



PHONEMIC-PHONOLOGICAL AND LEXICAL VARIATIONS OF BATAK-TOBA LANGUAGE: A CASE OF LANGUAGE USE IN TAPANULI TENGAH AND TAPANULI UTARA

Elsen Agustina Sigalingging¹, Jufrizal²

English Department

Faculty of Languages and Arts

Universitas Negeri Padang

email: sigalinggingelsenagustina@gmail.com

Abstract

The study aims to determine the phonemic-phonological and lexical variations of the Batak Toba language as spoken in Tapanuli Tengah and Tapanuli Utara. The research focuses on identifying linguistic differences between two sub-dialects that share the same linguistic origin but develop distinct features due to geographical and social factors. The study employed descriptive qualitative method. The data were collected through interviews with ten native speakers selected based on the NORMS criteria, using Swadesh list of 207 basic lexical items and 54 additional daily vocabulary. The data were analyzed by classifying and comparing phonological and lexical forms across the two sub-dialects. The findings indicate that 24 phonemic-phonological variations, including 3 initial consonant shifts, 3 initial dropping, 5 gemination, 2 palatalization, 1 initial vowel change, 2 diphthongization, 7 vowel shift, and 1 nasal variation, as well as 73 lexical variations in which different words express the same meanings. These findings show that geographical and social factors play a significant role in shaping dialectal variation within the Batak Toba language.

Keywords: *Batak Toba language, dialectal, phonemic-phonological, lexical, Tapanuli Tengah, Tapanuli Utara*

A. INTRODUCTION

Language is dynamic and constantly varies due to the geographical, social, and cultural differences between its speakers. In sociolinguistic and dialectological studies, language variation is viewed as a natural phenomenon that reflects the identity of a community of speakers (Wardhaugh, 2006). This variation can occur at various linguistic levels, particularly in phonological and lexical aspects, which are often the main indicators in distinguishing one dialect from another. Previous research published in linguistics journals indicates that this variation is strongly influenced by social mobility and the intensity of language contact (Putri & Jufrizal, 2022).

¹ English ELLSP of English Department of FBS Universitas Negeri Padang graduated on March 2026

² Lecturer of English Department of FBS Universitas Negeri Padang



The Batak Toba language, one of the main regional languages in North Sumatra, exhibits quite rich internal variation. Although spoken by the same ethnic group, this language exhibits regional differences in sound and vocabulary. Tapanuli Tengah and Tapanuli Utara adjacent regions, yet they have distinct environmental and social characteristics. Tapanuli Tengah, as a coastal region, is more open to inter-ethnic interaction, while Tapanuli Utara, as a mountainous region, has a relatively homogeneous community and tends to maintain a more conservative form of language. These differences contribute to the emergence of dialectal variations between the two regions.

Several studies have examined regional language variation in Indonesia, from both lexical and phonological perspectives (Anjelia, 2019; Simanullang & Surip, 2023). However, studies that combine these two aspects simultaneously, particularly on the Batak Toba sub-dialects in Tapanuli Tengah and Tapanuli Utara are still limited. Yet, analyses that encompass both phonemic-phonological and lexical variation simultaneously are crucial for a more comprehensive understanding of language variation patterns.

Theoretically, lexical variation refers to differences in understanding of the same meaning, while phonological variation relates to differences in sound realization without changing meaning (Chambers & Trudgill, 1998). This variation is generally non-contrastive and serves as a marker of regional identity. In addition to its contribution to dialectological studies, documenting regional language variation is also important in the context of language preservation amidst the dominance of the national language.

This study discusses the phonemic-phonological and lexical variation of the Batak Toba language in Tapanuli Tengah and Tapanuli Utara, emphasizing the forms and types of variation that emerge as a reflection of the geographical and social differences between the two regions.

B. RESEARCH METHOD

This research used descriptive qualitative method because it is the most suitable approach for this study. Furthermore, this method used to examine the different lexical and phonological variations in the Tapanuli Tengah and Tapanuli Utara as sub-dialect of Batak Toba dialect. According to Creswell (2014), qualitative method relies on text and visual data, involved data analysis procedures, and incorporated diverse research designs. Therefore, that is the reason why the researcher used the descriptive method to describe the data.

The data of this research were the list of words obtained through interviews with the selected informants by used the *Swadesh list* and particular word list as the instrument. The data were in the form of vocabulary used and spoken by native speakers of the Batak-Toba language in Tapanuli Tengah and Tapanuli Utara. There were 207 lexicons and 54 personal data that use in daily life by Tapanuli Tengah and Tapanuli Utara people.

In determining the informants, the researcher was selected a number of native speakers who live in Tapanuli Tengah and Tapanuli Utara. From each group of speakers, a sample of informants were selected who are considered representative and met certain criteria. Based on Teeuw's opinion (1948)

informants in phonetic-phonological and dialectological research must meet the NORMS (Non-Mobile, Older, Rural, Male) criteria, namely being a local native who is not influenced by other dialects, age between 35 and 70 years, having basic to secondary education (elementary school, junior high school, high school), and being male. Meanwhile, according to Francis (1983) important characteristics that must be considered in selecting informants include age, education level, mobility, and other factors.

To collect the data, the researcher used an instrument in the form of *Swadesh list* table with total number 207 of words and 54 added word lists. According to Creswell (2014) a research instrument refers to a tool used to measure a concept or variable in a study. It includes surveys, interviews, and observation protocols will be used to collect data from participants. This research used four instruments to collect data because this is a descriptive qualitative. The researcher gathered the data directly by using the following instruments:

1. *The first instrument was the researcher herself.* The researcher collected the data from 10 informants for both sub-dialects, with five from the Tapanuli Tengah sub-dialect and the remained five from the Tapanuli Utara sub-dialect.
2. *The next instrument was a list of words.* The list included both basic words from the *Swadesh List* with total number 207 lexicon and 54 additional words from the basic vocabulary used in everyday conversation by speakers in Tapanuli Tengah and Tapanuli Utara.
3. *Note-taking by using notebooks and writing tools,* which used to record the specific words spoken by the informants.
4. *Recording device,* such as a smartphone or video camera, which used to document the interviews between the researcher and the informants.

Techniques of data collection that had been used by researcher were (1) went to the field and have conversation with the informants, (2) direct observation and interviewed to collect the data, (3) interviews based on list of words containing 207 lexicons from the *Swadesh list*, (4) taking a note about original word that unique and local vocabulary in Tapanuli Tengah and Tapanuli Utara sub-dialect, (5) recorded the communication and takes notes to ensure the accuracy of the information provided by the informants.

The next, technique of data analysis that had been used by researcher were (1) identifying the data (2) classifying the data (3) analyzing the data (4) drawing conclusion.

C. RESULT AND DISCUSSION

1. Research Finding

Based on the descriptive data analysis, it is found in this research that there are similarity and differences of phonological and lexical. According to the research questions, there are the several findings of the research:

4.2.1 Phonemic-Phonological Differences found of Batak Toba Language in Tapanuli Tengah and Tapanuli Utara Sub-dialects

Moreover, in the phonemic-phonological differences in the Tapanuli Tengah and Tapanuli Utara sub-dialects, there are several types of vowel changes found by the researcher; initial consonant shift, for example the word '*dia*' which is pronounced *imana* /imana/ in the Tapanuli Tengah sub-dialect changes to *ibana* /ibana/ in the Tapanuli Utara sub-dialect, initial dropping, for example the word '*kapan*' which is pronounced *nandigan* /nandigan/ in the Tapanuli Tengah sub-dialect changes to *andigan* /andigan/ in the Tapanuli Utara sub-dialect, double consonants and gemination, for example the word '*panjang*' which is pronounced *ganjang* /ganjan/ in the Tapanuli Tengah sub-dialect changes to *gajjang* /gajjan/ in the Tapanuli Utara sub-dialect, palatalization and change in place of articulation, for example the word '*mencuci*' which is pronounced *manucci* /manucci/ in the Tapanuli Tengah sub-dialect changes to *manussi* /manussi/ in the Tapanuli Utara sub-dialect, initial vowel change, for example the word '*bulu*' which is pronounced *imbulu* /imbulo/ in the Tapanuli Tengah sub-dialect changes to *ambulu* /ambulo/ in the Tapanuli Utara sub-dialect, diphthongization, for example the word '*water*' which is pronounced *aek* /aek/ in the Tapanuli Tengah sub-dialect changes to *aik* /aik/ in the Tapanuli Utara sub-dialect, vowel shift, for example the word '*pendek*' which is pronounced *jeppek* /jeppe?/ in the Tapanuli Tengah sub-dialect changes to *joppok* /joppo?/ in the Tapanuli Utara sub-dialect, and nasal variation, for example the word '*semut*' which is pronounced *porhis* /porhis/ in the Tapanuli Tengah sub-dialect changes to *porngis* /porngis/ in the Tapanuli Utara sub-dialect. These variations do not result in changes in meaning but merely indicate differences in sound realization between Tapanuli Tengah and Tapanuli Utara speakers.

4.2.2 Lexical Differences found of Batak Toba Language in Tapanuli Tengah and Tapanuli Utara Sub-dialects

The researcher found that 73 differences in vocabulary at the lexical level in the Tapanuli Tengah and Tapanuli Utara subdialects based on the *Swadesh list* and additional words lists. For example, the word '*mereka*' *halaki* /halaki/ 'they' in Tapanuli Tengah sub-dialect totally changes into '*mereka*' *nasida* /nasida/ 'they' in Tapanuli Utara sub-dialect. Another example, the word '*rumah*' *bagas* /bagas/ 'house' in Tapanuli Tengah sub-dialect totally changes into '*rumah*' *jabu* /jabu/ 'house' in Tapanuli Utara sub-dialect. These words have different forms but convey the same meaning. This lexical variation is primarily due to geographical differences, migration history, and intensity of language contact.

4.2.3 Lexical Similarities found of Batak Toba Language in Tapanuli Tengah and Tapanuli Utara Sub-dialects

The researcher found a very high level of lexical similarity, with 164 of the 261 words having identical forms in the Tapanuli Tengah and Tapanuli Utara sub-dialects. For example, the word '*kamu*' *ho* /ho/ 'you' in Tapanuli Tengah sub-dialect is the same as '*kamu*' *ho* /ho/ 'you' in Tapanuli Utara sub-dialect. Another example, the word '*kuku*' *sasilon* /sasilon/ 'nail' in Tapanuli Tengah sub-dialect is the same as '*kuku*' *sasilon* /sasilon/ 'nail' in Tapanuli Utara sub-dialect. The similarity shows that Tapanuli Tengah and Tapanuli Utara are still in a very close linguistic system. This high level of similarity indicates that the differences found are not structural divergences, but rather natural variations due to geographical and social differences.

The findings reveal that the two regions share a high degree of linguistic similarity, yet several lexical and phonological differences exist due to geographical, cultural, and social factors. The overall results are presented in the table below.

Category	Number of Items	Description
Lexical Similarities	164	Words that are identical in both form and meaning across both regions.
Lexical Variations	73	Words that differ in form but express the same meaning.
Phonological variations	24	Words with different sound realizations but identical lexical roots.
Total	261	

2. Discussion

The results of this study provide important insights into how the Batak-Toba language, as spoken in Tapanuli Tengah and Tapanuli Utara, exhibits both linguistic uniformity and regional variation. The analysis of 261 lexical items revealed that the majority of the lexicon remains identical across both regions, demonstrating a strong linguistic unity. Therefore, by making these items as guideline, it was make us understand why language has many variations, especially in phonological variation. According to Trudgill (2000), phonological variation in dialects is a form of free variation when two different phonetic forms are used for the same semantic function without

causing ambiguity of meaning. Some differences were found in the form of phonemic-phonological variation.

Point one is differences in phonemic-phonological level. First, the findings in this study indicate that sound shifts in initial positions such as the alternation of /m/ and /b/. For example, the word '*dia*' which means 'she, he, it' in Tapanuli Tengah is pronounced /*imana*/, but in Tapanuli Utara the word '*dia*' is pronounced /*ibana*/. Then, the omission of the phonemes /n/ and /s/ in initial positions. For example, the word '*kapan*' which means 'when' in Tapanuli Tengah is pronounced /*nandigan*/, but in Tapanuli Utara the word '*kapan*' is pronounced /*andigan*/. This phenomenon is reinforced by Roach (2009) who calls elision a natural phonotactic strategy to simplify articulation in everyday speech. In the context of the Batak Toba, speakers in Tapanuli Tengah and Tapanuli Utara display this form of variation consistently, thus becoming a marker of dialectal identity.

Next, the differences in phonemic-phonological level that is gemination. According to Fromkin et al. (2011), gemination is a form of phonetic emphasis based on the articulatory habits of a language community, which does not necessarily change the phonological structure. The results of this study indicate that variations in medial position such as gemination of the consonant /n/ to /jj/, /dd/, or /tt/ are forms of articulation strengthening that function as dialectal phonetic markers, not changes in the phoneme system. For example, the word '*panjang*' which means 'long' in Tapanuli Tengah is pronounced /*ganjan*/, but in Tapanuli Utara the word '*panjang*' is pronounced /*gajjan*/. Another example is the word '*tanduk*' which means 'horns' in Tapanuli Tengah is pronounced /*tanduk*/, but in Tapanuli Utara the word '*tanduk*' is pronounced /*tadduk*/. This kind of gemination pattern was also found in previous research by Simanullang and Surip (2023) in the Barus dialect, which showed that the Toba Batak community in the coastal area displayed a form of gemination due to intensive contact with speakers of other regional languages. Thus, the data of this study indicate that Tapanuli Utara maintains a stronger form of articulation, while Tapanuli Tengah tends to display sound reduction due to higher exposure to cross-ethnic interactions.

Furthermore, a phonemic-phonological difference is the vowel shift. Labov (1994) states that vowel shift occurs when a vowel system repositions sounds to maintain internal phonological contrasts within a dialect. The results of this study found that changes in vowel quality such as /i/ to /a/, /e/ to /i/, and /o/ to /e/, including the formation of diphthongs, are part of this systemic mechanism. For example, the word '*cacing*' which means 'worm' in Tapanuli Tengah is pronounced /*goya*/, but in Tapanuli Utara the word '*cacing*' is pronounced /*geya*/. This shift was found repeatedly in various lexemes, indicating that the vowel system in the Batak Toba language in both regions did not add new phonemes, but made phonetic adjustments to maintain contrast. This phenomenon is in line with previous research by Putri and Fitrawati (2020) on the Minangkabau dialect, which showed that

geographic location accompanied by active interaction through trade and community mobility resulted in stable vowel shifts within the dialect.

Next, point two is differences in lexicon or vocabulary. The differences between the Tapanuli Tengah and Tapanuli Utara also occurs in the lexicon totally. For example, the word '*mereka*' which means 'they' in Tapanuli Tengah is pronounced /*halaki*/, but in Tapanuli Utara the word '*mereka*' is pronounced /*nasida*/. It means all vocabularies, pronunciations and phonemes are changed. It supported by the researcher that have been conducted the research about lexicon between Tapanuli Tengah and Tapanuli Utara sub-dialects. Lyons (1995) explains that lexical variation reflects the cultural background and social experiences of a community, so that differences in understanding between the two regions indicate not only linguistic divergence but also divergence of cultural identity.

According to Holmes (2013), linguistic variation serves not only as a medium of communication but also as a symbol of social and cultural identity within a community. The findings of this study clearly demonstrate that phonological and lexical variation in the Batak Toba language spoken in Tapanuli Tengah and Tapanuli Utara does not hinder mutual understanding among speakers, but rather serves as a marker of local identity for each region. This variation also represents the vitality of regional languages that develop dynamically in relation to geographical factors, historical migration, and social contact. Therefore, the overall findings support theoretical frameworks in dialectology and sociolinguistics, which suggest that variation is an inherent characteristic of language development in geographically separated speech communities that remain within the same linguistic system.

In comparison with studies on Minangkabau dialects conducted by Nesti (2016), Mujahidaturrahmah & Saun (2018), and Suksio & Jufrizal (2024), this research shows a similar pattern in which lexical variation outnumbers phonological variation. Nevertheless, this study differs in that it combines phonemic-phonological and lexical analysis simultaneously, whereas most previous studies focused on only one linguistic aspect. This integrated approach allows for a more comprehensive understanding of dialectal variation in the Batak-Toba language.

From the analysis and research findings, it clears that the Tapanuli Tengah and Tapanuli Utara has own characteristics. It makes the words pronounced differently but still has same meaning. The differences can occur in vocabulary and pronunciation. Overall, this research strengthens previous findings that dialectal variation is shaped by geographical conditions, social interaction, and cultural identity.

D. CONCLUSION AND SUGGESTIONS

1. Conclusion

In the evolution of the languages, lexical and phonemic-phonological differences between the *Tapanuli Tengah* and *Tapanuli Utara* sub-dialect is used commonly occur. One of the phenomena of lexical and phonetic differences are found in the *Tapanuli Tengah* and *Tapanuli Utara* sub-dialect used which has many differences. From the discussion and finding in this research, it can be concluded that:

- (1) In initial position sound changes occur, such as initial consonant shift and initial dropping.
- (2) In medial position, a pattern of consonant gemination and palatalization are found.
- (3) There are core vowel shifts, such as the replacement of /i/ and /a/, /e/ and /i/, /o/ and /e/.
- (4) Nasal Variation that is phoneme /h/ changes to be /ŋ/

Furthermore, it is found that there are 73 different word or vocabulary at lexical level and 24 different word or vocabulary at phonemic-phonological level in *Tapanuli Tengah* and *Tapanuli Utara* sub-dialects, from 261 words in total on this research. These variations reflect the influence of social and environmental factors in the language's use. *Tapanuli Tengah*, located in the coastal region, has received more foreign language influence, thus exhibiting more innovative forms, while *Tapanuli Utara* maintains more traditional forms, which is seen in the choice of comprehension and phonological features such as geminate.

2. Suggestions

Based on the findings and conclusions of this study, several suggestions can be drawn. Future researchers are encouraged to expand the scope of similar studies to other Batak-speaking regions and to examine additional linguistic aspects such as morphology, syntax, or semantics for a more comprehensive understanding of language variation. The use of quantitative methods such as dialectometric or phonetic analysis is also recommended to measure variation more precisely. Furthermore, for the Batak Toba community, it is important to continue using and appreciating their regional varieties in daily communication.

BIBLIOGRAPHY

- Anjelia, T., & Rosa, R. N. (2019). Dialectal Comparison Between Simarasok Sub-Dialect and Padang Tarok Sub-Dialect: The Lexical Variation. *English Language and Literature*, 8(3). <https://doi.org/10.24036/ell.v8i3.105774>
- Banjarnahor, A. S. (2017). Afiks-Afiks Derivasi Bahasa Inggris Dan Bahasa Batak Toba (Suatu Analisis Kontrastif). *Jurnal Elektronik Fakultas Sastra Universitas Sam Ratulangi*, 1(4).
- Chambers, J.K, Trudgill, P. (2004). *Dialectology*: Second Edition. United States of America. Cambridge Press University.
- Creswell, J. W. R. (2014). *Research design: qualitative, quantitative, and mixed methods approaches*. In SAGE Publications, Inc. All (Vol. 13, Issue 1).
- Fishman, J.A, (2001). *Handbook of Language and Ethnic Identity*. New York: Oxford university press.
- Francis, W.N. (1983). *Dialectology: An Introduction* New York: Longman Linguistic Library
- Fromkin, V., Rodman, R., & Hyams, N. (2011). *An introduction to language* (9th ed.). Boston, MA: Wadsworth Cengage Learning.
- Holmes, J. (2013). *An introduction to sociolinguistics* (4th ed.). New York, NY: Routledge.
- Jufrizal. (1999). *Introduction to General Linguistic*. Padang: DIP Universitas Negeri Padang.
- Kim, L. S. (2003). Exploring the relationship between language, culture and identity. *GEMA Online Journal of Language Studies*, 3(2).
- Kridalaksana, H. (1985). *Tata Bahasa Deskriptif bahasa Indonesia: Sintaksis* (Dendy Sugono (ed.)). Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Mujahidaturrahmah, F., & Saun, S. *The Lexical Comparison Between Tebing Tinggi Sub-Dialect in Dharmasraya and the Standard Minangkabau*. *E-Journal of English Language and Literature (JELL)*, 7(1), 83–91. <http://ejournal.unp.ac.id/index.php/jell>
- Nesti, M. R. (2016). Variasi leksikal bahasa Minangkabau di kabupaten pesisir selatan. *Jurnal Arbitrer*, 3(1), 46-61.

- Putri, M. Rahma. Jufrizal. (2022) Dialectal Comparison Between Muaro Takung and Maloro Sub-dialects: The Lexical Variations. *E-Journal of English Language and Literature*. <http://ejournal.unp.ac.id/index.php/jell>
- Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.
- Simanullang, R., & Surip, M. (2023). Analisis Sinkronik Dialek Bahasa Batak Toba di Barus Tapanuli Tengah: Kajian Dialektologi. *Jurnal Pendidikan Sosial dan Humaniora*, 2(3).
- Suksio, A. Jufrizal. (2024). Phonetic-phonological and Lexical Comparison of Lubuk Basung Sub-dialect with The Standard Minangkabaunese. *E-Journal of English Language and Literature*. <http://ejournal.unp.ac.id/index.php/jell>
- Swadesh, Morris. 1955. *Towards greater accuracy in lexicostatistic dating*. *International Journal of American Linguistics* 21. 121-137.
- Tarigan, A. R., Simanjuntak, F. A., Harahap, S. H., Malau, S., & Manullang, T. L. (2024). *Fenomena Kedwibahasaan Pada Masyarakat Tutur Perbatasan Batak Simalungun-Batak Toba Kajian Sociolinguistik Di Kabupaten Simalungun*. *Jurnal Review Pendidikan dan Pengajaran (JRPP)*, 7(3), 8915-8921.
- Trudgill, P. (2000). *Sociolinguistics: An introduction to language and society* (4th ed.). London: Penguin Books.
- Wardhaugh, Ronald & Fuller, Janet M. (2015). *An Introduction to Sociolinguistics*. Wiley-Blackwell.
- Yule, G. (2017). *The study of language* (6th ed.). Cambridge University Press.