
LEXICOSTATISTICS STUDY OF MANDAILING AND ANGKOLA LANGUAGES

Harianto II*, Zulfitri, Teguh Satria Amin

Universitas Muslim Nusantara Al-Washliyah, Medan, Indonesia

*Corresponding Author: hariantoi@umnaw.ac.id

ABSTRACT

This research was conducted to discuss the relationship between the Mandailing and Angkola languages by studying the lexicostatistics side to discuss the degree of kinship of the two regional languages in North Sumatra. This research makes comparisons based on 200 Morris Swadesh vocabulary which is the main reference in research and studying the relationship between the Mandailing language and the Angkola language. The purpose of this study was to determine the percentage of kinship between the Mandailing and Angkola languages in North Sumatra. Then determine the time of separation between the two languages with the study of the lexicostatistics. And also explains the sound correspondence between the Mandailing and Angkola languages. The research method used in this study is a qualitative method and also a quantitative method. From the results of this study it was explained that the Mandailing and Angkola languages have a fairly close kinship side. Especially if you look at the similarity comparison of the number of words consisting of 98 words. Then the sound correspondence itself consists of, 1) replacement of sounds or phonemes consisting of 10 words, 2) reduction of sounds or phonemes with a total of about 16 words and also 3) addition of sounds or phonemes with a total of 18 words. The kinship relationship that occurs between the Mandailing and Angkola languages accounts for around 71% of the 200 Glossary of Morris Swadesh's vocabulary obtained from the two languages. The year of separation of these two languages, namely Mandailing and Angkola, was around 1,235 years ago, starting from 2023. The two languages are in the same sub-family and are in the same family.

Keywords: Lexicostatistics, Mandailing Language, Angkola

Submitted Jan 03, 2021 | Revised Mar 19, 2021 | Accepted Mar 31, 2021

INTRODUCTION

In a developmental process in terms of linguistics, of course it cannot be separated from the kinship factor or close relationship between the two languages. This kinship relationship, which is also a family relationship, is one of the relationships on the human side that has the same origin and term. Both in terms of biological descent as well as social and cultural. This is stated by Ermanto (2020: 15). The kinship relationship of language itself has the meaning of a group of languages that are in a linguistic group that presents the historical development of that language. So that these languages can be concluded to have a kinship or kinship relationship and also the history of the development of languages that come from the same family or family.

According to Syaf'i (2019: 6), languages that have a kinship side and also the same relationship come from the same proto-language. In the proto-language there will be several branches which will later become various languages that exist around the human environment in which they live. And the branches of these proto-languages will branch back into child languages which are referred to by experts in the field of linguistics as the classification of languages into various parts.

There are around 14 major categories of languages in total. The 14 language categories are divided into 1) Niger – Cordofania, 2) Austronesian, 3) Trans – New Guinea, 4) Indo – European, 5) Aphroasiatic, 6) Sino – Tibetan, 7) Australian, 8) Nilo – Shara, 9) Oto – Manguea, 10) Austro – Asiatic, 11) Sepi – Ramu, 12) Dravidian, 13) Tai – Kadai, and 14) Tupi. According to Syaf'i's research (2019: 15) that the languages in the archipelago are included in the proto-language category of Austronesian.

The Austronesian language itself has 5 parts consisting of the West, Northwest, North and East as well as the central part. For each of these sections there are areas that have their own language structure. As in the western region which includes Indonesia and Sarawak and Madagascar. Then for

the northwestern region consisting of Taiwan, the Philippines and also North Kalimantan and Brunei. The North and East regions will include Micronesia and Polynesia. Then the central region consists of East Irian or Papua and also Melanesia. The division of the regional categories is stated in Keraf's thinking (1996: 192).

Meanwhile, according to the thoughts of Dyen in Parera (1991: 15) states that the Archipelago has a language division which is divided into 3 categories, namely the Nusantara language which is in the western part or also called Hesperonesia, the Maluku group category which consists of the Maluku, Sumba and also the Maluku groups. East Timor, as well as the Polynesian and Micronesian groups which are in 1 linguistic group in the archipelago.

The archipelago language group in the western part belongs to the Hesperonesia group where Andalas or Sumatra Island has various ethnic groups and various languages. Northern Sumatra itself has the dominance of the Malay tribe and also a variety of languages. The dominance of the Batak language is one of the languages that can be found in North Sumatra. The Bukit Barisan area became the initial base for the Batak tribes and languages that spread there. Then the Batak language began to spread and develop around North Sumatra and was divided into various languages and dialects according to the environment of the local community.

Each tribe has its own language with different accents and articulations, although there are some linguistic similarities. Among the Batak languages that are spread in North Sumatra, namely Mandailing and also Angkola, which are one of the Austronesian language families spoken by people in North Sumatra.

According to KBBI (2008), kinship in linguistics has a meaning as a fairly close relationship between 2 or more languages that originate from the same language structure and source. And related languages mean that the language has a genealogical relationship with other languages. The language is in 1 sub-family or the same family so that it can have its own kinship side. The level of kinship in the language does not necessarily have kinship and similarities in the language.

The degree of closeness of the language relationship can of course be seen by the similarities in vocabulary and also the differences in the words contained in the structure of the language if the two or more languages are compared. If these languages are juxtaposed and have similar vocabulary and pronunciation, then the level of kinship between these languages is quite high. And vice versa if the vocabulary presented in the word structure of the language has differences, then the kinship relationship will be quite tenuous.

Based on the Kridalaksana Linguistic Dictionary which states that the Batak language including Mandailing and also Angkola has a designation as the Batak Dialect with the Batak language which has many implied variations of the language. The Batak languages in the Sumatra region consist of Karo/Alas/Pakpak, Simalungun, Toba, Mandailing and also Angkola languages.

And this division is supported by research from the Medan Language Center which conducted a mapping of all languages throughout the archipelago including North Sumatra which was carried out with dialectological calculations which is one way to categorize the Batak language and also its variants into these dialect forms. The data used in the language mapping was obtained in 1990 which has the possibility that there will be changes in every word in the language with a span of 10 years. And the fact is of course the words in the language are no longer the same.

The Batak languages that have variants in the Batak language group mean that the Batak people will think that their tribe is different and also stands alone. Likewise with people who use the Mandailing and Angkola languages who think that they are a different society and are not part of the Batak community. Likewise with Simalungun, Karo and Toba who stated this. The fact that these languages may be related at the language level should be investigated further.

Lexicostatistics is a technique that will make it easier for humans to determine the level of kinship of a language using fairly easy steps. This is done by comparing the vocabulary of the two languages or more which are allegedly related. This is stated in the narrative of Crowley (1992: 168). This can find out how far the kinship of the language studied is.

Crowley (1987: 191 – 192) states that this Lexicostatistical method is used using basic assumptions which consist of 2 parts. The first assumption is that the vocabulary of the language will be difficult to change compared to other parts. It is stated that the basic vocabulary becomes the core vocabulary for identifying the level of kinship because words are used quite often in people's lives,

especially in the set of languages which are an important element in life and death, a language that is also spoken by Keraf (1991: 123).

Changes in the replacement of words with non-relative words caused by the original form will change the meaning if it refers to something else. Or it could also be that the word is borrowed from another language to express certain meanings in that language. Then the second assumption itself states that changes to the basic vocabulary in all languages will have the same nature. This assumption has been made in 13 languages that have written scripts.

The results show that every thousand years will display vocabulary that survives, especially in the basic vocabulary between 74.4% to 86.4%. And the average is at 80.5%. And this cannot be interpreted that all languages will have the same percentage according to this average number. This is of course due to the experimentation of the language, namely Indo-European which does have differences.

If the second assumption is accepted, then the retention of the average basic vocabulary of that language can be expressed by the formula $80.5\% \times N$. Which means that the symbol N is the number of basic vocabulary that existed at the beginning of multiples of 1 millennium or one thousand years. Two hundred basic vocabulary according to Swadesh in a particular language, over the first 1,000 years, you will get, $80.5\% \times 200$ which is 161 words. Then after the second 1,000 years, the remaining words are $80.5\% \times 161$ which means 140 words have been made. So it continues after going through a period of 1,000 years.

According to Mahsun (1995: 115), Lexicostatistics is a method of grouping languages that will calculate the percentage of sets of relatives or cognates of that language. And in the lexicostatistics calculation there will be words that have similarities in the phonetic sector and also in morphology which will be referred to as related words and also referred to as *cognate* or cognate. From these related words, a separation time calculation will be carried out which comes from the proto language or also known as the age of the language.

Based on research from Keraf (1984: 121) that this lexicostatistics is a technique that will group languages into statistical observation of words. And this will make it easier to define these groupings on the basis of the percentage of similarities as well as differences between the languages studied. From this narrative, the working methods of statistical lexicon are described which are the reference and guidelines for conducting research on the kinship side of the language sector. In lexicostatistics, there are sub-group levels which are divided into the following:

1. Subgroup : 81% - 100%
2. Family/ family : 36% – 81%
3. Family/ stock : 12% - 36%
4. Microphyllum : 4% - 12%
5. Mesophyllum : 1% - 4%
6. Macrophyllum : 0% - 1%

In the classification in the lexicostatistics sector, the similarity at the 81% to 100% level will be referred to as language. Meanwhile, for the lower level, namely 36% to 81%, this is also called a family. Then at the 12% - 36% level this is referred to as a clump. Then at the microphyllum level it has a percentage of between 4% and 12% and also the mesophyllum at the level of 1% to 4%. And the last one is macrophyllum at the 0% to 1% level.

Based on Keraf's thoughts (1991: 127 – 130) it is stated that in the process of word comparison to determine basic words that are related and also non-relative, this will assume that proto-language phonemes that have developments in these related languages will develop consistently. in the linguistic environment of each of these related languages. In this comparison, the phonemes in the relative position will be compared with each other. While the genetic relationship is mentioned, this phoneme pair appears in many other pairs. The pair that appears in this relationship is a reflection of a phoneme and allophone in the proto-language. This is stated in Crowley.

Language migration is a way to collect data, limitations, assumptions and hypotheses that discuss the movement and direction of national migration in the prehistoric era. The purpose of knowing about this migration is to recognize and research the origins of languages that migrate to an area. And of course the distribution area will also be known by researching the migration of these languages.

Based on the thoughts of Keraf (1984: 172) says that the area of a language will be identified with the area of its speakers. There are 2 terms in the migration theory which consist of region and

region. A language area is a place where there are speakers of a language which also consists of a continuous geographical place and there are also a number of areas that are geofreely separated from other places.

For areas that are geographically separated from one another, but are still inhabited by speakers of that language, they are referred to as language areas. And this language area can occur due to speakers of that language moving to other areas. And this migration theory is carried out with 2 propositions which consist of the region of origin of a relative language from a continuous area. This is the basis for determining that a region has unity with other areas determined by a separate language.

While the second proposition is to determine the number of migrations that are reconstructed and will be inversely proportional to the number of migrations that occurred in each language. This postulate means that the movement of the language is minimal if the number of moves in the 2 migrations has a large enough chance of being the migration that has occurred.

The Batak language itself has several levels of kinship that are not the same between one and the other. Likewise with the separation time. The Toba Batak language as well as Angkola and other Batak languages have different levels of kinship. The Mandailing and Angkola languages are still very closely related and both have related dialects. The Mandiling and Angkola languages are in the same group and also in the same family. So that the level of kinship between the two languages is quite close with a time of separation of more than 1,000 years.

METHODS

This research uses qualitative and quantitative systems and methods. In order to study language comparisons in a quantitative way, the researchers will use the percentage of kinship and also the language categories of the Mandailing and Angkola languages. And both of these languages will use the percentage of related words in the research. From the percentage of related words, namely from the Mandailing and Angkola languages, the researcher can get the length of time and also the year of separation from the Mandailing and Angkola languages.

Meanwhile, in order to examine the language comparison between Mandailing and Angkola using this qualitative method, evidence of phoneme or sound correspondences contained in the basic vocabulary will be used as a reference for categorizing the Mandailing and Angkola languages as the research material. In addition, the lexicostatistics technique is the main material in this study with the addition of glotochronology techniques as auxiliary techniques.

The Mandailing and Angkola languages, which are in their respective areas and language areas, are the objects of research in the linguistic sector. The data contained in this study and also used as a reference are 200 Swades vocabularies that are usually spoken by the respondents who are the subjects of the research. The research location was in North Sumatra with respondents and informants who were accustomed to using the Mandailing and Angkola languages in their daily lives.

According to Mahsun (2007: 105 – 106) states that people or people who can become respondents or informants are middle-aged, namely between the ages of 20 to 40 years, which is assumed that they have mastered the language and dialect perfectly. And of course not to the level of senility for old age. The next requirement for respondents or informants is to have complete articulators or speech tools. Which means the number of teeth is intact, correct and also valid in the articulation of the language and also the dialect.

The third requirement is having sufficient education. Informants as well as respondents do not need to be highly educated and also not too low. At least have graduated from elementary school to become a respondent or informant. The fourth requirement is that the people who become informants and also respondents are of course not illiterate. This is because the interview process and also the questionnaire conducted by the researcher certainly had to make the respondents write down the words they said and also some instructions from the researchers. And this is to ensure that the respondent masters what is given by the researcher.

Informants as well as respondents are products of indigenous or indigenous people in the area who have been in the area for at least 2 generations before. This is to make the recorded data more valid in order to examine the level of lexicostatistical studies of the Mandailing and Angkola languages. The next requirement is that informants or respondents can master the language purely without any influence from languages or dialects from outside the Mandailing and Angkola languages. This is closely related to the purity of the pronunciation and also the dialect of the community.

Mobility of informants out of the area is quite small. This is of course to maintain the level of purity of the Mandailing language and also the Angkola. Informants and respondents have never traveled to other areas for a long time. This is so that the pronunciation of the Mandailing and Angkola dialects and languages is not affected by other dialects and languages.

Informants who have been out of their area for a long time will certainly be quite affected by the other languages they receive. This of course affects the level of purity of the spoken vocabulary and also the dialect of the informant towards Mandailing and Angkola languages.

Based on Sudaryanti's narrative in Muhammad (2011: 212) states that data obtained from a study can use data collection techniques consisting of 1) fishing techniques where the researcher will provoke the conversation first so that the desired data will come out from the mouth of the informant and also the respondent so that researchers can immediately get the data. 2) the face-to-face technique is a technique that allows researchers to ask informants directly about the desired research object. 3) the recording technique is a way that can be done by researchers to listen back to informants' answers properly. 4) the next technique is the recording technique which is the answer from the informant or respondent heard by the researcher and also recorded as a data transcription.

technique is a technique for validating data carried out in this study. According to Moleong (2007: 330), this technique is a technique that checks the validity of data by utilizing something that is outside the data to check recorded and existing data. This technique is usually carried out by conducting interviews with informants where if 2 of the 3 informants have the same answer, then the answer is a valid answer.

The data obtained from Swadesh's 200 basic vocabularies from the Mandailing and Angkola languages is entered into the data analysis stage which can be done by: 1) calculating related vocabulary, 2) determining the percentage of relatedness levels based on Crowley's formula in Ermanto (2020: 71) with the formula, namely:

$$\text{Percentage of kinship} = \frac{\text{number of related vocabulary}}{\text{Number of basic vocabulary (200)}} \times 100\%$$

3) calculation of the separation time between the two languages, in this case, namely Mandailing and also Angkola with the following formula:

$$t = \frac{\log C}{2 \log r}$$

The lexicostatistics technique is the main technique in the research process. And this technique has several ways and steps to obtain data which will then be analyzed in the research process. Some of these steps can be done by:

1. Register the words found in the research

The data obtained in the research will be entered into the language table which can later be used as a reference. As an example of respondents from the Angkola language, some data will be obtained from the vocabulary used. And this word will later be entered into the table so that it is chosen which can be material for research on the data.

2. Selecting words as reference material for research data from each language to be analyzed

From the words that have been registered, words that will be used as further research data will be selected. The choice of this word can of course be done with various kinds of considerations that these words have similarities and similarities between the three languages that are juxtaposed in each of these languages.

3. Determining words that have kinship and also analyzed

In this step, words that have a kinship side between the Mandailing and Angkola languages will be determined based on similarities, similarities in sound and dialects that present distinct similarities.

4. Calculate the degree of kinship of the language.

The next step is to calculate the percentage level of related words according to Keraf's formula (1984: 127), namely:

$$C = \frac{K}{X} \times 100\%$$

C = Cognates or words that have a degree of kinship

K = Number of vocabularies that are related between the two languages

G = gloss or vocabulary from Swadesh

5. Determine the proto phoneme

This method is of course to determine from the proto-dialect that corresponds to the proto-phoneme for each gloss. Of the 200 Swadesh vocabularies, elements of the proto-phoneme will be found first and will be calculated from the dialect, which can later produce the most data to become the proto-phoneme and can also become the proto-dialect.

6. Calculation of split time between 2 languages

The separation time between 2 or more languages can be determined by calculating the percentage of words that are related. Crowley (1992: 178) has a formula that Keraf (1984: 130) agrees with.

7. Calculates the age of the language

In order to calculate the error term, standard error will be used which means 70% of the estimated truth. Of the standard errors, we will use calculations from Kerat (1984: 132) in the following way:

$$s = \sqrt{\frac{c(1-c)}{n}}$$

S = standard error in the percentage of related words

C = Percentage of words that have a kinship side between the following words

n = the number of words compared to those who are related and also not related at all.

The calculation can be done in the following way

- a. 1 minus C
- b. C will be multiplied by the result of (a)
- c. The result of (b) will be divided by n.
- d. Draw the upper root of the calculation result ©
- e. And the result of (d) is the error term and also the percentage of related words on the basis of 0.7 estimates of the actual data.

By calculating the results in point (d), of course, the calculation of the standard error in years is carried out in the following ways:

- a. The error range of the percentage of relatives in point (d) is added with C
- b. The number in point (a) is treated as C or the percentage of words that are the most recent relative with the time calculation formula.
- c. The formula obtained in point (b) will be reduced by the first number of times. And this number will be added and subtracted by the first number to get a basic error period of 0.7 from the actual conditions.

The results are added up with the percentage of related words to get a new value of C so that the separation time will be calculated again using the separation time formula. And the results of this calculation can be seen regarding the age of the language and the time of separation.

8. Define language migration

In order to account for language migration, a distribution unit will be used. If this related language spreads and people can move directly to an area, then the language will form a chain that becomes 1 link in the chain. And in 1 link there will be at least 2 languages. To determine your own language migration using several ways such as:

- a. The starting point for migration is by looking at the proto-dialect of a language by looking at the number of proto-phonemes in that language. The larger the phoneme of the language, it can determine the starting point of the migration.

- b. This known starting point will determine the continuation of kinship relations between dialects with lexicostatistical techniques where kinship relations have become provisions based on Mahsun (2012). The higher the percentage of kinship, the direction of early-stage migration. However, the lower the percentage of kinship, the next migration direction.
- c. Dialect migration can also be seen by the order of the age of the dialect and also by the percentage of kinship. Language age is also one of the things that can be seen regarding sub-dialect migration with the provision that the younger the sub-dialect is, the closer the relationship between the language and the direction of migration will be. And the older the sub dialect is, the language will be further away and also the direction of migration will also be further away.

RESULTS AND DISCUSSION

The data taken in this research method is data based on 200 vocabulary sources derived from the basic words of Morris Swadesh. And this data will be analyzed in order to get the level of kinship based on the percentage of the language. In this case, namely the Mandailing and Angkola languages. Separation time, evidence of sound correspondence from the Mandailing and Angkola languages is also one of the things analyzed.

Before calculating the percentage of kinship and also the differences in these languages, a cognitive vocabulary calculation will be carried out in these languages. In order to calculate this cognate language, the researchers used marker codes, namely (+) and (-) which means that if the word has a cognate or degree of kinship, it will be marked (+). Meanwhile, if there is no cognate or degree of kinship between the two languages, a sign (-) will be indicated. In order to get the calculation of the number of words, it is done by looking at the similarity of markers between words that are related between the two languages.

Of the 200 Swadesh vocabularies, it can be obtained, classified and concluded, 142 vocabularies that are related to Mandailing and Angkola languages and 58 languages that do not have any level of kinship. From calculations using the lexicostatistics technique obtained from Swadesh's 200 basic vocabularies, the kinship between the Mandailing and Angkola languages can be determined at the percentage level using the formula:

$$\text{Percentage of kinship level} = \frac{\text{Amount of kinship vocabulary}}{\text{Number of basic vocabulary (200)}} \times 100\%$$

$$\begin{aligned} \text{Percentage of kinship level} &= \frac{142}{200} \times 100\% \\ &= 0.71 \times 100\% = 71\% \end{aligned}$$

From the calculation data, the percentage of both languages is 71%. And determining the language level from calculating the percentage using this lexicostatistics technique when juxtaposed with references based on narratives from Crowley in Farida (2013), it is found that Mandailing and Angkola languages are languages that come from sub families. Reference categories of Isolect or language groupings based on percentages are:

1. Dialect of one language : 81% - 100%
2. Language from 1 sub family : 55% - 80%
3. Sub-family of 1 family : 28% - 54%
4. Families of 1 offspring : 13% - 27%
5. Derivative of 1 pilum : 5% - 12%

According to Crowley, from this grouping and category, the kinship level and kinship of the Mandailing and Angkola languages in North Sumatra are found in 1 sub-family. And after getting the

level of grouping, you can also get the percentage of cognates that determine the separation time between the 2 languages using the glotochronology technique with a formula derived from Crowley.

$$t = \frac{\log C}{2 \log r}$$

From the calculation of the separation time between the two languages, it can be described with data information, $C = 71\%$ and $r = 80.5$. So if you put it in the formula you will get it

$$t = \frac{\log 71}{2 \log 80.5}$$

$$t = \frac{342}{2 \times 217}$$

$$t = 788 \text{ years}$$

From calculations using the glotochronological technique, the separation time between the Mandailing and Angolan languages, which are in North Sumatra, will be 788 years from the current calculation year, namely 2023. Then 1,235 AD will be the separation time for the two languages so that now it still has sides relationship between the two languages.

Cognates or related words are pairs of words that are quite well related in terms of phoneme forms and meanings that have exactly the same content. This vocabulary similarity is viewed from the dialect, articulation, writing and also the pronunciation of the vocabulary which does not experience a change in pairing between the two languages, namely Mandailing and Angkola. And these are about 98 similar pairs of forms.

For example, in Swadesh's basic glossary, namely water, in the Mandailing and Angkola languages it is also referred to as aer or written as ae?. Meanwhile, for the word flow or flow, in the Mandailing and Angkola languages it is called manalir. Likewise for the word child referred to as ana? In the Mandailing and also Angkola languages. In line with the word wind and dog and fire. The Mandailing and Angkola languages have the same spelling and writing of the vocabulary.

While the vocabulary corresponds to sound, the data obtained and researched has variations in the type of change. As an example of the word "smoke". In the Mandailing and Angkola languages, the writing becomes "Timbus - Timus". Then in the word "see," the Mandailing and Angkola languages refer to it as "digin" and also "ligin." While for the word "name," in the Mandailing language it is referred to as "goar" and also "golar."

Quantitative and qualitative comparisons obtained from this study, the researchers determined the percentage of kinship and also the classification of the Mandailing and Angkola languages according to the percentage of relatives or cognates. From this data and research, it is also possible to calculate the length of time of separation and also the year of separation of the Mandailing and Angkola language kinship. While studying comparative language qualitatively, the researchers were able to establish evidence of sound and phoneme correspondence in the basic vocabulary used for the classification of the Mandailing and Angkola languages included in the study. Using the lexicostatistical technique and also the glotochronological technique, the relationship between these two languages is obtained which turns out to be the language of one sub-family.

The kinship of a language can be done using various methods and methods as well as reviews. And the lexicostatistical method which uses 200 basic vocabulary from Morris Swadesh, is the technique used in this study. The basic vocabulary will be spoken by the respondents from the language to be studied. Each respondent and informant will speak 200 vocabularies according to their language based on Swadesh's basic vocabulary. The Mandailing language will be spoken by 5 respondents or informants. Likewise with the Angkola language spoken by 5 respondents.

From the calculation of the lexicostatistical technique of the Mandailing and Angkola languages, as mentioned earlier, it can be concluded that there are 142 cognate words with a total of 71% of the kinship as seen from the classification of the percentage of cognate level and kinship. Meanwhile, if you look at the calculation of the glotochronology technique, you will get the separation time between

the Mandailing and Angkola languages, namely 1,235 AD, which is calculated from the current year, which is 2023.

The number of language levels that have this kinship can be determined by the high value of kinship between the Mandailing and Angkola languages. These two languages have similarities in pronunciation in the spoken and written sectors, although there are differences in dialects and accents. The similarity of phoneme forms is one proof that words are related if the similarity in form of 98 vocabulary and sound correspondences is divided into 1) sound and phoneme changes up to 10 vocabularies, 2) reduction of sounds and phonemes of around 16 vocabularies, 3) addition of sounds and phonemes up to 18 vocabularies.

The Mandailing and Angkola languages are languages that are related to each other and also come from 1 sub-family. As for tradition, the Mandailing and Angkola Batak people also have the same tradition and are also familiar with the clan system, which is a fairly strong patrilineal system among North Sumatra, especially in Mandailing and Angkola. The migration process is one aspect of the kinship between the two languages.

CONCLUSION

From the results of the analysis and discussion of the lexicostatistics study of the Mandailing and Angkola languages in North Sumatra, it was found and also concluded that according to the lexicostatistics technique, out of 200 basic Swadesh vocabulary in the Mandailing and Angkola languages in North Sumatra, 142 words were found which are related to the two languages. . Then there are 58 vocabularies that are not related.

By using lexicostatistical techniques as well as glotochronological techniques as well as qualitative and quantitative methods, information is obtained about the kinship of the Mandailing and Angkola languages. For the percentage of kinship of the two languages, namely Mandailing and Angkola, it is worth 71% which indicates that there is a kinship relationship between the two languages. The relationship between the Mandailing and Angkola languages is concluded as a relationship between languages that are in the same sub-family. This is due to language migration and other influences.

According to glotochronological calculations, between the Mandailing and Angkola languages, there is a gap of about 788 years between the two languages. And the separation time of the Mandailing and Angkola languages was 1,235 AD last year, and was calculated since 2023. These two languages are in the same sub-family, and have been separated since 788 years which is a sign of a kinship between the two languages, namely Mandailing and Angkola.

After the research conducted, there are various kinds of evidence that can help the conclusions of the research. The correspondence of sounds or phonemes in the Mandailing and Angkola languages, from the 200 basic vocabulary based on the Swadesh reference, can be categorized into several classifications. The similarity of word forms between the Mandailing and Angkola languages includes around 98 vocabularies. And the sound correspondence is divided into:

1. Alternation of sounds or phonemes totaling 10 vocabulary words
2. Reduction of sounds or phonemes totaling 16 vocabularies
3. The addition of sounds or phonemes totaling 18 vocabularies

This research is quite important to do because it is to check and also extract information about the level of language kinship in North Sumatra, especially the Mandailing and Angkola languages. The level of kinship of regional languages in Indonesia has not been studied much regarding kinship and also family relations in the linguistic sector between one language and another. Especially for languages that are in the area and also adjacent areas. Even many people who do not know about the level of kinship. Differences in race and ethnicity in society have indeed made regional languages a means of communication for the local community. However, this language certainly does not stand alone and is different from other languages in the areas near where the community lives. There is a kinship side between the regional languages.

Language migration is one of the links that can connect the kinship side between regional languages in Indonesia. This language migration involves the Mandailing and Angkola languages which serve as the origin of the migration. This migratory movement created synergy between these languages and also created a kinship between the two regional languages in North Sumatra, which has a fairly high degree of kinship.

To find out that the regional languages in Indonesia are related, of course research must be carried out to find out the level of kinship, the time of separation from these languages. It is hoped that this research can provide and improve the integrity of society in Indonesia, especially in the language sector, so that people in the archipelago know their language and also their previous ancestors who already had kinship in the linguistic sector and have survived to this day.

Of course research in conducting linguistic studies on regional languages must be carried out frequently. And of course other researchers can take a reference to conduct a more in-depth study of the language and also the cultural vocabulary in each region, in order to get the kinship side between the languages in the archipelago. Do not forget to also conduct studies on the historical side of population migration in the archipelago. This is because population migration is not only related to culture and tradition, but in the language sector there is also a strong role for migration from society. With this migration, of course, the vocabulary of the language will change once every 1,000 years.

REFERENCES

- Azhar, Ridwan. 1989. *History of Language Studies in Indonesia*. Medan: USU Faculty of Letters
- Blust and Collins. 1985. *Comparative Study of the Languages of the Western Archipelago*. ILDEP. Jakarta: Djbatan.
- Crowley, T. 1992. *An Introduction to Historical Linguistics*. Oxford: Oxford University Press.
- Crowley, Terry. 1987. *An Introduction to Historical Linguistics*. Papua New Guinea: University of Papua New Guinea Press.
- Ermanto. 2020. *Comparative Historical Linguistics: Theory and Practice of Determining Language Kinship in the World*. Depok: PT Raja Grafindo Persada.
- Fernandes, Inyo Yos. 1994. *Comparative Historical Linguistics: An Introduction to Field Theory*. Yogyakarta: Unpublished Book
- Halima, Amran. 1988. *National Language Politics I*. Jakarta: Balai Pustaka.
- Indrani, Ika. 2007. *Lexicostatistics of Toba Batak language with Pakpak Dairi language*. Thesis.
- Hutabarat, Farida Meliana. 2013. "Toba Batak Language Kinship with Mandailing Batak Language". *Journal of Language and Literature*, 2(1): 59- 71.
- Keraf, Gorys. 1984. *Historical Comparative Linguistics*. Jakarta: Gramedia Pustaka Utama.
- Keraf, Gorys. 1991. *Typological Comparative Linguistics*. Jakarta: Gramedia Pustaka Utama.
- Keraf, Gorys. 1996. *Comparative Historical Linguistics*. Yogyakarta: Gadjah Mada University.
- Parera, Jose Daniel. 1991. *Comparative Historical General Linguistic Studies*. And Jakarta's Structural Typology: Erlangga Publisher.
- Kridalaksana, Harimurti. 1963. "Lexicostatistical calculations of the eight languages of the Western Archipelago and the determination of the distribution centers of these languages based on the theory of migration". in the *Literary Sciences Magazine* 1963-1973. Jakarta.
- Kridalaksana, Harimurti. 2008. *Linguistic Dictionary*. Jakarta: PT Gramedia Pustaka Utama.
- Syafi, Imam. 2019. "Lexicostatistics of Five Archipelago Languages: Javanese, Madurese, Sundanese, Balinese, and Madurese". *Journal of language studies, Indonesian Literature, and Learning*, 3(1).
- Mahsun. 2007. *Language Research Methods*. Jakarta: PT Raja Grafindo Persada.
- Mahsun. 1995. *Dialectology Diachronic: An Introduction*. Yogyakarta: Gadjah Mada University Press.
- Mahsun. 2005. *Language Research Methods*. Jakarta: Raja Grafindo Persada.
- Masyuri and Zainuddin. 2008. *Research Methodology: Practical and Applicative Approach*. Malang: Refika Aditama.
- Mbete, Aron Meko. 2002. *Diachronic Linguistic Method*. Bali: Udayana University.
- Moleong, Lexi J. 2007. *Qualitative Research Methodology*. Revised Edition. Bandung: PT Juvenile Rosdakarya.
- Parera, Jose Daniel. 1991. *Comparative Historical General Linguistic Studies*. Jakarta: Erlangga. (Digitalized April 25, 2008.)
- Sudaryanto. 1986. *Linguistic Method 1: In the Direction of Understanding the Linguistic Method*. Yogyakarta: Gadjah Mada University Press.

- Sudaryanto. 1988. *Linguistic Method 2: Methods and Various Data Collection Techniques*. Yogyakarta: Gadjah Mada University Press.
- Sudaryanto. 1993. *Methods and Various Techniques of Language Analysis*. Yogyakarta: Sanata Dharma University Press.
- Sugiyono. 2008. *Quantitative, Qualitative Research Methods, and R&D*. Bandung: Alfabet.