filipino, Indonesian, and Malaysian languages

Languages	Ethnic	Similarity
Gai Rai	Amis	6
	Paiwan	6
	Puyuma	14
	Bunun	11
	Thao	6
	Rukai	10
	Yami	8
	Saisiya	2
	Atayal	6
	Taroko	7
	Seediq	1
	Tsou	8
	Kavalan	7
	Sakilaya	4
	La Aluwa	3
	Kanakanavu	10
Filipino	Tilaossen-Sinkan	7
	Formosan	3
	Favorlang	3
	Lower Tamsui	19
	Tagalog	16
	Ifugao	13
	Cebuano	18
	Pampanga	9
	Ilocano	20
	Bikol	13
	Waray	12
	Hiligaynon	20
	Ibatan/Ivatan	16
Indonesian	Indonesia	21
	Sundanese	12
	Minangkabau	21
	Toba Batak	14
	Ternatenes	3
	Bugis	16
	Sangir	17
	Acehnese	24
	Javanese	16
	Madurese	9
	Balinese	16
	Banjar	18
	Makasar	15
	Manggarai	12
	Lamaholot	11
	Alor	11
	Bima	3

	Tetun	16
Malaysian	Malay	20
	Sea Dayak	22
	Kadazan Dusun	19
	Bisaya	16
	Timugon Murut	12
	Bau Bidayuh	8
	Central Melanau	10

Indonesian language, Aceh, the region farthest from Gia Rai, shares the highest similarity with Gia Rai, with a total of 24 similar words. Following this are Minangkabau and Bahasa Indonesia, each having 21 similar words. In the context of the Malaysian language, Sea Dayak exhibits the greatest similarity to Gia Rai, with 22 shared words. As for the four old Taiwanesian languages, Tilaossen-Sinkan stands out, sharing the most similarities with Gia Rai, equating to 10 similar words. Among the 16 modern Taiwanesian languages, Puyuma shows the closest resemblance to Gia Rai, with 14 words in common.

This research identifies two distinct types of Taiwanesian languages: one is an early Taiwanesian language documented in the 17th century, and the other is a group of 16 contemporary Taiwanesian languages still spoken today. The old Taiwanesian language was primarily used in regions between Chiayi County and Kaohsiung County. During the Dutch colonial period, efforts were made to standardize the Sinkan language as the formal means of communication for the local tribes. However, following the Dutch withdrawal and the Sinicization policies implemented during the Qing Dynasty, the Sinkan people became Sinicized, resulting in the gradual decline and eventual extinction of their language. In contrast, the 16 modern Taiwanesian languages are spoken by tribes located outside the historical extent of the Sinkan language area. The old Taiwanesian languages contained elements of Papuan influence, suggesting both linguistic and ancestral ties. Meanwhile, the 16 Taiwanesian languages show fewer connections to the Papuan language, demonstrating stronger affiliations with the languages of the Philippines and Indonesia.

The historical book *Chu Fanzhi* from 1225 documents that the people of "Pí- shě- yé" in Taiwan traveled to "Sanyu" (interpreted as three islands) in the Visayas region of the Philippine Islands for trade (Feng, 1975). This interaction likely took place prior to the time this record was written.

The term "Pí- shě- yé" (also known as Visaya) emerged in southern Taiwan during the 13th century, in northern Taiwan by the 17th century, and was also associated with the Visaya ethnic group in the central Philippines and Sarawak. The repeated appearance of the same name in different regions suggests a potential ethnic connection between these groups.

Japanese scholar, Inō Kanori, proposed in his study that the "Pí-shě-yé" people migrated from the Philippines, traveling via Siaoliuqiu Island to reach the Tainan region (Inō, 1993). His theory can be understood as an attempt to explain the migration of populations between Borneo, the Philippines, and Taiwan.

Among the Filipino languages, Hiligaynon is the one most closely linked to the Taiwanesian languages. This language family, primarily spoken in the western parts of the Visayan region and Soccsksargen, ranks as the second-largest in the Visayan linguistic group. It may correspond to the "Sanyu" region referenced in the *Zhu Fanzhi*. According to the text, Taiwan's "Pí-shě-yé" people traveled to "Sanyu" for trade, exchanging goods such as sand gold, dried leopard meat, and honey. Based on this record, it is thought that the early indigenous peoples of Siaoliuqiu Island and the Dà Yuán region (now Tainan) may have been the "Pí-shě-yé" people, originating from the "Sanyu" area in what is now considered the central region of the Philippine Islands.

In the context of Indonesian languages, those most closely related to Taiwanesian languages include Sangir, Bugis, Toba Batak, Minangkabau, Makasar, Javanese, Balinese, Manggarai, Lamaholot, Alor, and Tetun. The Puyuma language, for instance, shares 13 similar words with Toba Batak, 10 with Minangkabau, and 8 with Acehnese. Geographically, the Toba Batak and Minangkabau peoples inhabit the northern and central parts of Sumatra, respectively. To explore the historical ties between these ethnic groups and Taiwan, we can look at historical records on the Nanwuli kingdom which is located in Banda Aceh, neighboring to Batak kingdom. Yang Yikui's historical work, *Yicheng* (Family Multiply), notes: "In 1282... numerous countries, such as Nanbuli, Malandan, Nawang,

Dingheer, Lailei, Jilan Yihu, and Sumudula, dispatched envoys to pay tribute (to China)" (Yang, 1981). Nanwuli continued its tributary missions in the years 1286, 1294, and 1375, underscoring its role in regional diplomacy.

In ancient Chinese records, the Batak people were referred to as Baihua, meaning "tattooing a hundred flowers on their faces," due to their custom of facial tattoos resembling floral patterns. The Bataks sent envoys to China for trade in the years 1375, 1378, 1394, and 1397. They were also called Hua Mian in old Chinese texts (Chen, 1986), highlighting their tradition of adorning their faces with tattoos reminiscent of flowers. It is possible that during one of their voyages to China, the Bataks encountered strong winds that caused them to drift off course and arrive in Taiwan by accident. Interestingly, some Taiwanesian groups, including the Amis, Atayal, Taroko, and Seediq peoples, share the practice of facial tattooing, suggesting the possibility of a historical or cultural connection between them.

The Bataks who journeyed to China might have departed from the Aru Port, located south of Medan in the northeastern region of Sumatra Island. According to records from Emperor Ming Taizu's era (mid-14th century), it was noted that the Baihua people practiced Buddhism. In 1378, the Baihua Kingdom sent an envoy to present tributes, and its ruler was named Lading Lazhe Wangsha. Baihua, identified as the ancient Chola Kingdom, was also considered a country within the South China Sea region. This kingdom spanned approximately 2,500 miles from east to south, and its people were known for their wealth and their adherence to Buddhism (He, 1971). If individuals from Baihua had unintentionally ended up in Taiwan, it most likely occurred prior to that era, as the Taiwanesian population lacked any form of Buddhist beliefs until the arrival of Dutch colonial rule. Alternatively, those Baihua individuals who drifted to Taiwan may have followed animistic religious practices, sharing similarities with the early Taiwanesian community in their rituals, such as facial tattooing and headhunting.

Minangkabau lies to the south of the Batak Kingdom. The Minangkabau language shares a close relationship with the Sinkan and Puyuma languages, marked by comparable phonetic words and a shared matriarchal social structure. Between the 7th and 14th centuries, the Sri-Vijaya kingdom governed the eastern and southern regions of Sumatra, while the central and western areas remained under the authority of the Minangkabau. In 1347, Adityawarman, originating from the Dharmasraya kingdom in Jambi, founded a new kingdom located in Suruaso and Pagaruyung, situated along the upper Batang Hari River. This kingdom embraced Mahayana Buddhism (The Spice Route End, 2019). Sri-Vijaya also practiced Buddhism, whereas the central part of Sumatra, which stayed outside Sri-Vijaya's sphere of influence, is believed to have continued practicing fetishism until 1347. As such, prior to this time, people from that region who migrated to Taiwan or the Gia Rai area of Vietnam might have still adhered to those fetishistic traditions.

The Minangkabau people have a longstanding tradition of migrating to other regions. Originating from the mountainous areas stretching from Jambi to Padang, they began moving abroad as early as the mid-15th century, with migrations to places like the Malay Peninsula and East Asia. For instance, they settled in areas such as Naning and Sungai Ujong in Negeri Sembilan. Known for their matriarchal way of life, they thrived in agriculture and thus gravitated towards inland locations (Hooker, 2003). Under their social norms, marrying within the same tribe was strictly forbidden. Men were expected to marry women from different tribes and reside with their wife's family. In this matriarchal household structure, women were the primary figures, holding ownership of family assets. Upon the mother's death, ownership would pass down to her daughter (Hooker, 2003). Similarly, cultures such as the Sinkan, Amis, and Gia Rai also practiced matrilineal systems of inheritance.

In *The Aborigines of Formosa* by George Taylor, a British consular officer in the Qing Dynasty, he recorded that an English ship was wrecked in eastern Taiwan by a storm, a crew from Kelapa (referring to Batavia or Java) was rescued by the natives for his language similar to local natives, and other white members were all killed. After that, the Kelapa married a local girl (Liu, et. al., 1998).

Based on the aforementioned cross-comparative study analyzing strong and weak phonetic similarities among ethnic languages, the connection between Taiwanesian languages and Southeast Asian languages becomes apparent when there are at least nine similar words. Examples of such languages in Taiwan include Amis, Paiwan, Puyuma, Bunun, Rukai, Yami, Sakilaya, Kavalan, and La

Aluwa, which show close phonetic ties with certain Southeast Asian languages. These Southeast Asian related languages, listed in order, are Ivatan, Hiligaynon, Sangir, Bugis, Sea Dayak, Kadazan Dusun, Bisaya, Timugon Murut, Minangkabau, Toba Batak, Javanese, Balinese, Banjar, Makassar, Manggarai, Lamaholot, Alor, and Tetun.

From this analysis, it can be inferred that emigration towards Taiwan may have followed two potential routes. One route likely extended from northern Sumatra to Minangkabau, Sarawak, the Visayan Islands, and southern Taiwan. The other route began at Timor Island, passed through Alor, Manggarai, Bali, Java, Makassar, and Bugis, merging with the previously mentioned route, or alternatively, it traveled directly from Timor Island via Alor and Manggarai to Taiwan without involving the Moluccas Islands. Consequently, the linguistic connection between the ethnic groups of the Moluccas Islands and the Taiwanesians appears to be minimal.

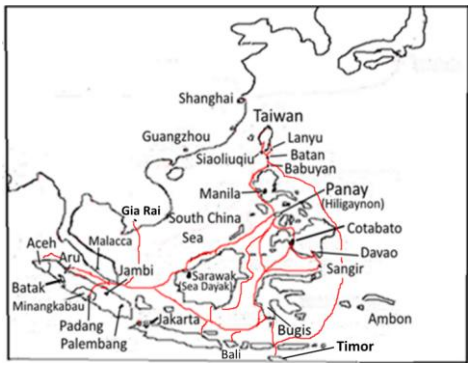

Figure 1. The migration route of Austronesian people to Panay Island and then to southern Taiwan (Source: Drew by the author,)

To determine that the linguistic group's direction of movement was from south to north, historical records and experimental evidence can provide support for this hypothesis. One example is the mention of the "Pí-shě-yé" people in the 13th-century text *The Zhu Fanzhi*. This term refers to an ethnic group and place name in the central Philippines, rather than a location in Taiwan. The "Pí-shě-yé" people were believed to have resided in southern Taiwan and may have migrated there from the Visayan region, traveling through northern Luzon, the Babuyan Islands, and the Batanes Islands before reaching southern Taiwan.

The indigenous Hiligaynon people of Panay Island are believed to have migrated from Borneo. According to Filipino historian Gregorio F. Zaide, the Borneo kingdom once had dominion over Panay Island, with its influence centered in two primary areas: Sulu and Visaya. At the time, Borneo was a tributary state under the Sri-Vijaya Kingdom. Officials from Borneo were sent to Sulu, where one of them married a princess from the Sulu royalty. By the mid-13th century, a group of 10 Bornean chiefs, led by Datu Puti, migrated to Panay Island to establish settlements (Zaide, 1957). Historical accounts also refer to the area surrounding Panay Island as “Sanyu,” which was known to engage in trade with Taiwan. These connections suggest a historical linkage between Borneo, Panay, and Taiwan.

Additionally, the oral legends passed down through generations are crucial for uncovering ancestral roots. Japanese anthropologist Torii Ryuzo researched the Yami people and discovered accounts of their ancestors traveling by canoe from the Batanes Islands in the Philippines to Lanyu (Botel Tobago). Their traditions and language bear a strong resemblance to those of the Batanes Islanders (Ryuzo, 1996). It is believed that the Batanes people likely reached Lanyu by the Kuroshio Current.

Not to mention, on October 15, 1897, Inō Kanori, a Japanese anthropologist, visited the Sinkan community in Kaohsiung. While there, he learned from the local aborigines about an oral tradition recounting their migration to Taiwan. According to their account: "Our Sinkan ancestors originally drifted to Taiwan from islands overseas, eventually landing on Sǐ-Shù-Zǐ-Yuán, now called Sì-Kūn-Shēn (Fourth Island), a small islet located roughly ten miles south of the Tainan River. We initially settled there temporarily before gradually spreading towards what is now the Tainan City area" (Inō,

1996). This suggests that the indigenous inhabitants around Tainan originally came from islands to the south.

In addition, the Amis people of Hualien share a comparable oral tradition. According to their legend, their ancestors brought breadfruit trees along with other fruit-bearing trees from overseas to Taiwan (Ryuzo, 1996).

This study questions the long-standing assumption that the Taiwanesian people speak Bahasa Malay or Bahasa Indonesia. When the Dutch arrived in southern Taiwan, they mistakenly believed the indigenous population spoke Bahasa Malay. However, this study suggests otherwise. The four old Taiwanesian languages show very little resemblance to Bahasa Indonesia or Bahasa Malay, and the same holds true for the current sixteen Taiwanesian languages. Of these, the Yami language shares 9 words with Bahasa Indonesia, while the other fifteen Taiwanesian languages share even fewer. Similarly, Yami contains 7 words similar to Bahasa Malay, whereas the remaining fifteen Taiwanesian languages have even fewer similar terms. These misconceptions can be corrected through comprehensive research.

Bahasa Malay historically served as the language of the Malay people residing in northeastern and northern central Sumatra. The Sri Vijaya Dynasty, which flourished in Palembang and Jambi, reached its zenith during the 12th and 13th centuries. Prior to Malay emerging as the primary trade language in the Southeast Asian islands in the same era, various ethnic groups from the Indonesian archipelago had already migrated to the Philippine Islands and even reached Taiwan. As a result, the four old Taiwanesian languages spoken during the period of Dutch rule in Taiwan exhibit minimal connection to Malay.

The Gia Rai language holds significant importance for this study due to its close connections with the ethnic languages under examination. Nonetheless, its phonetic resemblance to Luzon appears weaker when compared to Aceh. Consequently, it seems unlikely that the migration route would have followed the path from Gia Rai to Luzon, then to Borneo, and ultimately to Aceh. Instead, it is more plausible that the migration path extended from Aceh and Minangkabau to Borneo, where it split into two branches, one leading to Gia Rai and the other toward the Visayan region in the central Philippines. Another possibility is a direct movement from Sumatra to Gia Rai. Based on historical records, one theory suggests that the Gia Rai people might have been moved to southern Vietnam by groups such as the Nanwuli or Baihua tribes during their trade expeditions to China. They could later have been relocated to Gia Rai from the south following invasions by the Cham or Vietnamese, which then pushed them inland to the central highlands. Alternatively, an even earlier hypothesis links the Gia Rai to the remnants of the Sailendra Dynasty's military forces, which were dispatched from central Java to occupy Zhen-la (Cambodia) in the eighth century (Briggs, 1951). The region currently settled by the Gia Rai people, Gia Rai province, historically formed a part of Zhen-la's territory.

Based on this research, the Okinawans situated north of Taiwan and the Vietnamese located west of Taiwan speak different languages from the Taiwanesians. Additionally, the Taiwanesians do not share close ethnic ties with either group.

CONCLUSION

Based on the analysis, the four old Taiwanesian languages exhibit strong connections to Filipino languages such as Ivatan, Hiligaynon, Ilocano, and Tagalog. They show a moderate degree of connection to Indonesian languages like Bugis, Sangir, Toba Batak, and Minangkabau, as well as to Malaysian languages such as Sea Dayak and Bahasa Malay. However, they share a more distant relationship with the languages of New Guinea. Similarly, among the sixteen modern Taiwanesian languages, this pattern persists, as they all fall within the same language family. I propose that this linguistic relationship may potentially hold some racial significance.

This study examines the basic vocabulary of 67 languages spoken across Taiwan, Southeast Asia, and New Guinea, revealing a notable similarity between the languages of Taiwanesian groups and Southeast Asian ethnic groups. These findings point to a linguistic and ethnic link among these groups. Additionally, analysis of historical records supports the conclusion of migration from south to north, which seemingly occurred before the 13th century. Certain ethnic groups in Southeast Asia, before the prevail of Hinduism, Buddhism, and Islam, migrated to Taiwan. As a result, when the Dutch arrived on the island in the 17th century, Taiwanesians practiced animistic beliefs. Building on this

analysis, it presents a valuable opportunity for scholars from both Taiwan and Indonesia to delve into the fascinating realms of racial and linguistic relevance. This exploration could yield insights that enrich our understanding of both cultures and foster meaningful connections between them.

ACKNOWLEDGMENT

Acknowledgments for Prof. Dr. Hj. Nadra, M.S., Study Program of Linguistics, Faculty of Humanities, Universitas Andalas, Padang, Indonesia, and Vincent Kewibu, senior lecturer in Archaeology/Anthropology at the University of Papua New Guinea, both linguists, provided me with professional linguistic opinions.

REFERENCES

Ager, Simon. (2024). *Hiligaynon (illonggo).* Omniglot. https://www.omniglot.com/writing/hiligaynon.htm

Bellwood, Peter. (1992). Southeast Asia before history. In N. Tarling (Ed.), *The Cambridge History of Southeast Asia, Vol. One, From Early Times to C.1500* (pp. 53-136). Cambridge University Press.

Bellwood, Peter. (1991). The Austronesian dispersal and the origin of languages. *Scientific American*, *265*(1), 88-93.

Benedict, Paul K. (1975). *Austro-Thai language and culture, with a glossary of roots.* Human Relations Area Files.

Blussé, L., Natalie E., & E. Frech, (Eds.), The Chinese translated by Lin Weisheng. (2010). *The Formosan encounters, Taiwanesian society documentary: A summary of the Dutch archives.*1, 1623-1635. Council of Indigenous Peoples, Shunyi Museum.

Blust, Robert A. (1985). The Austronesian homeland: A linguistic perspective. *Asian Perspectives*, *26*(1), 45-67. https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/8ecf4de6-7b99-40aa-b5fe-0d829b68998c/content

Campbell, Lyle. (1999). *Historical linguistics: An introduction*. MIT Press.

Chen, J., Xie, F. & Lu, J., (Eds.) (1986). *An interpretation of ancient South China Sea place names.* Zhonghua Book Company.

Dang N. V., Chu, T. S., & Luu, Hung. (1993). *The ethnic minorities in Vietnam.* The Gioi Publishers.

Dalisdis, J. (2018). Igorot (People of the mountains). Cprdilerran Culture. https://inyamet00.wordpress.com/2018/04/13/igorotpeople-of-the-mountains/

Dodd, J. (1882). A few ideas on the probable origin of the hill tribes of Formosa. *Journal of the Straits Branch of the Royal Asiatic Society.* 69-84.

Elliot. (2019). *Minagkabau culture: History.* The Spice Route End. https://thespicerouteend.com/minangkabau-culture-sumatra-indonesia-history/#an-ethymology-of-the-word-minangkabau

He, Q. (1971). *Mingshan Zang* (Records on Famous Materials), Volume 63. Wang Xiangji, Three Southeastern Yi. Chengwen Publishing Company.

Feng, C. (1975). *Zhu Fanzhi annotation.* Taiwan Commercial Press Company.

Hooker, V. M. (2003). *A short history of Malaysia: Linking East and West*. Allen & Unwin.

Inō, K. (1993). The approximation of Bisaya in the Philippine Islands and Siraiya in Taiwan. In Huang Xiumin's translation, *A compilation of Japanese-Chinese translations of Taiwan's South Island language research papers* (pp. 34-35). Preparatory Office of the National Taiwan Museum of Prehistoric Culture.

Inō, K, Annotation by Yang Nanjun. (1996). *Taiwan tacha diary* (*A diary of investigation of Taiwan*) (Part 1). Yuanliu Publishing Company.

Jiang, S., translated/annotated. (1999). *The diary of Kasteel Zeelandia.* Tainan City Government.

Jiang, S, translated/annotated. (2015). *Letters from the Dutch Taiwanese Governor to the Governor of Batavia* (V) (1629-1636[3]). Taiwan Historica.

Liu, H., Li, S., Huang, T., Huang, Z., &Wang, Z. (1998). *Searching for the Kaitagelan People, the Culture, and current situation of the Kaidagrand People.* Taipei County Cultural Center.

Manuel, E. A. (1994). *Documenting Philippineasian: An inquiry into the ancestor of the Filipino people, their early culture, and prehistory before the Christian era.* The Philippineasian Society.

Statistics Indonesia. (2011). *Nationality, ethnicity, religion, and daily language of Indonesian population.* Statistics Indonesia. https://www.neliti.com/publications/49956/nationality-ethnicity-religion-and-daily-language-of-indonesian-population

Taborete, M.. (n.d.) Philippine ethnic group (Waray) reporter. *Academia.edu*. https://www.academia.edu/35646103/PHILIPPINE_ETHNIC_GROUP_WARAY_REPORTER

The Editors of Encyclopedia Britainnica. (n.d). *Kuroshio, oceanic current, Pacific Ocean* Britainnica. https://www.britannica.com/place/Kuroshio

Torii, R. (1996). The Chinese translated and annotated by Yang Nanjun. *Expedition to Taiwan.* Yuanliu Publishing Company.

Wirth, A. (1897). The Aborigines of Formosa and the Liu-Kiu islands. *American Anthropologist*, *10*(11), 357-370. https://www.jstor.org/stable/pdf/658522.pdf

Yang, Y. (1981). *Yicheng* (Family Multiply), Volume 8(I), Nanyi Volume II. National Central Library.

Zaide, G. F. (1957). *Philippine political and cultural history. Vol.1*. *The Philippines since pre-Spanish times.* Philippine Education Company.