LEXICOSTATISTICS OF MALAY, TAGALOG AND ILOCANO LANGUAGES: A COMPARISONAL HISTORICAL LINGUISTIC STUDY

Oleh:
Muhammad Kiki Wardana[1)], DwiWidayati[2)], RostinaTaib[3)], Muhammad Iqbal[4)]
[1]Departemen Sastra Inggris Universitas Harapan Medan
[2]Departemen Linguistik Universitas Sumatera Utara
[3,4]Departemen Pendidikan Bahasa Indonesia, Universitas Syiah Kuala

Abstract

This study tries to explore the relationship between three languages, namely Malay, Tagalog, and Ilocano. The purpose of this study is to find cognates (word relatives), estimate when these three languages are a single language and find the time apart from the three languages. This study uses a language grouping method with a Lexicostatistical approach. The results of this study indicate that Malay and Tagalog share about 28% of words that are related and fall into the category of Stock Clumps. Other results also show that Malay and Ilocano historically and linguistically are closer and classified as category of language stockwith the cognate or related words as much as 31%.

Keywords: *Malay, Tagalog, Ilocano, Lexicostatistics, Cognate*

1. INTRODUCTION

The Malay language which is the forerunner of the unifying language of the Indonesian nation has been widely used by the people of the archipelago for a long time. The existence of Malay (BM) as a trading language and vernacular, prestige language, and lingua franca in several regions of the archipelago strengthens the reason why the people of the archipelago can easily accept Malay. Collins (2005) also explains that Old Malay is not only used on the peninsula of Sumatra but is also widely used in Java as evidenced by the discovery of several inscriptions also written in Old Malay in Central Java (Gandasuli Inscription, year 832).) and in Bogor (Bogor Inscription, year 942). (Burke, 2015) also confirms that the Old Malay language has also been widely used in the Sriwajaya Kingdom which is strengthened by the discovery of several ancient inscriptions such as the Kedukan Bukit Inscription in Palembang, in 683; (2) the TalangTuoInscription in Palembang, 684; (3) Kota Kapur Inscription in West Bangka, year 686; and (4) the Karang Brahi Inscription between Jambi and the Musi River, 688.

The discovery of artifacts from the Sriwijata Kingdom in the form of copper chips written in Old Malay was also found in Laguna, Philippines (Putrayasa, 2018). The historical evidence found further strengthens the influence of the Srivijaya Kingdom, which was very wide reaching to the Philippines. Therefore, this finding is also one of the reasons why this research was conducted, namely to determine the relationship between Malay and Tagalog and Ilokano with Lexicostatistics and Glotochronology methods. Tagalog which is also a member of Proto Austronesian (PAN) has also become the National Language in the Philippines. But (Arizo et al., 2020) also emphasize that unlike Indonesian, Tagalog only has speakers one third of the Philippines' population. While Ilocano is one of the 4 largest languages in the Philippines. Many Ilokano speakers migrate to the islands of Luzon and Manila to find work and mingle with the local community (Tamayao, n.d.).

What these two Filipino regional languages have in common is the fact that they are also included in Proto Austronesian (PAN). When viewed from the side of history and development, PAN has undergone many changes, both from the phonological, morphological and syntactic levels. (ChurmatinNasoichah et al., 2020) in his research revealed that the PAN in the archipelago underwent changes due to the influence of other nations who brought new cultures, different socio-cultural systems, government systems and the use of different scripts. (Blust, 2010) also argues that PAN speakers have interacted with immigrants, especially from India, starting from 2000 years ago. This fact shows that even though Malay, Tagalog and Ilokano originate from one family, they have undergone changes, shifts and erasures at all levels of language, both in terms of sounds, morphemes and syntactic labels. Several related studies have shown that to determine the duration of the relationship between languages and to find the separation time of languages from one proto family, the Lexicostatistics Technique is very suitable to be used. According to (Dardanila et al., 2020) lexicostatistics is a technique that allows linguists to determine the degree of relationship between two different languages.

Previous research that has been conducted regarding the lexicostatistics of these three languages was carried out by (Nelson, 2001) who broke the assumption that Tagalog has more words related to Ilokano than Bikolano, but it turns out that Tagalog,

Ilokano, and Bikolano have a cognate proportion. the same one. Meanwhile, research conducted by (Sofiyatunnida&Hendrokumoro, 2021) on lexicostatistics of Malay and Mandailing revealed that Batak Mandailing and Malay have a percentage of 58% kinship. Based on lexicostatistical calculations of 200 lists of Swadesh vocabulary in Batak Mandailing and Malay, it was found that 114 kinship vocabularies and 84 non-relative vocabularies. The conclusion of this research is that Batak Mandailing and Malay are related and belong to the language family level. Another study on Malay was conducted by (Istiqamah, 2017) which found that (1) the kinship level of Acehnese language with Malay was 48.4%, (2) Acehnese language and Malay began to separate from its proton language (parent language) since 1,635 years. ago, (3) the Acehnese language with the Malay language was included in the classification of the language family.

Lexicostatistics research is a development of comparative language research using the Comparative Method which was originally applied to Indo-European languages which was pioneered by Ludwig Karl (Nurhayati, 2017).

The lexicostatistic technique not only serves to determine the percentage of kin words and calculate language age, but can also be used for grouping kin languages. Languages that show a high percentage of kinship are groups that are closer in membership, while those with a low percentage of kinship are groups whose membership level or kinship is more distant.

Swadesh proposed a classification of language kinship systems, namely:

Tingkatan bahasa	Waktu pisah dalam abad	Persentase kata kerabat
Bahasa (*Language*)	0-5	100-81
Keluarga (*Language Family*)	5-25	81-36
Rumpun (*Stock*)	25-50	36-12
Mikrofilum	50-75	12-4
Mesofilum	75-100	4-1
Makrofilum	100-ke atas	1- kurang dari 1 %

(Sumber: Keraf, 1996:135)

Therefore, based on the description above, the formulation of the problem in this study is:

1. What is the cognate (kinship) of the two languages?
2. When are these two languages thought to be a single language?
3. When are these two languages separated from their proto-languages?
4. What is the error time between the two languages?

2. RESEARCH METHODS

This research was studied using language grouping methods and lexicostatistical techniques. The first stage, collects 207 basic vocabularies compiled by Morris Swades. The method used in providing this data is a referential method, while the technique used is a note-taking technique (Kesuma, 2007:48; Sudaryanto, 1993:13-16; Sudaryanto, 1988:5). Second, determine the word kin (cognate) to find the percentage of kinship from the three languages by classifying based on: (a) identical pairs, (b) phonemic correspondence pairs, (c) phonetically similar pairs, (d) different pairs. one phoneme. Third, calculate the age and separation time of the three languages and also calculate the error range to determine a more precise separation time. Fourth, the preparation of the classification of kinship systems, whether as one language (language), language family (subfamily), language family (stock), microphylum, mesophyllum, or macrophylum (Keraf, 1996:126-128).

3. HASIL DAN DISKUSI

a. Leksikostatistik pada Bahasa Melayu dan Tagalog

Seusaimenetapkan kata kerabat, langkahselanjutnyayaknimencaripersentasekerabatdenganrumus:

$$C = \frac{Vt}{Vd} \times 100\%$$

Keterangan:
C = kata kerabat;
Vt = jumlahkosakatakerabat;
Vd = jumlah gloss yang diperhitungkan

$$C = \frac{Vt}{} \times 100\%$$
$$= \frac{59}{} \times 100\%$$
$$= 0{,}28 \times 100\%$$
$$= 28\ \%$$

Vd 207

Setelah persenta sekerabat diketahuihasilnya, kita dapat menghitung waktu pisah dari Bahasa Melayu dan Bahasa Tagalog

diketahui:
C = 28 %
log r = 80,5 %
ditanya : W = ...?
jawab : W = $\frac{\log.C}{2 \log.r}$

$$W = \frac{\log 0{,}28}{2 \times \log 0{,}805}$$
$$W = \frac{-1{,}273}{2 \times (-0{,}217)}$$
$$W = \frac{-1{,}273}{-0{,}434}$$
$$W = 2{,}933$$

The split time is multiplied by 1000 so that the result is 2,933. So, the calculation of the initial separation time for Malay and Tagalog is 2,933 years ago. After the results of the separation of Malay and Tagalog are known, the next step is to calculate the error term. This is done to avoid miscalculations and to set a more precise separation time. It should be remembered that to anticipate errors in statistics is to give an estimate, ie not in a certain time, but in a certain time period. To calculate the error term can be

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

using the formula:
S = Standard error in the percentage of relative
C = Percentage of relatives kata
n = Number of words compared, both relatives and non-relatives
is known:
C = 0.28
n = 207
asked: S = …..?
answer: $S = \sqrt{C(1-C}$

$$S = \sqrt{\frac{0{,}28\,(1-0{,}28)}{n}}$$

$$= \sqrt{\frac{0{,}28x\ 0{,}72}{207}}S$$

$$= \sqrt{\frac{0{,}2016}{207}}$$

207
= 0,030 (rounded to 0, 03)

The result of this standard error (0.03) is summed with the percentage of initial relatives (C1) to get C2 (C2 = C1 + S). So C2 the result is 0.28+0.03=0.31. With C2, the separation time can be calculated again, using the same formula:
Is known:
C2 = 31%
log r = 0,805
asked: W2= .....?
answer: $W2 = \frac{\log 0{,}31}{2 \log 0{,}805}$

$$W2 = \frac{\log 0{,}31}{2 \times \log 0{,}805}$$

$$W2 = \frac{-1{,}171}{2 \times -0{,}217.}$$

$$W2 = \frac{-1{,}171}{-0{,}434}$$

This split time is multiplied by 1000 to get 2,698.W2= 2,698.
To calculate the error period, the long time is reduced by the blue time. Thus, we can calculate the error period is = W1 – W2 = 2,933 – 2,698 = 225. This number must be added and subtracted by the long time to get the age and separation time of the two. the language

(1) Malay and Tagalog are estimated to be a single language at 2933 +225 = 3158; 2933 – 225= 2708 years ago
(2) Malay and Tagalog were a single language in 3158-2708 last year.
(3) Malay and Tagalog are thought to have started to separate from the Proto language circa 1136 - 686 BC - (calculated in 2022).

2. Lexicostatistics in Malay and Ilocano

After determining the word relatives, the next step is to find the percentage of relatives with the formula:

$$C = \frac{Vt}{Vd} \times 100\%$$

Information:
C = relative word;
Vt = number of relatives' vocabulary;
Vd = amount of gloss calculated

$$C = \frac{Vt}{Vd} \times 100\%$$

$$= \frac{65}{207} \times 100\%$$

= 0,27 x 100%
= 31 %

is known:
C = 31 %
log r = 80,5 %
asked: W= .....?
answer: $W = \frac{\log.C}{2 \log.r}$

$$W = \frac{\log 0{,}31}{2 \log 0{,}805}$$

$$W = \frac{-1171}{2 \times -0{,}217}$$

$$W = \frac{1171}{434}$$

W = 2.698

Malay and Ilocano separated 2,698 years ago. If calculated in the current year (2022) then the two languages separated around 676 BC

After the results of the separation of Malay and Ilocano are known, the next step is to calculate the error term. This is done to avoid miscalculations and to set a more precise separation time. It should be remembered that to anticipate errors in statistics is to give an estimate, ie not in a certain time, but in a certain time period. To calculate the error term can be using the formula:

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

S = Standard error in the percentage of relative
C = Percentage of relatives kata
n = Number of words compared, both relatives and non-relatives
is known:
C = 0.31
n = 207
asked: S = …..?
answer: $S = \sqrt{C(1-C}$

$$S = \sqrt{\frac{0{,}31\,(1-0{,}31)}{n}}$$

207

$$= \frac{\sqrt{0,31 \times 0,69}}{207}$$

= 0,0321
(rounded to 0, 03)

The result of this standard error (0.03) is summed with the percentage of initial relatives (C1) to get C2 (C2 = C1 + S). So C2 the result is 0.31+0.03=0.34. With C2, the separation time can be calculated again, using the same formula:

is known:

C2 = 34%
log r = 0,805
asked: W2= .....?
answer: $W2 = \frac{\log 0,34}{2 \log 0,805}$

$W2 = \frac{1079}{2 \times 217}$

$\frac{1079}{434}$

W2= 2.486

Thus, we can calculate the error range is = W1 – W2 = 2.698 – 2.486 = 212. This number must be added and subtracted by a long time to get the age and separation time of the two languages.

So, the age of Malay and Ilocano can be expressed as follows:

(1) Malay and Ilocano are estimated to be a single language at 2,698+212=2,910; 2.698 - 212 = 2.486 BC

(2) Malay and Ilocano were a single language in 2698-2486 BC.

(3) Malay and Ilocano separated from the Proto language circa 676 BC – 464 BC (calculated in 2022).

4. CONCLUSION

The Proto Austronesian languages play an important role in the map of the world's language distribution. Malay, Tagalog and Ilocano are also languages that have strategic functions in the countries that use them, namely Indonesia and the Philippines and several other countries such as Malaysia, Brunei Darussalam and Singapore. The relationship between these languages needs to be known and studied to get a bigger picture of this language family. Lexicostatistics is a method that can be used to measure the closeness between proto-languages, regardless of their type. This study only focuses on determining cognates or related words, as well as the separation time between languages. From the results of this study it is known that Malay and Tagalog are estimated to have started to separate from their mother tongue around 1136 -686 BC - (calculated in 2022). Meanwhile, Malay and Ilocano are thought to have started to separate from their proton languages around 676 BC – 464 BC (calculated in 2022). From this interesting finding, it can be concluded that Malay which is the forerunner of Indonesian has a higher affinity with Ilokano which is one of the regional languages in the Philippines than Tagalog itself which is the national language in the Philippines.

Appendix I

No U Data	Bahasa Melayu	Bahasa Tagalog	Bahasa Ilokanos	Glossary
	Aku/saya	Ako	siak	I
	Dia	Siya	-	He/She
	kami	Kami	Dakami	We
	kapan	Kalian	kano	When
	Satu/Esa	Isa	Maysa	One
	Dua	-	Dua	Two
	Tiga/elu		tallo	Three
	empat	Apat	Uppat	Four
	lima	Lima	Lima	Five
	Engkau	Ikaw	-	You
	tipis	Nipis	Ingpis	Thin
	Anak	Anak	Anak	Child
	Asu	aso	aso	Dog
	Kutu	Kutu	kuto	Leech
	Ikan	-	Ikan	Fish
	Ulat	Uot	-	Worm
	Ular	-	Uleg	Snake
	Kayu	-	Kayo	Wood
	Benih	Binhi	Bini	Seed
	Daun	Dahoon	-	Leaf
	Rumput		Ruot	Grass
	Kulit	-	kudil	Skin
	Tali	Tali	Tali	Rope
	Hulu	Ulo	Ulo	Head
	Telinga	Tainga	-	Ears
	Mata	Mata	Mata	Eyes
	Hidung	Ilong	Agong	Nose
	Lidah	Dila	Dila	Tongue
	Kuku	Kuko	Kuko	Fingernail
	Tulang	-	tulang	Bone
	Leher	Leeg	-	Neck
	Susu	Suso	Suso	Breast
	Hati	Atay	-	Liver
	Minum	Inom	Inom	To drink
	Ludah	Luwa	-	To spit
	Makan	-	Mangan	To eat
	muntah	-	Uta	To vomit
	Tawa	tawa	Katawa	To laugh
	Takut	Takot	-	To fear
	Darah	-	Dara	Blood
	Bulu	Buok	-	Hair
	Mati	Patay	Matay	To die
	dengar	-	dengngeg	To listen
	Lawan	Laban	-	To fight
	Pisah	-	Bisak	To split
	Gali	-	Kali	To dig
	jahit	-	dait	To sew
	Datang	Dating	-	To come
	Tolak	Tulak	-	To push
	Bilang	Bilang	Bilang	To count
	Bulan	Buwan	Bulan	Moon
	Bintang	Bituin	bituin	Star
	Hujan	Ulan	-	Rain
	danau	-	Danaw	Lake
	Laut	Laot	-	Sea
	Masin	Asin	Asin	Garam
	Batu	Bato	Bato	stone
	darat	-	Daga	earth
	Langit	Langit	Langit	sky
	Angin	Hangin	Angin	wind
	Asap	Aso	Asok	smoke
	Api	Apoy	Apoy	fire
	Abu	abo	Dapo	ash
	Jalan	daan	Dalan	road
	Putih	puti	Puraw	white
	Hitam	itim	-	black
	Tahun	taon	Tawen	year
	baharu	-	Baharo	new
	Penuh	puno	napno	full
	Lama	luma	-	old
	Terus	-	Natarus	straight
	tajam	-	Natajam	sharp
	Busuk	bulok	-	rotten
	Basah	basa	Basa	wet
	Kanan	kanan	Kannawan	right
	Bersama	kasama	-	together
	Jauh	-	Dayuh	far

	Di	-	Diay	at
	Nama	-	Nagan	name
	Habuk/de	-	Tapok	dust
	bu	-	Lawa	wide
	Luas	pandak	-	short
	Pendek	asawa	asawa	husband
	Suami			
Total		59	65	

5. DAFTAR PUSTAKA

Arizo, C. D., Palayon, A. L., Tornito, A. P. V., & Sukma, B. P. (2020). Comparative Analysis of Filipino and Indonesian Monophthongs. *Lexeme : Journal of Linguistics and Applied Linguistics*, *2*(2), 59. https://doi.org/10.32493/ljlal.v2i2.8103

Blust, R. (2010). Austronesian languages. In *A Survey of Word Accentual Patterns in the Languages of the World*. https://doi.org/10.1515/9783110198966.2.691

Burke, P. (2015). *Sejarah dan Teori Sosial* (2nd ed.). Yayasan Pustaka Obor Indonesia.

Churmatin Nasoichah, Dwi Widayati, & Mulyadi. (2020). Jejak Bahasa Proto Austronesia dan Keberlanjutan Bahasa Melayu Kuno pada Penulisan Prasasti Sitopayan 1 dan Sitopayan 2. *Talenta Conference Series: Local Wisdom, Social, and Arts (LWSA)*, *3*(3), 143–147. https://doi.org/10.32734/lwsa.v3i2.879

Collins, J. T. (2005). *Bahasa Melayu Bahasa DUnia*. Yayasan Obor.

Dardanila, ., Mulyadi, ., & Tantawi, I. (2020). *Lexicostatistics of Gayo Language with Mandailing Language*. *August*, 1199–1203. https://doi.org/10.5220/0010069511991203

Istiqamah. (2017). KEKERABATAN BAHASA ACEH DENGAN BAHASA MELAYU (MALAYSIA): KAJIAN LINGUISTIK KOMPARATIF. *Seminar Nasional II USM 2017*, *1*, 222–226.

Nelson, H. (2001). Lexicostatistics applied to the historical development of three languages of the Philippines. *Deseret Language and Lingustic Society Symposium*, *27*(1).

Nurhayati. (2017). Kinship of Indonesia Language and Bugis Language (Lexicostatistics Analysis). *International Journal of Science and Research (IJSR)*, *6*(11), 1082–1086. https://doi.org/10.21275/ART20177991

Putrayasa, I. G. N. K. (2018). sejarah bahasa indonesia - Universitas Udayana. *Sejarah Bahasa Indonesia*, *Jurnal Uda*, 7–8. https://simdos.unud.ac.id

Sofiyatunnida, & Hendrokumoro. (2021). Leksikostatistik Bahasa Mandailing dan Bahasa Melayu. *NUSA*, *16*(2), 165–180.

Tamayao, A. I. (n.d.). *The Role of Linguistic Capital in Filipino Ethnic Intermarriage and Identity : A Bourdieuan Analysis*. 35–53.

Collins, J. T. (2005). Bahasa Melayu bahasa dunia: Sejarah singkat. Yayasan Obor Indonesia.

Burke, P. (2001). Sejarah dan teorisosial. Yayasan Pustaka Obor Indonesia.