Quest Journals
Journal of Research in Humanities and Social Science
Volume 10 ~ Issue 1 (2022)pp: 01-10
ISSN(Online):2321-9467
www.questjournals.org

Research Paper

Innovation of Protophonems of Bataknese Languages and Their Subgroups

Himpun Pangggabean
Department of English Language and Literature,Universitas Methodist Indonesia

ABSTRACT
*This research deals with proto-phonemes, phonemic correspondence sets, rules of sound change, cognate levels, and subgroups of five sister languages in Batak languages (BLs) comprising Toba language (TL), Angkola language (AL), Simalungun language (SL), Karo language (KL), and Dairi language (DL) spoken in North Sumatera, Indonesia. The method of the research is comparing variants of sounds stemming from innovation of proto-phonemes of BLs or p(BLs) via reconstruction. It is revealed that there are eight phonemic correspondence sets in BLs namely /tt-nt-nt-nt-nt/ of which proto-phoneme is /*nt/, /Ø-Ø-h-h-h/ of which proto-phoneme is /*h/, /b-b-b-mb-mb/ of which proto-phoneme is /*b/, /ɔ-ɔ-ə-ə-ə/ of which proto-phoneme is /*ə/, /ɛ-ɛ-ei-ɛ-ɛ/ of which proto-phonemes is /*ɛ/, /k-k-g-k-k/ of which proto-phoneme is /*k/, /i-i-i-i-e/ of which proto-phoneme is /*i/, and /d-d-nd-nd-nd/ of which proto-phoneme is /*nd/. The subgroups of BLs are BT and AL belong to one subgroup, BK and BD belong to one group and BS is separated from the two subgroups.*
***KEYWORDS**: Proto-phonemes, Reconstruction, Innovation, Comparative, Subgroups*

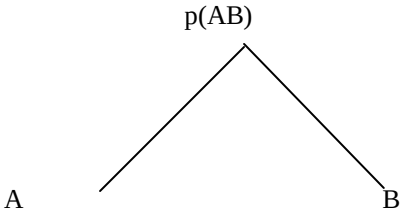
***R**eceived 10 Jan, 2022; **R**evised 24 Jan, 2022; **A**ccepted 25 Jan, 2022 © The author(s) 2022.*
P**ublished with open access at **www.questjournals.org

I. INTRODUCTION

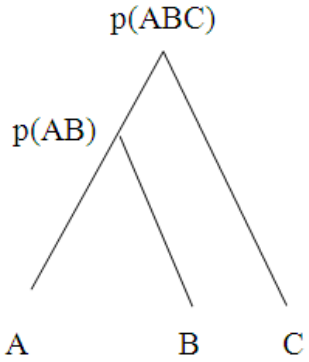
In this globalization era, many regional languages all over the world, including those of Indonesia's are prone to extinction. Patji (2011) suggests that 169 regional languages are threatened to be dead. The phenomenon prompts me to do research into *Bataknese* languages (BLs) as parts of around 750 regional languages of Indonesia spoken in North Sumatera, Indonesia, to help preserve the languages. One of the ways of preserving languages is to do research into them in terms of all aspects of linguistics.

The issues being dealt with in this research are proto-phonemes, phonemic correspondence sets, reconstruction, and subgroups of BLs which are the areas of Histrorical Linguistics or Diachronic Linguistics suggesting that sister languages are inherited from the same language called protolanguage. The principle underlying the issues is language changes gradually and regularly resulting in variants of languages.

Some sounds are lost due to innovation of protolanguage. Crowly (1992) uses symbol /Ø/ for lost sound, Lehman (1972) uses asterisk (*) for protolanguage or proto-phoneme, and Hyme (1960) uses p(AB) for protolanguage of A and B. These three symbols are used in this research.
Crowly (1992) shows the protolanguage of sister languages by using the following diagram:



The diagram shows that p(AB) is the protolanguage of A and B. The diagram can be extended by the presence of meso language or intermediate protolanguage depending on the number of sister languages after subgrouping is determined,



This kind of diagram will be used to show the closeness between p(BLs) and its reflexes in sister languages to make subgrouping.

The changes of language are regular and can be seen by comparing the sister languages or so called Comparative Method. Schleicher (1871) in McManis et.al (1987) proposed, such changes occur regularly and recognizably and can be seen in genetically related languages called sister languages.

The regularity of the changes in BLs can be seen in the following example:

Gloss: *bamboo* ´bamboo´

TL	AL	SL	KL	DL
b	b	b	b	b
u	u	u	u	u
l	l	l	l	l
u	u	u	u	u
Ø	Ø	h	h	h

In this example, it is found out that in TL and AL, phoneme /h/ in final position is lost due to innovation of the proto-phoneme of phonemic correspondence set in sister languages whereas in each SL, KL, and DL, phoneme [h] is inherited from p(BLs) linearly or has retention in the same position. This is the evidence that the five languages are genetically related.

In the above data, there are five correspondence sets or Gudschinsky (1986) calls it cognate sets, namely /b-b-b-b-b/, /u-u-u-u-u/. /l-l-l-l-l/, /u-u-u-u-u/, and /Ø-Ø-h-h-h/. The first through the fourth correspondence sets are the result of linear inheritance and the fifth set is the result of innovation, /h/ changes into /Ø/ in the final position before vowel or /h/→/Ø/___ #.

V

To determine the proto-phonemes of the sets of correspondence, the phonemes inherited linearly, are automatically the proto-phonemes since there is no change in proto-phoneme's reflexes. Thus, the proto-phonemes of /b-b-b-b-b/ is /*b/, /u-u-u-u-u/ is /*u/, /l-l-l-l-l/ is /*l/, and /u-u-u-u-u/ is /*u/ whereas the proto-phoneme set of correspondence due to innovation is determined by the most dominant phonemes or *majority wins* meaning that the proto-phoneme of /Ø-Ø-h-h-h/ is phoneme /*h/. However, there are other criteria to determine proto-phonemes of the latter to be discussed in Literal Review.

Regarding subgrouping of BLs, the same phonemes shared by two or more languages the most show that they are closer one another in comparison to the other languages.

With the preliminary data of sets of phonemic correspondence, it is predicted that TL and AL belong to the same subgroup while BS, BK, and AL might belong to another subgroup(s). However, it should be supported by rigorous data where the same sets occur regularly to be discussed in the following part.

There are some phonemic clusters in TL are orthographically written equal to its counterparts but pronounced differently. For instance gloss *tangkap* ´catch´, orthographically is written the same, but it is pronounced differently in TL, /takkup/.

TL	AL	SL	KL	KD
takkup	taŋkup	taŋkap	taŋkap	taŋkap

In TL, /ŋk/ is pronounced with /kk/, but in AL,SL, KL, and KD, it is pronounced with /ŋk/. There are many other equal phenomena of this thing as will be seen in the Results and Discussions.

II. LITERATURE REVIEW

Phonemic Corresspondence or Cognate Sets

Crowly (1992) argues that phonemic correspondence is sounds sequence in sister languages that is reflected by protolanguage. Gudschinsky (1956) uses the term cognate set to mean the same thing by suggesting the criteria

that the sets are phonetically identical like [a]:[a] and [c]:[c]) or phonetically similar like ([p]:[b], [t]:[d], and occur regularly.

Further, Crowly says that suspicious pairs of sound correspondences can be determined by firstly, to find out sound sets that are phonetically similar, and secondly to find out whether they are distributed complementarily or contrastively.

Keraf (1991) shows the example of correspondence sets as follows:

Gloss	Greek	Latin	Sanskrit	Gotic
six	hex	sex	sas	saih
two	dýo	duo	dva	twai
ten	dėka	decem	dāsā	tăihum

There are three sets of phonemic correspondence in glosses six, two, and ten namely /h-s-s-s/, /d-d-d-t/, and /d-d-d-t/. To make sure that they are phonemic correspondence sets, they have to be strengthened by other data that can show phonemic correspondence set recurrence taking place repeatedly and regularly to prevent sounds sequences which are not actually correspondence sets from being determined as sets of phonemic correspondence and coincidence.

The example in Austronesian languages is shown below:

Melayu : [hiduŋ]
Batak : [iguŋ]
Sunda : [iruŋ]

The correspondence set in the data is /d-g-r/. This set of correspondence is predicted to occur repeatedly, regularly, and complementarily if there are adequate data to prove that it is phonemic correspondence set.

According to Crowly (1956), language change may be conditioned sound change and unconditioned sound change. Conditioned sound change occurs is due to the effect of adjacent sounds whereas unconditioned sound change is the change in initial, middle, and final positions without the effects of adjacent sounds. The changes can be seen by comparing sounds of sister languages in comparable sets. The comparison is used to conduct internal reconstruction of protolanguage.

Langacker (1972) argues that comparative method is implied by discovering sets of sound correspondence and adds that phonemic changes are systematic. Sounds that have phonemic sets of correspondence do not have to be the same but occur regularly in the same positions of words having similarity of forms and meanings.

Reconstruction

Crowly (1992) suggests that reconstruction is the estimate of protolanguage by discovering sound changes of the contemporary genetically related languages. Internal reconstruction is conducted by: 1. Excluding non-cognates from cognates, 2. Determining the phonemic correspondence sets, 3.Examining the differences of sounds in phonemic correspondence sets.

Where there is no phonetic difference between two sounds, to determine their proto-phoneme as in the following, the most likely change into another sound is the one which is more natural. In

Tonga	Samoa	Rarotong	Hawai
k	ʔ	k	ʔ

/*k/ is more likely and naturally to change into /ʔ/ rather than /*ʔ/ into /k/ thanks to lenition that is common in all languages of the world.

Meanwhile, *The Comparative Method and Linguistic Reconstruction*, retrieved from *http://en.wikipedia.org/wiki/Comparative_method* suggests that phoneme with widest distribution also called *majority wins* is determined as proto-phoneme.

Crowley (1992) in his researches into languages sum up sound changes as follows

1.	t→k	{t} becomes	{k}
2.	ŋ → Ø	{ŋ} is lost	
3.	t→ s/___ front	{t} becomes	{s} in initial position
4.	x→k/s___	{x} becomes	{k}after {s}
5.	p→v/V___V	{p} becomes	{v} between vowels
6.	p→w/#___	{p} in initial position	becomes {w}
7.	Voiced→Voiceless /___# C	Voiced Consonant becomes Voiceless Consonant	
8.	V→Ø/___#	vowels in final positions are lost	
9.	V→/V (C)___	{nas} {nas}	

Language Subgrouping

Langacker (1972) proposes that the basic criterion for establishing subfamilies is shared innovation. He states that if two or more languages have undergone a substantial number of common changes that have not occurred in any other daughters, it is likely that these languages constitute a subfamily and derive from a common pattern that does not underlie the other daughters.

Language and Dialects

Gleason (1985) proposes that if two people can understand one each other, then they speak the same language, if not, then they speak different languages. This means to say that variants of one language are dialects for there is mutual intelligibility between the speakers.

The research questions :

1. What phonemic correspondence sets stemming from innovation are found out in BLs?
2. How are the phonemic correspondence sets of BLs reconstructed?
3. What are the proto-phonemes of the correspondence sets?
4. What sound rules do account for sound changes of p(BLs) ?
5. How are *Bataknese* languages or BLs sub-grouped?
6. What phonemic clusters of sounds written the same orthographically but pronounced differently in TL?

III. METHODOLOGY

The data are gathered by listing basic core vocabulary in TL, AL, SL, KL, and DL relying on principles of basic core vocabulary used to gather data used in comparative study of sister languages being investigated to determine their proto-phonemes and sound change rules.

Swadesh (1952), suggests that basic core vocabulary covers personal pronouns, numerals, physical organs and their activities, nature, and instruments for daily use.

Similarly, (Keraf 1984) says the same thing by adding the words must be universal meaning that they exist in all sister languages.

The numbers of words proposed by linguists vary. According to Wikipedia, retrieved from https://en.wikipedia.org/wiki/Swadesh_list November 15, 2021, Swadesh created several versions of his list, 215 and reduced to 165 words. In 1952 he published a list of 215, then 200 words, and finally 100-words. Lees (1953), Rea (1958), Hymes (1960), and so on have their own versions. Dyen (1920) used a list composed of 200 words to investigate 95 language variants. Teeter (1963) argues that in gathering the data, loan words and taboo words as well as affixes have to be excluded.

In this research, since there are only five languages under investigation, I limit the number of basic core words into 23 words representing the universal words in BLs on the basis of what is previously mentioned.

The list is as follows:

	Indonesian glosses	English Words
1.	pintu	door
2.	bintang	star
3.	belah	split
4.	bamboo	bambu
5.	basi	stale
6.	dalam	deep
7.	mati	die
8.	kunyit	turmeric
9.	mayat	corpse
10.	betis	leg
11.	bibir	lips
12.	tahu	know
13.	hitam	black
14.	datang	come
15.	gigi	teeth
16.	jatuh	fall
17.	tongkat	stick
18.	makan	eat
19.	takut	fear
20.	jilat	lick
21.	nyamuk	mosquito
22.	tangkap	catch
23.	kentut	fart

The data are gathered by involving two native speakers of each TL, AL, SL, KL, and DL in the areas where each of the languages is used. In order for the data to be valid, the locations of the research are not in the borders of the areas where BLs are spoken to prevent mutual intervention between the languages.
Each of the words is transcribed phonetically to find out sets of phonemic correspondences by comparing phonemes and phonemes, phoneme clusters and phoneme clusters, and phonemes and phoneme clusters in comparable positions. Based on the result of comparative study, the proto- phonemes of each phonemic sets are analyzed. The finding will be used to determine the levels of similarity between the sister languages in BLs.
There are so many phonemes in BLs that are linearly reflected and consequently become the proto-phonemes of this kind of phonemic correspondence sets. The sets of linear phonemic correspondence sets will not be dealt with but the sets due to innovation will be what this article is about.

IV. RESULTS AND DISCUSSIONS

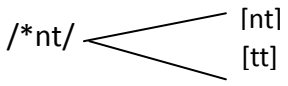
1. *Phonemic Correspondence Set* /tt-nt-nt-nt-nt/

Gloss: pintu ´door´

TL	AL	SL	KL	DL
p	p	p	p	p
i	i	i	i	i
t	n	n	n	n
t	t	t	t	t
u	u	u	u	u

From the data , it is found out that there are four phonemic correspondence sets namely /p-p-p-p-p/, /i-i-i-i-i/, /tt-tt-nt-nt-nt/, and /u-u-u-u-u/. The proto-phonemes of the sets are respectively /*p/, /*i/, /*nt/, and */u/. There are two types of changes of proto of BLs or p(BLs) in this data namely reflexes that are linearly inherited and reflexes resulted from innovation. The sets of phonemic correspondence inherited linearly are automatically the proto-phonemes of the sets like /*p/, /*i/, and /*u/ and will not be analyzed in this research. Only changes of sounds due to innovation like /tt/ and /nt/ need to be analyzed as how the changes take place and how to determine the proto-phonemes of phonemic correspondence sets in the sister languages.
The sound /*nt/ changes into /tt/ is due to conditioned sound change. When /n/ as nasal precedes voiceless stop dental /t/ in TL, it is conditioned by /t/ to become similar to it as voiceless stop dental in middle position and after vowel. This accounts for what Pike (1968) argues that neighboring sounds tend to affect one another, For the change of sound does not result in difference of meaning, [nt] and [tt] are the allophones of /*nt/.
The sound change may be set in this rule

/*nt/ → /tt/#___C
V



The sound correspondence set is distributed in another data.

Gloss: *kentut´* fart´

TL	Al	SL	Kl	DL
muttut	muntut	muntut	muntut	muntut

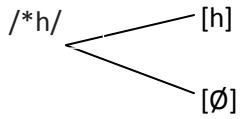
2. *Phonemic Correspondence Set* /Ø-Ø-h-h-h/

Gloss: *belah* ´split´

TL	AL	SL	KL	DL
b	b	b	b	b
o	o	o	o	o
l	l	l	l	l
a	a	a	a	a
Ø	Ø	h	h	h

In this data, there are five correspondence sets namely /b-b-b-b-b/, /o-o-o-o-o/, /l-l-l-l-l-l/, /a-a-a-a-a/, and /Ø-Ø-h-h-h/. The proto-phonemes of the correspondence sets are /*b/,/*o/,/*l/./*a/ and /*h/. The proto-phoneme of /Ø-Ø-h-h-h/ is /*h/, on the basis that it is the dominant phoneme or *majority wins* and the fact that /h/ loss is common in all languages, especially in initial and final positions as suggested by Crowly (1992).
As stated previously, to make sure that the suspicious phonemic set is a true phonemic correspondence set or cognate, it should be supported by its distribution in other data. That /Ø-Ø-h-h-h/ is cognate set is reinforced by its regular recurrence in other data.
The sound rule for the phonemic correspondence set is /*h / changes into /Ø/ in the final positions after vowels or /*h/.

/*h / →/Ø/___#
V



The phonemic correspondence set is recurrent in the following distributions.
Gloss:bambu ′bamboo′ and *jatuh* ′fall′

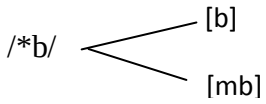
TL	Al	SL	Kl	DL
buluØ	buluØ	buluh	buluh	buluh
dʌbuØ	dʌbuØ	ndʌbuh	ndʌbuh	ndʌbuh

3. Phonemic Correspondence Set /b-b-b-mb-mb/
Gloss: basi ′stale′

TL	AL	SL	KL	DL
b	b	mb	mb	mb
ʌ	ʌ	ʌ	ʌ	ʌ
r	r	s	r	r
i	i	i	i	i

The phonemic correspondence sets stemming from innovation in this data to be tested are /b-b-b-mb-mb/ and /r - r- s- r- r/. However only the first can be proved to be phonemic correspondence set since it is supported by another regular recurrence and the second is not. The sound change rule is conditioned change for/b-b-b-mb-mb/, when /b/ is in initial positions in TL, AL, and SL, it conditions sound similar to it namely bilabial /m/ in front of it in KL and DL. It can be shown in the following formula:

/b/→/mb/___#
C



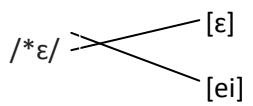
The phonemic correspondence set is endorsed by the following recurrence:
Glosses: *rumah* ′house′, *tahu* ′know, and *hitam* ′black’

TL	Al	SL	Kl	DL
bʌgʌs	bʌgʌs	bʌgʌs	mbʌgʌs	mbʌgʌs
bɔtɔ	bɔtɔ	bɔtɔ	mbɔtɔ	mbɔtɔh
birəŋ	birəŋ	birəŋ	mbiriŋ	mbiriŋ
biʌr	biʌr	biʌr	mbiʌr	mbiʌr

4. Phonemic Correspondence Set /ɛ-ɛ-ei-ɛ-ɛ/.
Gloss: mati ′die′

TL	Al	SL	Kl	DL
m	m	m	m	m
a	a	a	a	a
t	t	t	t	t
ɛ	ɛ	ei	ɛ	ɛ

Phonemic correspondence set under investigation is /ɛ-ɛ-ei-ɛ-ɛ/. Based on *the majority wins* principle, /*ɛ/ is the proto-phoneme of the set. Conditioned change may also account for it, /*ɛ/ changes into /ei/ in the final position in SL after consonants and remains the same in TL, AL, KL, and DL.
The rule of change is
/*ɛ/ → /ei/__#
C



The recurrence of this phonemic correspondence set is shown in the following data:
Glosses: *mayat* ′corpse′ and *padi* ′paddy′

TL	Al	SL	Kl	DL
bakkɛ	baŋke	bakkei	baŋkɛ	baŋkɛ

ɛmɛ	ɛmɛ	omei	ɛmɛ	ɛmɛ

5. *Phonemic Correspondence Set* /k-k-g-k-k/.

Gloss: kumis ´moustache´

TL	Al	SL	Kl	DL
k	k	g	k	k
u´	u	u	u	u
m	m	m	m	m
i	i	i	i	i
s	s	s	s	s

The suspicious set of phonemic correspondence is /k-k-g-k-k/. On the basis of *majority wins*, /*k/ is the proto-phoneme of the set and on the basis of conditioned sound, /*k/ as voiceless stop velar changes into voiced stop velar /g/ in initial position before vowel.

The sound change is due to conditioned change; It does not result in meaning change, meaning that /k/ and /g/are allophones of /*k/.

The sound rule is /*k/→/g/__#
V

The suspicious phonemic set is strengthened by another data as follows:

Gloss: *kuning* ´yellow´

TL	Al	SL	Kl	DL
kunik	kunik	guniŋ	kuniŋ	kɔniŋ

6. *Phonemic Corresspondence Set* /i-i-i-e-e/

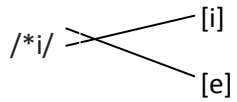
Gloss: betis ´leg´

TL	Al	SL	Kl	DL
b	b	b	mb	mb
i	i	i	i	i
t	t	t	t	t
i	i	i	e	e
s	s	s	s	s

In this data, it is predicted that there is a phonemic correspondence set due to innovation of its proto into its reflexes namely, /i-i-i-e-e/ and its proto-phoneme is /*i/ on the basis of *majority wins* principle and conditioned change of sound.

The rule of the change is /*i/ changes into /*e/ in KL and DL between two consonants and remains unchangeable in TL, AL, S. It can be formulated with

/*i/→/e/ C__C
V



This cognate set is endorsed by the following data.

Gloss: *bibir* ´lips´

TL	Al	SL	Kl	DL
bibir	bibir	bibir	mbiber	mbiber

7. *Phonemic Corresspondence Set* /ɔ-ɔ-ə-ə-ə/

Gloss: datang ´come´

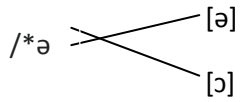
TL	Al	SL	Kl	DL
r	r	r	r	r
ɔ	ɔ	ɔ	ə	ə
Ø	Ø	h	h	h

In this data there are two suspicious phonemic correspondence sets namely /ɔ-ɔ-ə-ə-ə/ and /Ø-Ø-h-h-h/. However only the former will be analyzed, for the latter has been determined as phonemic correspondence set previously in gloss *bambu* ´bamboo´

The proto-phoneme of /ɔ-ɔ-ə-ə-ə/ is /*ə/ on the basis of *majority wins*. The sound change rule under conditioned change of sound for this data is /*ə/ changes into /ɔ/ in TL and AL after consonant in the final positions and between two consonants and does not change in SL, KL, and DL.

In can be shown in the following:

/*ə/ → /ɔ/#__V
C



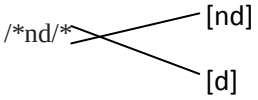
The phonemic correspondence set is also distributed in the following data:
Glosses: *nyamuk* ´mosquito´ and *gigi* ´teeth´

TL	Al	SL	Kl	DL
rəŋit	rəŋit	rəŋit	rəŋit	rəŋit
ipɔn	ipɔn	ipən	pən	ipən

8. Phonemic Corresspondence Set /d-d-nd-nd-nd/
Gloss: jatuh ´fall´

TL	Al	SL	Kl	DL
d	d	nd	nd	nd
ʌ	ʌ	ʌ	ʌ	ʌ
b	b	b	b	b
u	u	u	u	u
Ø	Ø	h	h	h

In this data, there are two sets of phonemic correspondence namely /d-d-nd-nd-nd/ and /Ø-Ø-h-h-h/. But only the first needs to be examined since the second one has been proved as phonemic correspondence set previously in gloss *bambu* ´bamboo´
The proto-phoneme of /d-d-nd-nd-nd/ is /*nd/ on the basis of *majority wins* principle. On the basis of conditioned sound change, the rule of sound change is /*nd/ changes into /d/ in TL and AL and remains unchangeable in SL, KL, and DL because alveolar nasal /n/ is lost before /nd/ to become alveolar /d/ in initial positions.
The change of sound is shown in the following:
/*nd/→/d/#___C
V



The cognate set can also be discovered in the following data:
Glosses: *jilat* ´lick´ *jauh* ´far´

TL	Al	SL	Kl	DL
dilʌt	dilʌt	ndilʌt	ndilʌt	ndilʌt
dʌɔ	dʌɔ	ndʌɔh	ndʌɔh	ndʌɔh

Based on the above analysis, the phonemic correspondence sets of BLs are as follows.
1. /tt-nt-nt-nt-nt/ of which proto-phoneme is /*nt/
2. /Ø-Ø-h-h-h/ of which proto-phoneme is /*h/
3. /b-b-b-mb-mb/ of which proto-phoneme is /*b/
4. /ɔ-ɔ-ə-ə-ə/ of which proto-phoneme is /*ə/
5. /ɛ-ɛ-ei-ɛ-ɛ/ of which proto-phonemes is /*ɛ/
6. /k-k-g-k-k/ of which proto-phoneme is /*k/
7. /i-i-i-e-e/ of which proto-phoneme is /*i/
8. /d-d-nd-nd-nd/ of which proto-phoneme is /*nd/

Based on the finding, it can be calculated which languages share the same phonemes and their frequency
1. /*nt/ is shared by AL,SL, KL, and DL
2. /*Ø/ is shared by TL and AL
3. /*b/ is shared by TL, AL, and SL
4. /*ə/ is shared by SL, KL, and DL
5. /*ɛ/ is shared by TL and AL, KL, and DL
6. /*k/ is shared by TL and AL
7. /*i/ is shared by TL, AL, SL, and KL
8. /*d/ is shared by TL and AL
9. /*nt/ is shared by AL, SL, KL, and DL
10. /*h/ is shared by SL, KL, and DL
11. /*mb/ is shared by KL and DL

12. /*k/ is shared by TL, AL, KL, and DL

According to frequency of sharing the same phonemes,

1. TL and AL share seven equal proto-phonemes namely /*Ø/, /*b/, /*ɛ/, /*k/, /*i/ /*d/, and /*k/.
2. TL and SL share two equal proto- phonemes namely /*b/ and /*i/.
3. TL and KL share two equal proto-phonemes namely /*ɛ/ and /*k/.
4. SL and KL share four equal proto-phonemes namely /*i/, /*nt/, /*h/, and /*ə/.
5. SL and DL share three equal proto-phonemes namely /*nt/, /*h/, and /*ə/.
6. KL and DL share six equal proto-phonemes namely /*ɛ/, /*nt/, /*h/, /*ə/, /*mb/, and /*k/
7. AL and SL, AL and KL, AL, and DL share one equal proto-phoneme namely /*nt/

With this in mind, it can be inferred that TL and AL have the highest level of sharing the same proto-phonemes namely /*Ø/, /*b/, /*ɛ/, /*k/, /*i/, /*d/, and /*k/. It means that both languages are closer one to another in comparison to the other languages. Consequently, they belong to one subgroup or subfamily.

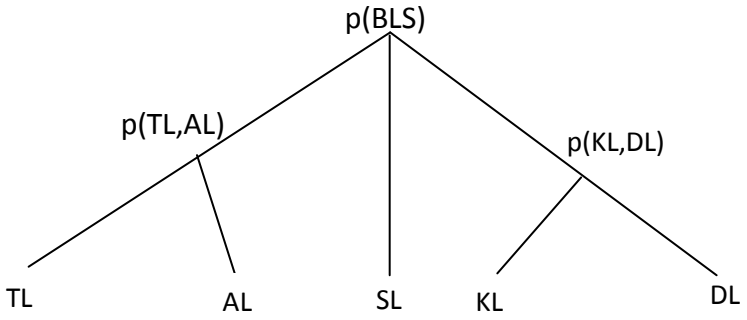
The second two languages that share the second highest level of sharing the same proto-phonemes are KL and DL namely /*ɛ/, /*nt/, /*h/, /*ə/, /*mb/, and /*k/. Thanks to it, the two languages have another subgroup or subfamily.

Meanwhile, SL is a distinct language that is not included in the two subgroups for it shares the lowest level of sharing the same proto-phonemes with other sister languages namely /*i/, /*nt/, /*h/, and /*ə/. Despite that it shares four equal phonemes with KL, there is no reason to put them in one group since SL is the only one language that has /ei/.

Since TL is my mother tongue, I know that TL speakers can understand one each other and so can the speakers of KL and DL and the speakers of TL and AL cannot understand KL, TL, and SL, whereas the speakers of SL understand none of TL, AL, KL, and DL. It can be stated that TL and AL are dialects of the same language namely p(TL,AL) and KL and DL are dialects of the same language namely p(KL,DL). Meanwhile, SL is a distinct language.

This fact is almost the same as lexicostatistic analysis as argued by Panggabean (1994) that cognate level of TL and AL is 85%, language, and cognate level of KL and DL is 76%, language whereas cognate level of SL and TL, AL, SL, KL, and DL is below 70%. In this research, out of seven proto-phonemes, TL and AL share all seven or 87,5% whereas KL and DL six or 75% and SL and KL, the closest language to it share only four or around 57% .

According to Dyen (1962), if cognate level between languages is over 81-100% they are dialects of one language and if it is below 36-80%, they are distinct languages. Although the result is not completely the same as Dyen's idea, mutual intelligibility as proposed by Gleason (1985) as the criterion of subgrouping, can account for establishing KL and DL as one subgroup since there is mutual intelligibility between the speakers of KL and DL.



As stated previously the way of many phoneme clusters are written equally in TL, AL, SL, KL, and DL such as /tt/ in TL is written /nt/ as in /muntut/ ´fart´ though it is pronounced with /tt/ as in /muttut/, and /kk/ is written /ngk/ orthographically and /ŋk/ phonetically as in /tangkap/´catch´ that is pronounced with /takkup/. If more data were available, there would be more evidences. Only in TL, /nt/ and /ŋk/ are pronounced differently, repectively /tt/ and /kk/.

V. CONCLUSIONS

On the basis of the results of analysis and discussion, the following conclusions can be made, there are eight-phonemes of BLs namely

1. /tt-nt-nt-nt-nt/ of which proto-phoneme is /*nt/
2. /Ø-Ø-h-h-h/ of which proto-phoneme is /*h/
3. /b-b-b-mb-mb/ of which proto-phoneme is /*b/

4. /ɔ-ɔ-ə-ə-ə/ of which proto-phoneme is /*ə/
5. /ɛ-ɛ-ei-ɛ-ɛ/ of which proto-phonemes is /*ɛ/
6. /k-k-g-k-k/ of which proto-phoneme is /*k/
7. /i-i-i-i-e/ of which proto-phoneme is /*i/
8. /d-d-nd-nd-nd/ of which proto-phoneme is /*nd/

The innovation resulting in variants of sounds are orthographically written the same but pronounced differently in TL. BLs can be put in three subgroups namely TL and AL subgroup KL and DL subgroup whreas SL is a distinct language separated from the two subgroups. In addition to it, /nt/ and /ŋk/ are pronounced the same in AL, SL, KL, and DL but differently in TL.

ACKNOWLEDGEMENTS

I would like to convey gratitude and thanks to all of my informants for providing me data of this research and Universitas Methodist Indonesia for giving me support to do this research and get the results published in international standard journal.

REFERENCES

[1]. Crowley, Terry 1992 *An Introduction to Historical Linguistics*. Oxford: Oxford University Press.
[2]. Dyen, Isodore 1956 *Language Distribution and Migration Theory*. in Language 32. *Language Divergence and Estimated Word Retention Rate*.
[3]. Gleason, H. A. Jr. 1961 *An Introduction to Descriptive Linguistics*. NewYork: Holt, Rinehart, and Winston.
[4]. Gudschinsky, Sarah S. 1956 *The ABC's of1exicostatistics (Glottochronology).* Word 12.
[5]. Haugen, E. 1972 *Dialect, Language, Nation. dalam Sociolinguistics*. New York: Penguin Book.
[6]. Hock, Hans Henrich 1988 *Principles of Historical Linguistics*. Berlin, New York, Amsterdam: Mouton de Gruyter.
[7]. Hymes, Dell H. 1959 *Genetic Classification: Retrospect and Prospect, Anthropological Linguistics,1*.
[8]. Keraf, Gorys 1990 *Linguistik Bandingan Tipologis.* Jakarta: Gramedia Utama. 1991 *Linguistik Bandingan Historis.* Jakarta: Gramedia Utama.
[9]. Langacker, Ronald W. 1972 *Fundamentals of Linguistics Analysis*. Sandiego: University of California.
[10]. Lehman, Winfred P. 1962 *Historical Linguistics.* Texas: The University of Texas.
[11]. McManis, Carolyn, Stollenwerk & Zheng-Sheng, Zhang 1987 Ohio: *Language Files*. Advocate Publishing Group.
[12]. Panggabean, Himpun 1994 *Telaah Bahasa-bahasa Batak dari Segi Leksikostatistik*. Bandung: Universitas Padjadjaran
[13]. Patji, Abdul Rachman. *Keragaman Bahasa Daerah Nusantara Terancam Punah*. Retrieved November 25, 2021 from http://www.antaranews.com/berita/
[14]. Pike, Kenneth. L. 1968 *Phonemics: A Technique for Reducing Language to Writing*. Michigan: The University of Michigan Press.
[15]. Swadesh, Morris 1952 *Lexicostatistic Dating of Prehistoric Ethnic Contacts.* dalam Proceedings of the American Philosophical Society 96.
[16]. Wikipedia, the free encyclopedia, retrieved from https://en.wikipedia.org/wiki/Swadesh_list November 15, 2021,