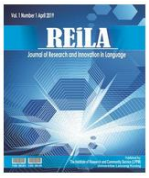
REiLA
Journal of Research and Innovation in Language
ISSN (Online): 2685-3906, ISSN (Print): 2685-0818
DOI: https://doi.org/10.31849/reila.17024
Vol. 7, No. 1, April 2025, pp. 29-46



The Patterns of Ablaut Reduplication in Toba Batak: A Construction-Based Approach

Esron Ambarita & Milisi Sembiring
Universitas Methodist Indonesia, Medan, Indonesia
esronambarita@gmail.com

ARTICLE HISTORY

Received : 2023-11-14
Revised : 2024-11-26
Accepted : 2025-03-05

KEYWORDS

Ablaut reduplication
Toba Batak language
Construction-based approach
Morphology
Endangered languages

ABSTRACT

The study addresses a critical gap in morphological research by investigating the understudied patterns of ablaut reduplication in Toba Batak (TB), an endangered Austronesian language, through a construction-based approach. While ablaut reduplication has been widely examined in Indo-European languages like English and German, its manifestations in TB remain largely unexplored, despite the language's cultural significance and vulnerability. This study fills this gap by employing a modified version of Halle's (1973) generative morphology model, tailored to TB's unique linguistic features, to analyze data collected from 12 native speakers in Samosir Regency. The findings reveal three distinct categories of ablaut reduplication in TB: verbal (VAR), nominal (NAR), and adjectival (AAR), each exhibiting specific vowel patterns. For instance, monosyllabic words typically alternate between [u]/[e] and [a]/[u], while disyllabic words follow a [u-a] to [a-I]/[a-u] sequence. Notably, the study demonstrates that TB's ablaut reduplication is unproductive, lacking grammatical meaning and affixation, with lexical meaning derived solely from vowel alternations. The modified Hallean model, incorporating orthographic and phonological rules, proves effective in capturing these nuances, offering a novel framework for analyzing non-Indo-European languages. By documenting these patterns, the study not only enriches the understanding of TB's morphological system but also contributes to broader linguistic typology and the preservation of endangered languages. The implications extend to applied fields such as language education and cultural revitalization, underscoring the urgency of safeguarding regional languages like TB amidst globalization. This research thus bridges theoretical linguistics and practical conservation efforts, advocating for further studies on underrepresented languages.

1. Introduction

From the playful rhythm of *tick-tock* to the lively cadence of *zig-zag*, ablaut reduplication captivates linguists and speakers alike. This phenomenon, where repeated words undergo systematic vowel changes (e.g., *flip-flop*, *chit-chat*), operates on an instinctive level that defies simple grammatical explanation (Downing, 2015; Körtvélyessy, 2016; Regjer, 2022). Its appeal is universal, appearing in childhood language acquisition (Downing & Inkelas, 2015) and popular culture (e.g., *Chitty Chitty Bang Bang*; Curto, 2024). Yet native speakers intuitively reject deviations like *tock-tick* without being able to articulate why (Puspani & Indrawati, 2021). As McLendon (2020) observes, this represents one of language's great mysteries: a rule everyone follows but no one can explicitly state.

The phenomenon extends far beyond English. We find it in Arabic *dam-dam* for rumbling sounds, Russian *Тарарам* for noise, German *schnick-schnack* meaning nonsense, and Japanese *doki-doki* mimicking a heartbeat (Lǐ & Ponsford, 2018; Mattes, 2017). While English has received the most scholarly attention (Minkova, 2002; Wallace, 2019; Weijer, 2020), significant work has also examined Germanic roots (Jasanoff, 2007), Slavic patterns (Körtvélyessy, 2020), and Austronesian languages like Javanese (Yip, 2009). However, this global perspective reveals a striking omission: endangered languages with unwritten morphological traditions remain largely undocumented.

Within the Austronesian family, Toba Batak (TB) represents a critical gap in reduplication studies. Though occasionally mentioned in broad surveys

Corresponding Author: Ambarita [Email: esronambarita@gmail.com].

(Blust, 2022, 2023; Wivell et al., 2024), this Sumatran language has never been the focus of dedicated ablaut research. TB's cultural significance as the language of Lake Toba communities (Sinaga, 2002) contrasts sharply with its endangered status, as speakers increasingly shift to Indonesian (Lubis & Bowo, 2022). Most existing TB research examines cultural artifacts like folktales (Simaremare et al., 2023) rather than its grammatical systems.

The morphological work of Ambarita, (2023) stands as a rare exception, having explored TB affixation and nominalization. Yet even this foundational research only touches on reduplication tangentially, with one study on adjective reduplication (Ambarita, 2018) and another on base full reduplication (Ambarita, 2023). The complete absence of work on TB's ablaut patterns represents a significant lacuna, particularly given the phenomenon's centrality to many Austronesian languages.

This study breaks new ground by applying a construction-based approach to TB's ablaut reduplication. We adapt Halle's (1973) generative morphology model, traditionally used for Indo-European languages, to accommodate TB's unique Austronesian structure. Our modified framework incorporates orthographic and phonological rules to analyze how TB's vowel alternations (e.g., [u-a] to [a-i]) compare with global patterns. This approach reveals previously undocumented aspects of non-concatenative morphology in endangered languages. This study holds both theoretical and practical significance. Theoretically, it challenges existing assumptions about reduplicative productivity in the Batak Toba language by presenting new evidence of unique linguistic patterns that deviate from established norms (Ambarita, 2018). Notably, it identifies distinct morphological constructions, including irregular verb affixations and prefixed forms that differ significantly from Indonesian (Cristy et al., 2023; Damanik & Mulyadi, 2020).

These findings contribute to the broader theoretical discourse on language structure and deepen our understanding of how such patterns shape verbal communication among Toba Batak speakers. Practically, the study aligns with Indonesia's constitutional mandate to preserve regional languages, emphasizing the crucial role of linguistic diversity in maintaining cultural heritage (Rajagukguk et al., 2022). For the Toba Batak community, language is more than communication; it preserves historical narratives, cultural values, and traditions, strengthening their connection to ancestry across generations (Rajagukguk et al., 2022).

This research highlights the importance of revitalization efforts, including the use of traditional expressions, or *umpasa*, particularly in ceremonies such as weddings, which serve as vessels for transmitting cultural knowledge across generations (Sitanggang et al., 2024).

Furthermore, language preserves cultural identity through traditional songs and narratives that reflect core societal values (Saragih, 2021). By documenting and analyzing these linguistic and cultural elements, this study not only contributes to linguistic scholarship but also reinforces the cultural significance of the Toba Batak language, offering insights that support both academic inquiry and practical preservation efforts (Rajagukguk et al., 2022).

Beyond academic circles, the research has tangible applications. The memorable quality of ablaut patterns, already exploited in advertising (Shariq, 2020), could inform TB-language media campaigns. Additionally, the findings may guide bilingual education programs aimed at maintaining intergenerational transmission (Lubis & Bowo, 2022). By illuminating TB's ablaut reduplication patterns, this study does more than fill a linguistic gap. It demonstrates how documenting endangered languages can simultaneously advance theoretical understanding and support cultural preservation. In an era of rapid language loss, such work takes on urgent importance, offering both scholarly insights and practical tools for language maintenance.

2. Literature Review

This section first clarifies the theoretical notions and assumptions that underlie this morphology research, as the understanding the concepts are necessary in order to make sense of the results of the research. After conceptualizing the relevant components of generative morphology, this section presents its review of the overall trend of the research to highlight how understudied Toba Batak language is.

2.1 Generative Morphology

Generative morphology focuses on word formation processes that generate actual and potential forms through rules and filters (Nida, 1949; Chomsky, 1968; Mathews, 1974; Haspelmath et al., 2010). To generate means to bring into existence or produce (Hornby, 1987), emphasizing the mechanism of word transformation (Samsuri, 1982; Suhadi, 2018). For instance, Templeton (2012) illustrates that learning the word *courage* leads to acquiring related words like *courageous*, *encourage*, *discourage*, *discouragingly*, *discourageable*, *undiscouraged*, *encouragement*, and *encouragingly*, totaling nine words. Wekker and Haegeman (1996) argue that understanding a language enables intuitive word transformation, a concept supported by Lipka (1975). In generative linguistics, Halle's 1973 model has been widely used for analyzing morpheme combinations (ten Hacken, 2020). The model is displayed in Figure 1.

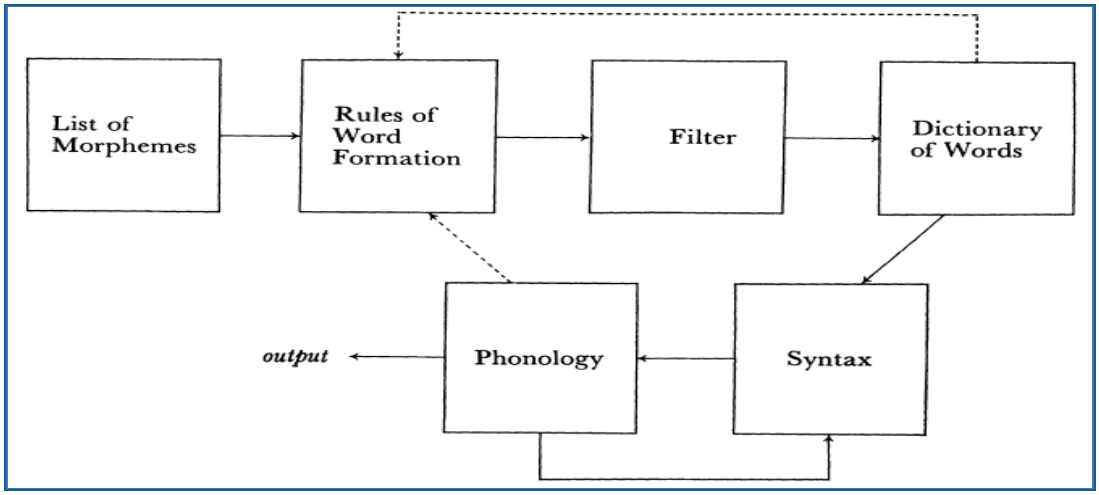


Figure 1. The Original Model of the Generative Morphology Theory (Halle, 1973, p. 8)

Manzini (2021, p. 1) states that Morris Halle established generative morphology as a computational model in which morphemes are the fundamental units and morphological merge is the core operation. This model dominates the generative field by offering a framework for understanding word formation through morpheme combinations. It is crucial for analyzing word structure and formation rules, making it suitable for this qualitative phenomenological case study, as it provides a detailed and nuanced data description. Researchers find it appealing due to its applicability across languages, enabling comparisons of morphological processes to identify universal principles and language-specific rules. Additionally, its clear distinction between morphemes, rules, and filters supports computational modeling for automatic word generation and analysis, benefiting language processing tasks.

2.1 List of Morphemes (LM)

Morphemes play a crucial role in word formation, encompassing both free and bound morphemes. They are categorized into basic lexemes and affixes (Crystal, 2008; Arbi et al., 2022). Lexical items consist of phonetic sequences with grammatical information, where basic lexemes are classified into free forms (bases) and bound forms (stems). In this study, basic lexemes fall into major lexical categories such as nouns (N), verbs (V), adjectives (ADJ), adverbs (ADV), and numerals (NUM). Affixes, the second category, are divided into derivational and inflectional types. Derivational affixes modify a base word's meaning, as in *happy (ADJ) → happiness (N)*, while inflectional affixes do not change the grammatical class but indicate grammatical rules, as in *study (V) → studying (V)*. As shown in Figure 1, the first step in generative morphology research is to register all morphemes from the data in the List of Morphemes (LM).

2.1.2 Word Formation Rules (WFR)

Once the List of Morphemes (LM) is complete, the next step is to establish the Word Formation Rules (WFR), which define how nouns (N), verbs (V), adjectives (ADJ), adverbs (ADV), and other categories are formed (Crystal, 2008). These rules guide the correct formation of morphemes, ensuring the production of acceptable words in a language. When a word follows WFR and is combined with different affix morphemes, it becomes a productive word. Word formation occurs entirely in the lexicon, with WFR serving as the specialized mechanism governing this process (Aronoff, 1976; Scalise, 1984). Morphologically, it results in grammatical meaning (GM) and lexical meaning (LeM), determining a language's potential words. However, while WFR can generate phonologically, syntactically, and semantically valid words, some of these words may never actually appear in the language or be used by its speakers (Scalise, 1984).

2.1.3 Filter

When words are formed, they may be phonologically, syntactically, or semantically acceptable or unacceptable, and their acceptability is determined by the filter (Mathews, 1974). Formed words undergo a morphophonological process involving phoneme assimilation, deletion, addition, and other modifications. If a word's structure is deemed unacceptable, it is intercepted and refined to ensure it conforms to the language's standards. However, not all words can be generated by Word Formation Rules (WFR) due to language-specific exceptions. The filter identifies these exceptions and assigns idiosyncratic characteristics, which can be categorized into semantic, phonological, and lexical idiosyncrasies. Klausenburger (1979) in Jensen (1995) describes the process of integrating phonological conditional formulas into morphological ones as morphologicalization. Acting as a repository

of idiosyncratic information, the filter documents unique characteristics of each lexeme, including potential words (Jackendoff, 1975).

2.1.4 Dictionary

Words generated through Word Formation Rules (WFR) undergo filtering, where only acceptable structures proceed to the final stage, which is the dictionary. This serves as the ultimate repository for all valid words, including those from the List of Morphemes (LM) and derivative words formed through WFR that have successfully passed the filter. Each entry is accompanied by its meaning and distinctive characteristics (Jensen, 1995). The dictionary stores both regularly formed words that remain unchanged by the filter and idiosyncratic formations that have been modified as needed (Sirulhaq et al., 2022).

2.2 Ablaut Reduplication

Ablaut involves a vowel change that alters a word's grammatical function, as seen in drink, drank, and drunk (Crystal, 2008). Reduplication, on the other hand, repeats all or part of a word to create a new meaning, such as mama, papa, or boo-boo (Crystal, 1993). Ablaut reduplication combines both processes, forming pairs where vowel alternation conveys specific meanings (Wivell, 2024). These words consist of a base and a reduplicant, following phonological patterns that, when disrupted, feel intuitively incorrect to native speakers. English contains at least 2,000 ablaut reduplications (Thun, 1968), typically structured as consonant-vowel-consonant sequences, following patterns found in Indo-European languages like Indo-Iranian, Latin, and Greek (Mathiassen, 1969). Ablaut reduplication exhibits a trochaic contour, with stress variations influenced by speaker emotions and competencies (Marchand, 1957), distinguishing rhyme-motivated compounds like namby-pamby from ablaut-motivated ones like shilly-shally (Bauer, 2006). Recent studies highlight its peripheral role in linguistic theory (McCarthy, 1992) and its relevance in typological research, as seen in Middle Welsh and Greek, where i-reduplication replaced e-reduplication in certain verb forms (Weijer et al., 2020; Weiss, 2010). Germanic languages present further complexities, with ablaut developing from reduplication in strong verbs (Durrel, 1975; Jasanoff, 2007). The positioning of vowels and consonants in Indo-European languages often reflects historical phonological developments (Penney, 1977; Vertegaal, 2020). Linguists continue to debate ablaut reduplication's linguistic function and its place within language systems, as orthographic rules evolve and its historical origins remain unclear (Dance, 2019; Guardiano, 2021). The phenomenon intrigues scholars because it is systematic in some languages while irregular in others (Giannakis, 1992).

2.3 Ablaut Reduplication in Indonesian and Toba Batak

Among Among the 26 language families, the Austronesian language group exhibits distinct characteristics of ablaut reduplication, particularly in the alternation between low and high vowels. Some examples include:

1) **Indonesian (Austronesian, Malayo-Sumbawan):**
   - *basa-basi* → 'polite language'
   - *jungkat-jungkit* → 'see-saw'

2) **Javanese (Austronesian, Javanese):**
   - *celak-celuk* → 'to call a name repeatedly'
   - *elang-eling* → 'to remember'

3) **Toba Batak (Austronesian, Northwest Sumatra barrier-islands):**
   - *mangabasmangebüs* → 'to strike right and left (like the tail of a crocodile)'

These examples illustrate how ablaut reduplication occurs across different languages, each following unique phonological patterns while maintaining the core principle of vowel alternation. Further, in English, ablaut reduplication has been categorized into three types by Jespersen (1965): (1) Ablaut reduplication (*riff-raff*), (2) Rhyme reduplication (*hocus-pocus*), and (3) Copy reduplication (*boo-boo*). According to McCarthy (2004), reduplication occurs when elements of the base have corresponding counterparts in the repeated form. It is a productive morphological process in Indo-European languages, where it can function as nouns, adverbs, verbs, and particles (Giannakis, 1992).

Languages are constantly undergoing sound changes, although the rate and nature of these changes vary. External influences, such as contact with other ethnic groups, can also impact linguistic evolution (Hickey, 2014). In Indonesian, reduplication often involves phoneme alterations within one of the syllables. Some examples include:

- *compang-camping* → 'ragged clothes'
- *gerak-gerik* → 'movements'
- *sayur-mayur* → 'all kinds of vegetables'

Keraf (1984) identified four types of reduplication in Indonesian based on the forms of repetition:

1. **Copy reduplication** → *meja-meja* ('tables')
2. **Partial reduplication** → *dedaunan* ('leaves')
3. **Affixed reduplication** → *menari-nari* ('dancing repeatedly')
4. **Ablaut reduplication** → *mondar-mandir* ('moving back and forth')

These classifications demonstrate the diverse ways in which reduplication manifests across languages, highlighting its significance in morphological structure.

3. Method

The function of this section is to describe all experimental procedures, including controls. The description should be complete enough to enable someone else to repeat your work. If there is more than one part to the experiment, describe the methods and present the results in the same order in each section. Decide what order of presentation will make the most sense to the readers.

This current study focuses on the patterns of ablaut reduplication in Toba Batak, specifically from a construction-based approach. This study applied a qualitative phenomenological case study method, which is scientifically advantageous for gaining a more in-depth comprehension of the patterns of ablaut reduplication in TB. Creswell and Poth (2017) argue that the phenomenological case study focuses on scrutinizing a language phenomenon through a group of individuals experiencing the phenomenon holistically. Bearing Sudaryanto's (1986) words in mind about how the condition and uniqueness of a language can be best known through the reality of the field where the language in question is studied, this method is also suitable because it approaches TB in a natural context (Djajasudarma, 1993), where the language pattern can be examined without the intervention of an experiment (Seliger, 1989).

3.1 Participants

This research was conducted in Samosir Regency, which consists of Pangururan District, Ronggur Nihuta District, and Harian Boho District. These locations were selected because they are part of *Tano Batak*, meaning "the Homeland of the Batak." This region is home to native Toba Batak (TB) speakers, where language use has remained relatively unaffected by external linguistic influences.

Given the homogeneity of the TB-speaking population, it was unnecessary to study all individuals. Considering the vast number of TB speakers, the extensive geographic coverage required, and the limitations of time, resources, and funding, this study employed a sampling approach. Following Hennink and Kaiser (2022), who reviewed 23 empirical qualitative studies, a sample size of 9 to 17 participants is considered ideal for a homogenous population with a narrowly defined objective. Samarin (1988, p. 52) outlined essential criteria for linguistic research participants, particularly for qualitative studies:

1) Must be native speakers of the language studied.
2) Born and raised in the research site.
3) Currently residing in the research site.
4) A fair representation of both male and female speakers.
5) Aged between 25 and 65 years, in good health, and not experiencing cognitive decline.
6) Adequately proficient in the language.
7) At least elementary school graduates, ideally junior high school graduates.
8) Physically and mentally healthy.
9) Have normal speech organs.
10) Patient and honest in using the language.
11) Willing to participate in the study.
12) Cooperative with the researchers throughout the process.

Since not all residents of Samosir Regency met these criteria, the study employed purposive and snowball sampling techniques. The first author, a native TB speaker, initially contacted TB-speaking acquaintances who met the criteria. These initial participants then recommended other eligible individuals. A total of twelve native TB speakers were recruited for the study. The group consisted of seven males and five females, ensuring representation across all three districts in Samosir Regency. To maintain ethical compliance, all participants remained anonymous to protect their identities. Their speech was recorded and observed, but any personal details embedded in their utterances were omitted. Participants were informed that only utterances containing ablaut reduplication would be analyzed and published.

3.2 Data Collection

This study desired verbal data on ablaut reduplication in TB as the primary data sources. Referring to Muhadjir (1989), the researcher ensured that the data collected is not artificial or manipulated. Semi-structured interviews for 25 to 30 minutes were carried out individually with the 12 participants. The verbal or oral data as primary data was recorded directly from TB speakers in real situations. This study also obtained secondary data from a visual compact disc, which contains video recordings of traditional TB wedding party between a couple participants who live in the same speech area, which were only used as quality assurance supplementary reference for the data analysis of the primary data.

3.3 Data Analysis

This study uses Halle's (1973) model of generative morphology as the foundation of the theoretical framework. This model is appropriate for this qualitative phenomenological case study since it can produce a richly comprehensive and intricate data description. However, each language has different rules in its morphological processes, and Dardjowidjojo (1983; 1988) has established very early that Halle's model can be problematic when applied to study Indonesian which is part of the Austronesian language family. This is further

validated by Luthfiani et al. (2020). After reading and understanding how the generative morphology theory of Aronoff (1976) and Scalise (1984), the researchers saw that Halle's model could not be applied in its entirety to dissect data or to analyze TB data where TB is also Austronesian group and for the specific purpose of analysing ablaut reduplication. Therefore, the researchers consider it necessary to modify Halle's generative morphology theory, creating a new model that still refers to Halle's model as the leading theory that would be suitable to analyse the rules of the morphological process of the TB. The modified theory is shown in Figure 2.

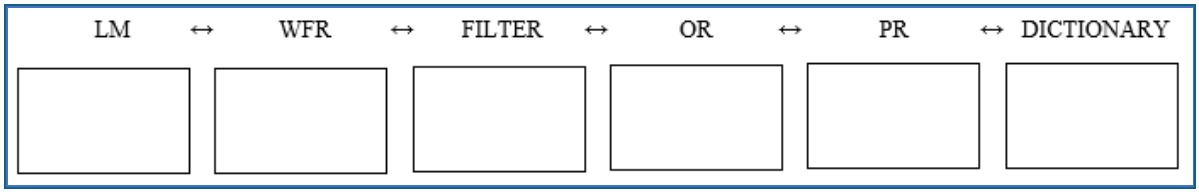


Figure 2. The Diagram of the Modified Halle's Theory Tailored to Toba Batak's Ablaut Reduplication

The modified model for studying TB within the framework of generative morphology theory consists of six components. Four of these components originate from Halle's model, namely the list of morphemes (LM), rules of word formation (WFR), filter, and dictionary of words. Two additional components, orthographic rules (OR) and phonological rules (PR), were introduced to address discrepancies between written and spoken forms in TB. Research by Ambarita (2018) and Purba et al. (2022) has shown that LM in TB includes free stems, word roots, bound forms, and four affix types, which are prefixes, suffixes, infixes, and confixes. Ambarita (2018) identified three WFR processes in TB, which consist of affixation, insertion of the premodifier *ni* between N and ADJ, and stress shift from the first to the second syllable in ADJ.

The study examined primary and secondary data to identify ablaut reduplication patterns by transcribing audio-recorded data into Microsoft Word, highlighting reduplicative forms in transcripts from twelve participants, and analyzing vowel and consonant order in repeated words. The researchers further reviewed transcriptions line by line to gain a comprehensive understanding (Ambarita, 2018; 2023) and the modified model to document single and multiple syllabic repetitions, vowel contrasts in height and backness, and the linear and relative ordering of vowels in reduplication. The findings are presented based on the model to provide an in-depth elaboration of the study's results.

4. Results

This study used a construction-based approach to study the ablaut reduplication in TB from the perspective of generative morphology. When registering the mophemes in the LM, this study found that the rule of vowel pattern of ablaut reduplication in TB dictates the order of vowels and consonants. For mono-syllabic words, the first vowel is almost always a [u] or an [e], and the second vowel is either an [a] or a [u], respectively. Meanwhile, for two-syllabic words, the vowel pattern of the first word is [u-a], and the vowel order of the second word is [a-I] or [a-u]. This indicates that the linear order of the vowels of ablaut reduplication is not fixed.

The data analysis yielded three categories of ablaut reduplication in TB, i.e. verbal ablaut reduplication (VAR), nominal ablaut reduplication (NAR), and adjectival ablaut reduplication (AAR). The findings for each category are disseminated in the following parts, where the process steps of ablaut reduplication refer to [[LM] ↔ [WFR] ↔ [OR] ↔ [PR] ↔ [DICTIONARY]] as the components of the modified generative morphology model for TB.

4.1 Verbal Ablaut Reduplication (VAR)

VAR which are found in TB modify the syntactic categories where the meaning of the first word is known and the meaning of the second word is known as well. Word reduplication with this phoneme variation generates verb (V) as presented below.

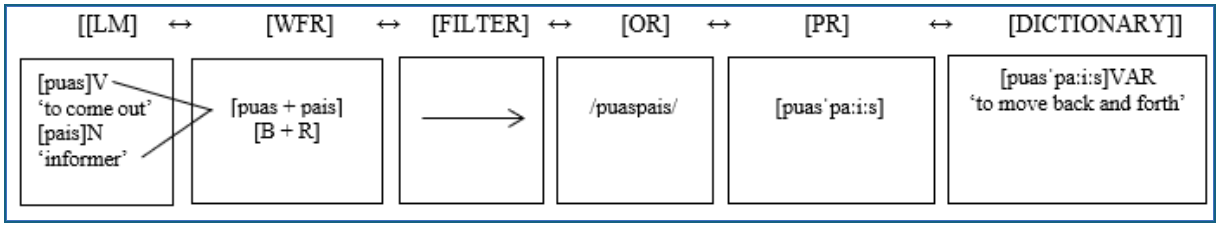


Figure 3. Data 1: *puaspais* [puasˈpa:i:s]

VAR [puasˈpaːiːs] is a reduplicative word i.e., a two-part word in which the first half [puas] is the base (B) and the second half [pais] is a repeat or rhyme as the reduplicant (R). For [puasˈpaːiːs] ablaut reduplication, the first vowels [u] and [a] in [puas] are low vowels, and the reduplicated ablaut variant of the first vowels are high vowels [a] and [i] in [pais]. Ablaut reduplication in [puasˈpaːiːs] involves a vowel shift from a short [u] and a short [a] in [puas] to a long [aː] and a long [iː] in [paːiːs].

The linear order of the vowels in [puasˈpaːiːs] are low vowels in the first element and high vowels in the second. The order of the B and the R in [puasˈpaːiːs] is from left i.e. *puas* [puas], to the right i.e. *pais* [paːiːs] as the R. Based on the data analysis above, it is very clear that phoneme change in reduplication is done by repeating syllable and making vowel sound vary from previous syllable.

VAR [puasˈpaːiːs] is formed from two different syntactic categories. The word *puas* [puas] 'to come out' is a V and *pais* [paːiːs] 'informer' is a N. The unique thing in this construction is that combining these two different word classes (V and N) generates [puasˈpaːiːs] as a VAR. Viewed from the original meaning, there is no meaningful relation between the meaning of the B [puas] and the meaning of the R [paːiːs] with the meaning of VAR [puasˈpaːiːs] as the product of ablaut reduplication. No grammatical meaning (GM) can be taken from [puasˈpaːiːs] because the VAR [puasˈpaːiːs] is the combination between two different B without affixation. The lexical meaning (LeM) of [puasˈpaːiːs] is 'to move back and forth'.

Based on the rule presented in the diagram above, the formation processes of VAR [puasˈpaːiːs] can be formulated as follows: [[LM] ↔ [WFR] ↔ [OR] ↔ [PR] ↔ [DICTIONARY]]. No process takes place in the filter in the formation of VAR [puasˈpaːiːs] because no idiosyncrasies are found, whether semantic idiosyncrasies, phonological idiosyncrasies, or lexical idiosyncrasies.

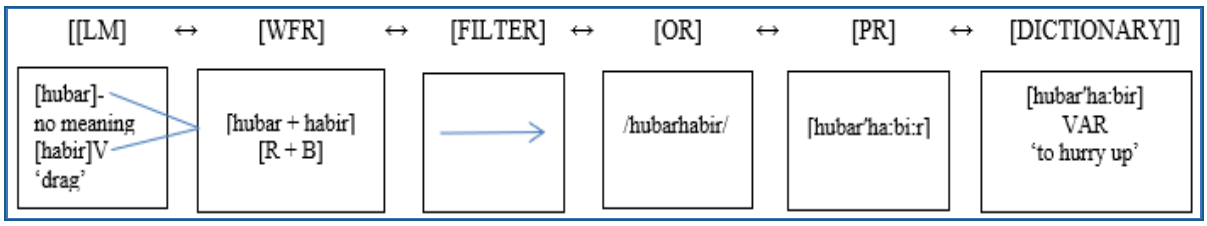


Figure 4. Data 2: *hubarhabir* [hubarˈhaːbiːr]

The VAR [hubarˈhaːbiːr] is a reduplicative word i.e., a two-part word in which the first half [hubar] is a repeat or rhyme as the R of the first [haːbiːr] as the B. In [hubarˈhaːbiːr], the first vowels [u] and [a] in [hubar] are low, and the reduplicated ablaut variant of the first vowels [a] and [i] are high vowels.

Ablaut reduplication in [hubarˈhaːbiːr] involves a vowel shift from a short [u] and a short [a] in [hubar] to a long [aː] and a long [iː] in [haːbiːr]. The linear order of the vowels in [hubarˈhaːbiːr] are low vowels in the first element and high vowels in the second. The order of the B and the R in [hubarˈhaːbiːr] is from left i.e. *habir* [haːbiːr] to the right i.e. [hubar] *hubar*. Repetition with change of vowels differs in the main form, the usual one in that it applies to the whole word, with the repeated word usually standing first (Tuuk, 1864).

VAR [hubarˈhaːbiːr] is formed from *hubar* [hubar] and *habir* [habir]. *Hubar* as a non-existence word in TB does not have meaning. *Habir* [habir] 'to drag' is a V. Viewed from the original meaning, there is no meaning relation between the meaning of the R [hubar] and the meaning of the B [habir] with the meaning of the VAR [hubarˈhaːbiːr] as the product of ablaut reduplication. While the LeM of [hubarˈhaːbiːr] is 'to hurry up', no GM can be taken because the VAR combines two different B without affixations.

- (**Data 3**) *rubasrabis* [rubasˈraːbiːs] 'to dangle and look like going to fall down' **rubasrubas*
- (**Data 4**) *bulangbaling* [bulaŋˈbaːliːŋ] 'to run helter-skelter' **bulangbulang*
- (**Data 5**) *mangabasmangebus* [maŋabasmaˈŋeːbuːs] 'to strike right and left' **mangabasmangabas*
- (**Data 6**) *rabarabu* [rabaˈraːbu] 'scattered (like corpses strewn about from both sides of wars)' **rabaraba*
- (**Data 7**) *talsatolsu* [talsaˈtoːlsu] 'to let go something which is bound' **talsatalsa*

Similarly to Data 1, no process takes place in the filter component for Data 2 due to the lack of idiosyncrasies, same with data (3-7) as well. It seesm that if the word to be repeated has an *a* in the penultimate syllable, this vowel then becomes the vowel in the ultimate syllable of the repeated word, whereas in the penultimate, it becomes a *u* (which can be an *o*), if the repetition of the word stands first as in (3) and (4). If the word that is being repeated has vowels other than *a* in the last two syllables, the vowels of these syllables in the first word must be *a* as presented in (5) (6), and (7).

4.2 Nominal Ablaut Reduplication (NAR)

Where VAR modifies the syntactic categories where the meaning of the first word is known and the meaning of the second word is known as well, NAR in TB modifies the syntactic categories where the meaning of both first and second words are not known. The following data presents the word reduplication with phoneme variation that generates V.

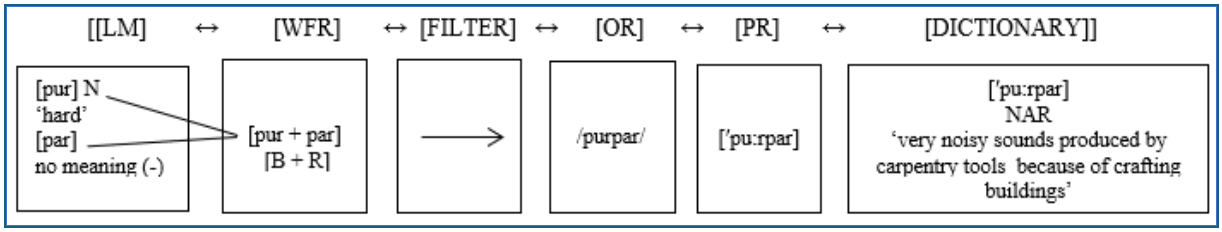


Figure 5. Data 8: *purpar* ['pu:rpar]

The vowel sounds in *purpar* ['pu:rpar] as an ablaut reduplication moves from a high front vowel [u:] in [pur] to a low back vowel [a] in [par]. To prove this, say ['pu:rpar] out loud and pay attention to where in your mouth you are making the vowel sound to find out how your organs of speech work to produce NAR ['pu:rpar]. In ['pu:rpar], the second half [par] is an R or rhyme of the first [pur]. The linear order of the vowel in ['pu:rpar] is a high vowel in the first element and a low vowel in the second. The order of the B and the R in ['pu:rpar] is from the right i.e. *pur* [pur] (as the B), to the left i.e. *par* [par] (as the R). NAR ['pu:rpar] is formed from the word *pur* [pur] and the form *par* [par]. The word [pur] as an ADJ means 'hard'. The latter is called form, not word, because [par] is not found in TB in isolation unless *par-* is, for example, used in word combination as in *parhoda* [parhoda]. The form [par-] in [parhoda] is a prefix where [par-] + *hoda* [hoda] 'horse' → *parhoda* [parhoda]. In this context, the GM of *par-* shows 'the owner of something stated in the base lexeme'. Therefore, the LeM of [parhoda] is 'the horse's owner'.

In NAR ['pu:rpar], [par] is not a suffix. Therefore, based on the non-existence word of the form [par] in isolation in TB, its syntactic category and its meaning cannot be identified. In other words, the B of the word repeated this way cannot be determined. Another variation of ['pu:rpar] in TB, which is very often used, is *parpur* ['pa:rpur]. Therefore, these two forms are interchangeable. It is very important to note that NAR ['pu:rpar] is derived from the onomatopoeia of artisans, such as the noisy sounds when people are building houses and other buildings.

In doing their activities, the craftsmen do their jobs with different roles; therefore, the sounds of their carpentries can be heard as [par...], [pur...], [par...], [pur...], and so on, and finally their activities result in noise. The LeM of ['pu:rpar] is 'very noisy sounds produced by carpentry tools because of crafting buildings'. This ablaut reduplication is also formed without affixations.

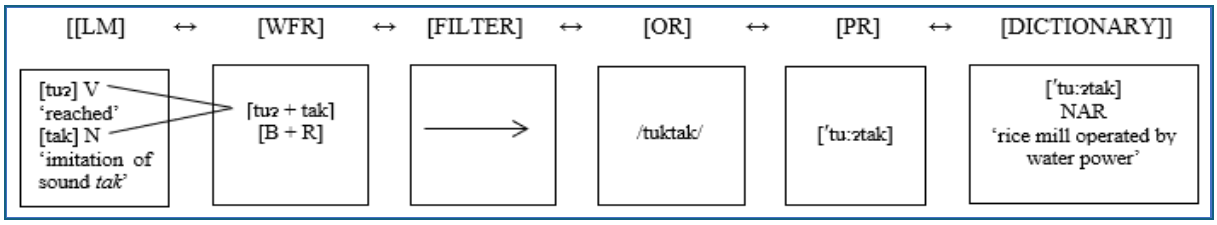


Figure 6. Data 9: *tuktak* ['tu:ʔtak]

NAR ['tu:ʔtak] is formed from two forms, i.e. the word *tuk* [tuk] and the form *tak* [tak]. The linear order of the vowels in ['tu:ʔtak] is a high vowel [u] in the first element [tuk] and a low vowel [a] in the second half [tak]. The order of the B and the R in ['tu:ʔtak] is from the right i.e. *tuk* [tuk] (as the B), to the left i.e. *tak* [tak] (as the R).

The original meaning that can be taken from the word [tuk] is 'reached'. However, the word ['tu:ʔtak] is unusual among its reduplicative kin because in TB society, it is believed that [tak] is the imitation of sound or onomatopoeia of rice mills when being operated and, therefore, produces the sound [tuk...tak...tuk...tak….], which means that there is no original meaning that can be taken from the form [tak]. This also means that the form [tak] is not found in TB, so its syntactic category and meaning cannot be determined. Without affixations, no GM can be drawn. These two forms generate a 'rice mill operated by water power' as its LeM, in which the syntactic category is N.

(**Data 10**) *hatahoti* [hataˈho:ti:] ‘gossip or scorn’ **hatahata*

If the word that is being repeated has vowels other than *a* in the last two syllables, the vowels of these syllables in the first word must be *a* as in data (10). The linear order of the vowels in [hataˈho:ti] is low vowels [a, a] in the first element [hata] and high vowels [o:] and [i:] in the second half [hoti]. The order of the B and the R in [hataˈho:ti:] is from the right, i.e. *hata* [hata] (as the B), to the left i.e., *hoti* [hoti] (as the R).

4.3 Adjectival Ablaut Reduplication (AAR)

AAR is a type of ablaut reduplication that specifically applies to adjectives. Unlike VAR and NAR, the meaning transparency of AAR can vary, with sometimes one component having a known meaning, while in other cases, neither may be clear.

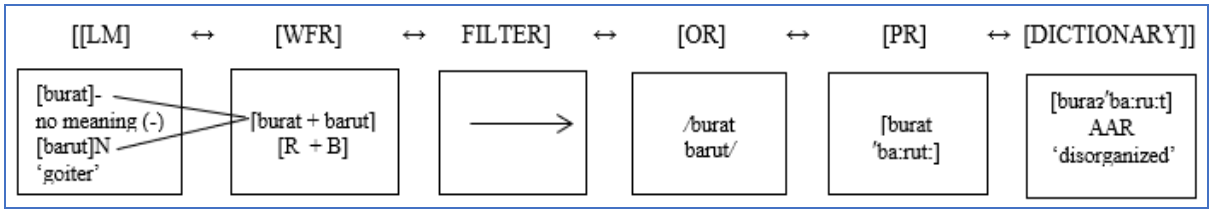


Figure 7. Data 11: *buratbarut* [buraʔˈba:ru:t]s

AAR [burʔˈba:ru:t], in which the syntactic category is ADJ, derives from the form [burat] as the R and [barut] as the B. The word *buratbarut* [buraʔˈba:ru:t] is a reduplicative word that consists of two-part words. The first half *burat* [burat] is a repeat or rhyme of the second half *barut* [barut]. The meaning of [burat] is unknown because the form [burat] is not found in TB. In [buraʔˈba:ru:t], the first vowels [u] and [a] in *burat* [burat] are low vowels. In [barut], the reduplicated ablaut variant of the first vowels shifts to high [a:] and high [u:]. The linear order of the vowels in [buraʔˈba:ru:t] is low vowels in the first element [burat] and high vowels in the second half [barut]. The order of the B and the R in [buraʔˈba:ru:t] is from the left, i.e. *barut* [barut] (as the B) to the right, i.e. *burat* [burat] (as the R), where the repeated word stands first as presented in the following diagram. This notion is consistent with Tuuk (1864) statement: Repetition with the change of vowels differs in the main form in that it applies to the whole word, with the repeated word usually standing first.

The form [burat] is a non-actual word in TB. Therefore, its syntactic category cannot be identified, given the symbol (-), and no meaning can be given, given the symbol (-). The second form is [barut] ‘goiter’ as a B in which the syntactic category is N. However, the combination of these two forms, [burat] and [barut], generates [burʔˈba:ru:t] as an AAR. Based on the original meaning, there is no meaningful relation between the form [burat] and [barut] with the meaning of [burʔˈba:ru:t] as an AAR generated by the two combinations. Besides, there is no GM that can be taken from [burʔˈba:ru:t] because it combines two forms without affixation. The LeM of [burʔˈba:rut] is ‘disorganized’. Like previous data, this word also did not have a filtered out idiosyncracy. The lack of filter seems to hold true for the remaining data found in this study.

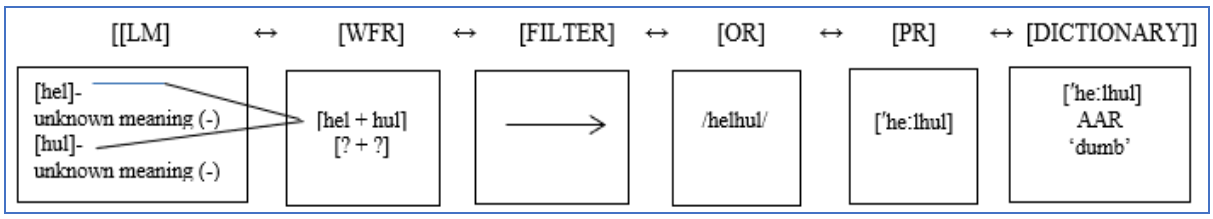


Figure 8. Data 12: *helhul* [ˈhe:lhul]

The vowel sounds in *helhul* [ˈhe:lhul] as an AAR involve a vowel shift from a high back vowel [e:] in the form *hel* [hel] to the low front vowel [u] in the form *hul* [hul]. The linear order of the vowels in [ˈhe:lhul] is a strong vowel in the first element and a weak vowel in the second half. The order of the B and the R in [ˈhe:lhul] cannot be determined because the forms [hel] and [hul] are non-actual words in TB. In other words, the B and the R cannot be identified in this manner.

Neither [hel] nor [hul] themselves are found in TB. Therefore, the syntactic categories and the meanings of [hel] and [hul] cannot be identified, and thus, the symbol (-) is given for the two forms. Surprisingly, these two non-actual words in TB can be combined to generate AAR [ˈhe:lhul] as an ADJ. It is unknown what onomatopoeia gives birth to the form [ˈhe:lhul]. The form [ˈhe:lhul] does not have GM; however, the LeM is ‘dumb.’

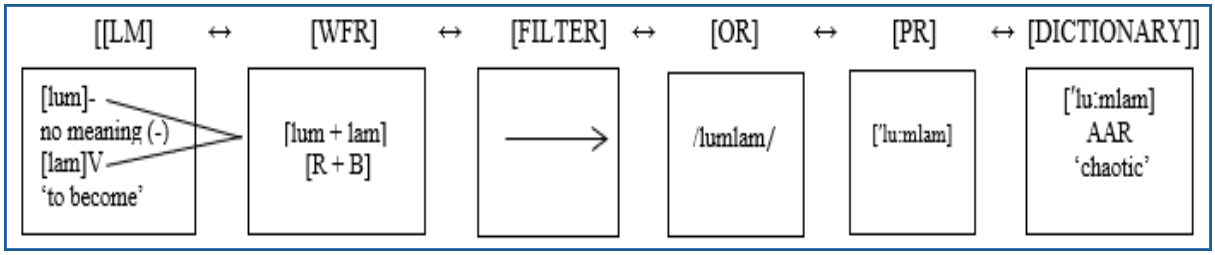


Figure 9. Data 13: *lumlam* ['lu:mlam]

AAR *lumlam* ['lu:mlam] is a reduplicative word that consists of two-part words. The first half *lum* [lum] is the R of the second half *lam* [lam]. The meaning of the first half *lum* [lum] as a repeated word is unknown because the form *lum* is not found in TB. In ['lu:mlam], the vowel [u:] in [lum] is high. In [lam], the reduplicated ablaut variant of the first vowel shifts to a low [a]. The linear order of the vowels in ['lu:mlam] is a high vowel in the first element and a low vowel in the second. The order of the B and the R in ['lu:mlam] is from left, i.e. *lam* [lam] (as the B), to the right, i.e. *lum* [lum] (as the R). In other words, the first half [lum] is a repeat or rhyme as the R of the second [lam].

AAR ['lu:mlam] with a phoneme change in which the syntactic category is ADJ is formed from the form *lum* [lum] and *lam* [lam]. Firstly, because of the non-actual word [lum] in TB, its syntactic category cannot be identified, therefore, given the symbol (-) and no meaning and, therefore, the symbol (-) as well. Secondly, the word [lam] means 'to become' as a V in isolation. The combination of the two forms [lum] and [lam] generates ['lu:mlam] as an AAR. No GM can be taken from ['lu:mlam] because the reduplication occurs without affixation. The LeM of ['lu:mlam] is 'chaotic.'

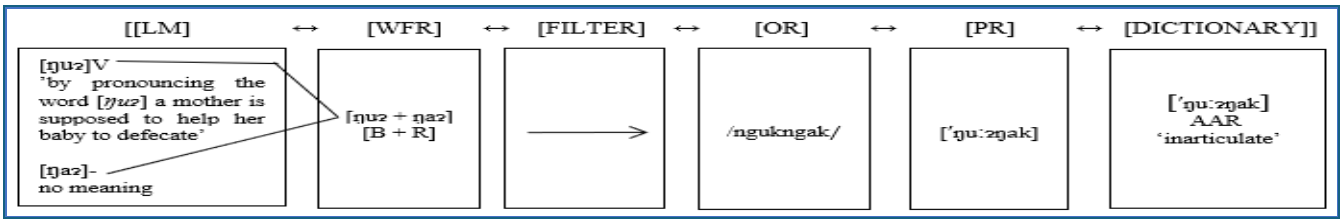


Figure 10. Data 14: *ngukngak* ['ŋu:ʔŋak]

AAR ['ŋu:ʔŋak] is formed from *nguk* [ŋuk] and *ngak* [ŋak]. The word [ŋuk] derives from *ngukнguk* ['ŋu:ʔŋuk] 'to defecate'. In TB traditions, especially those inland, ['ŋuʔŋuk] is pronounced by mothers to their babies repeatedly to help the baby defecate while the mother is accompanying her baby to defecate. The [ŋak] is a non-existence word in TB. TB speakers also use *ngaknguk* ['ŋa:ʔŋuk] as the variation of *ngukngak* ['ŋu:ʔŋak] that is by moving the B *nguk* [ŋuk] as the second element.

The vowel sounds in *ngukngak* ['ŋu:ʔŋak] as an AAR move from the high back vowel to the low front vowel of your mouth, i.e. from [u:] to [a] respectively. To prove this, say ['ŋu:ʔŋak] out loud and pay attention to where in your mouth you are making the vowel sound to find out how your organs of speech work to produce ablaut reduplication ['ŋu:ʔŋak]. The linear order of the vowels in ['ŋu:ʔŋak] is a high vowel in the first element and a low vowel in the second element. The order of the B and the R in ['ŋu:ʔŋak] is from the right, i.e. *nguk* [ŋuk] (as the B) to the left, i.e. *ngak* [ŋak] (as the R). In ['ŋu:ʔŋak], the second half [ŋak]] is a repeat or a rhyme of the first [ŋuk].

Viewed from the original meaning, there is no meaning relation between the meaning of each of the B [ŋuk] and [ŋak] with the meaning of the newly generated word ['ŋu:ʔŋak]. This word has no GM, but the LeM of ['ŋu:ʔŋak] is 'inarticulate', where someone is unable to speak distinctly or unable to speak clearly.

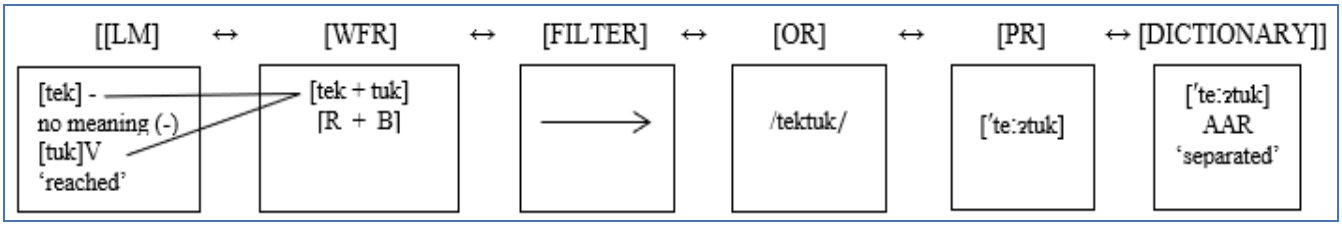


Figure 11. Data 15: *tektuk* ['te:ʔtuk]

AAR [ˈteːʔtuk] is formed from two different bases, i.e. *tek* [tek] and *tuk* [tuk]. The precategory [tek] is predicted to derive from the word *tektek* [ˈteʔtek] 'fall' as a V. The form [tek] in isolation does not have meaning unless another form is attached to it. The original meaning of [tuk] as a V is "reached". By this analogy, the combination of these two different forms generates [ˈteːʔtuk] as an AAR in TB. The vowel sounds in *tektuk* [ˈteːʔtuk] as an AAR involve a vowel shift from a high back vowel [e] in the form *tek* [tek] to the low front vowel [u] in the form *tuk* [tuk]. The first half [tek] is the R of the B [tuk]. The linear order of the vowels in [ˈteːʔtuk] is a high vowel in the first element and a low vowel in the second one. The order of the B and the R in [ˈteːʔtuk] is from the left i.e. *tuk* [tuk] (as the B) to the right i.e. *tek* [tek] (as the R). In other words, the R *tek* [tek] in [ˈteːʔtuk] stands first.

There is no meaningful relation between the AAR [ˈteːʔtuk] and the original meaning of the B *tuk* [tuk] as one of the elements to form [ˈteːʔtuk]. Probably, the form [ˈteːʔtuk] does not derive from either the form *tek* [tek] or the form t*uk* [took] at all. The LeM is 'separated' but it is not known what onomatopoeia generates [ˈteːʔtuk] either.

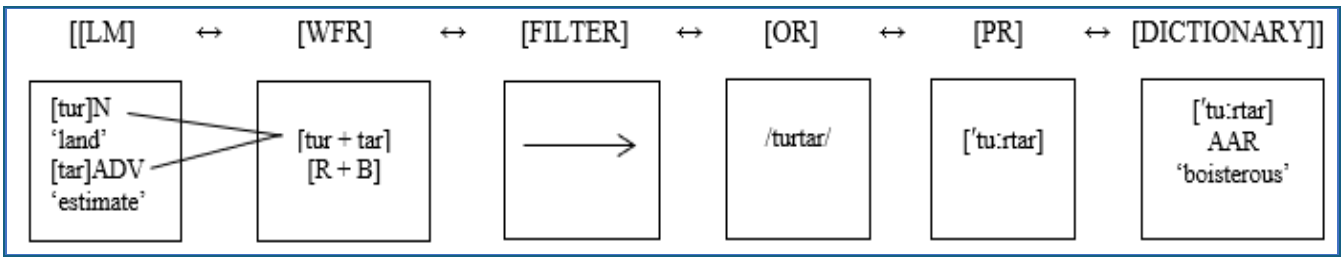


Figure 12. Data 16: *turtar* [ˈtuːrtar]

AAR [ˈtuːrtar] is formed from the R *tur* [tur] and the B *tar* [tar]. The first half *tur* [tur] is the R of the second *tar* [tar]. In [ˈtuːrtar], the first vowel [uː] in [tur] and the second vowel [a] in [tar] is a high vowel and a low vowel respectively. In [tur], the reduplicated ablaut variant of the first vowel shifts from a high long [uː] to a low short [a]. The linear order of the vowels in [ˈtuːrtar] is a high vowel in the first element and a low vowel in the second half. The order of the B and the R in [ˈtuːrtar] is from left i.e. *tar* [tar] (as the B) to the right i.e. *tur* [tur] (as the R).

The original meanings of [tur] and [tar] in TB are 'land' as N and 'estimate' as ADV of degree, respectively. Surprisingly, the combination of these two existence bases to generate AAR [ˈtuːrtar] does not have a meaningful relation with the R [tur] and the B [tar]. While there is no GM of [ˈtuːrtar] because there is no affixation process in the word formation, [ˈtuːrtar] does have LeM which is 'boisterous'.

- (**Data 17**) *burbar* [ˈbuːrbar] 'noisy' **burbarburbar*
- (**Data 18**) *hurhar* [ˈhuːrhar] 'broken' **hurharhurhar*
- (**Data 19**) *gurgar* [ˈguːrgar] 'damaged' **gurgargurgar*
- (**Data 20**) *sursar* [ˈsuːrsar] 'disorganized' **sursarsursar*
- (**Data 21**) *randatrundut* [raddatˈruːddut] 'chaotic' **randatrandat*
- (**Data 22**) *patarpotir* [patarˈpoːtir] 'very terrible' **patarpatar*
- (**Data 23**) *jabajobi* [jabaˈjoːbi] 'incoherent and untidy' **jabajaba*

Other examples of AAR in TB are presented in data (17-20). If the word that is being repeated has vowels other than *a* in the last two syllables, the vowels of these syllables in the first word must be *a*, as shown in data (21-23).

5. Discussion

This study aimed to investigate the patterns of ablaut reduplication in Toba Batak using Halle's (1973) model of the theory of generative morphology. With this modified model, this study found that the ablaut reduplications in TB are formed without affixations, so all the data lack any process taking place in the [FILTER] component. This means that the process steps of ablaut reduplication in TB only refer to [[LM] ↔ [WFR] ↔ [OR] ↔ [PR] ↔ [DICTIONARY]] as the components of generative morphology. This is an original knowledge on the morphological system of TB from this study thanks to its construction-based approach following the suggestions from Minkova (2002) and Weijer et al. (2020).

McCarthy (2004) states that in reduplication, each of the elements of the base (B) has a corresponding part in the reduplicant (R). This is linear with the findings of this study because the morphological process that occurs in ablaut reduplication in TB was done by repeating the base with sound changes. In other words, phoneme change in ablaut reduplication is done by repeating syllables and making vowel sounds vary from the previous syllable. However, combining the two correspondent existence bases to generate ablaut reduplication in TB does not seem to have a meaningful relation between the word repeated and the base. While reduplication is a productive morphological element in the Indo-European languages that can be used as N, ADV, V, and particles (Giannakis, 1992), this pattern is actually less valid in the ablaut reduplication in TB. As an Austronesian language, ablaut reduplication in TB is unproductive and is neither found in adverbs nor particles in TB, unlike Giannakis' discoveries. Instead, it is only found in three categories.

The first category, verbal ablaut reduplication (VAR), only applies to verbs where the meaning of both the first and second words is known (Crystal, 1993). The first vowel of the VAR can be high, and the reduplicated ablaut variant of the first vowel is a low vowel. This study found that VAR in TB can involve a vowel shift from a short [u] and a short [a] to a long [a:] and a long [i:]. As for the linear order of the vowels in VAR in TB, a high vowel in the first element and a low vowel in the second element, in other words, VAR involves a vowel shift from a short to a long vowel. This pattern is in line with Mathiassen's (1969) studies on Celtic VAR.

However, the order of the B and the R in VAR is from left to right, which differs from Wivell's (2024) investigation who concluded that Austronesian langauges, including TB, has a low vowel – high vowel property in ablaut reduplication. The different interpretation can be attributed to the fact that Wivell only investigated a very limited number of data in TB since they tried to cover 26 language families, so this study's interpretation regarding TB specifically is more valid. The interview data revealed that the linear order of the vowels in TB ablaut reduplication is actually not fixed, as sometimes a low vowel can be in the first element and a high vowel in the second and vice versa. The results of this study also support Durrel's (1975) and Jasanoff's (2007) research findings in Germanic dialects that the strong verb is formed almost exclusively by vocalic alternation or the ablauting strong verbs. The data shows that VAR in TB is the combination of two different word classes (V and N), where the first syllable is stronger than a short one.

The second category, nominal ablaut reduplication (NAR), only modifies nouns where the meaning of the both the first and second words is typically unknown. Some forms of the ablaut reduplication in TB are not found in isolation; therefore, its syntactic category and meaning cannot be determined and some of the data for NAR vary where the two forms are interchangeable. This is clear from data that were derived from onomatopoeia. This empirical statement aligns with what Körtvélyessy (2020) states about how every linguistic tradition's concept of onomatopoeia takes into account its relation to the language phenomenon and its place in the system of a language.

The third category, adjectival ablaut reduplication (AAR), applies to adjectives and the meaning of hte individual components may or may not be known. When the form is a non-actual word in TB, its syntactic category cannot be determined. Surprisingly, combining the two correspondent existence bases to generate AAR does not have a meaningful relation with the B and R. This finding affirms McCarthy's (2004) observation on how B elements have corresponding parts in R. If the word being repeated has vowels other than *a* in the last two syllables, the vowels of these syllables in the first word must be *a*. This is attributed to external factors such as the speaker's emotions, since it has been described by Hickey (2014) that emotions can determine the rhyme-motivated sound or flexible stress pattern of ablaut reduplication in TB. Marchand (1957) and Bauer (2006) have also asserted that the words in AAR belong to rhyme-motivated with the term echo word being a less technical label with a flexible stress pattern that can be influenced by external factors, for instance, by contact with other ethnic groups.

The modification of Halle's (1973) generative morphology model for TB has proven to be successful in uncovering nuances of ablaut reduplication in TB. This study found that one of the most notorious problems in ablaut reduplication is the creation of the ablauting strong verbs from earlier reduplicating verbs. Viewed from the original meaning, there is no meaningful relation between the meaning of the B and the R and the meaning of the generated ablaut reduplication.

Therefore, the LeM of the VAR, NAR, and AAR cannot be determined based on one or two of the elements and no GM can be interpreted because the words were combined without affixations. This might be something relatively unique in TB because other languages' ablaut reduplication involve ablauting an affix, such as the Greek's ablauting the infix **-né-* in the singular active *-n-* (Weiss, 2010). This lack of GM also indicates that the phoneme change in TB ablaut reduplication is formed by repeating syllables and making vowel sounds vary from the previous syllable, not by adding affixes. By applying and modifying Halle's model, this current research revealed that no process occurs in the filter in forming VAR, NAR, and AAR because no semantical, phonological, or lexical idiosyncrasies are found.

Regarding the vowels and consonants in ablaut reduplication in this study, their historical origins determine the position of long and short vowels and consonants in words. This study has not found that long stop vowels commonly represent the reflexes of the tenues, which supports Vertegaal's (2020) that long stop vowels commonly represent the reflexes of the Proto-Indo-European tenues, whereas short stop vowels generally show the Proto-Indo-European aspirate. Minkova (2002) proposes her arguments for short and long vowels as trochaic contours where she states that ablaut reduplication has characteristics of trochaic contour; therefore, there will be a huge number of independence due to the word stress in the second half of the neologism. This finding also confirms Padgett's (2011) statement that the empirical territory is not simple and attempts to understand consonant-vowel place interactions are still a much-unresolved debate.

Many of the past studies exploring the langauge of TB people seems to have mostly examined a linguistic aspect within a specific piece of the culture, such as the traditional oral rhymes *Umpama* and *Umpasa* (Siahaan & Barus, 2022; Sitorus & Lubis, 2023), a folktale called *Sitagran Bulu* (Simaremare et al., 2023), the *Poda* book (Nadeak, 2023), or the *Sauar Matua* death ceremony (Ginting et al., 2023; Tampubolon et al., 2024). One of the earliest reference on the repetition in TB is published in Tuuk (1864), who states that ablaut reduplication in TB differs in the main form from the usual one in that it applies to the whole word, with the repeated word usually standing first. Tuuk's statement

does not align with the findings of this study which found that the repeated word stands first or second. In other words, the reduplicant or the repeated words cannot be determined based on their positions. This difference somewhat indicates the complexity of ablaut reduplication in TB, but more studies will need to be done on TB to be certain. Unfortunately, only two other studies on reduplication in TB could be found in public databases, which are Ambarita (2018) and Ambarita (2023).

The results of this study are significantly different from Ambarita (2018) that examined adjective reduplication with a structural approach, identifying that TB speakers either fully or partially reduplicate adjectives. While this previous study managed to identify and group the various prefixes, infixes, suffixes and confixes that are involved in repetition of adjectives, this study enhanced the knowledge on AAR's word formation rules. The result of this study also reinforces the findings of Ambarita (2023) which had also applied Halle's model to explore base full reduplication in TB. Base full reduplication involves the exact repetition of a morpheme (e.g. *bye-bye, goody-goody*), which means no vowel changes occurred in this type of reduplication.

In TB, Ambarita (2023) found six types of base full reduplication. On the other hand, the present study focused on ablaut reduplication where the repetition involved a vowel change (e.g. *zig-zag, tick-tock*), and thus contributed original findings with its identification of three types of ablaut reduplication in TB.

The examination of ablaut reduplication and its implications for vowel patterns provides significant insights into linguistic structures, particularly in syllabic arrangements. In single-syllable words, the first vowel is typically "u" or "e," while the second vowel often includes "a" or "u," a pattern aligning with prior research on vowel prominence and phonetic function (Wivell et al., 2024; Minkova, 2002). In two-syllable constructs, the first word follows a "u-a" sequence, while the second adopts "a-i" or "a-u," reinforcing the role of phonetic transitions in speech segmentation (Wivell et al., 2024). When deviations occur, such as the absence of "a" in the last two syllables of the repetition, the first word compensates by maintaining an "a" in the corresponding position, a phenomenon linked to phonological constraints on reduplicative patterns (Minkova, 2002).

Moreover, if an "a" appears in the penultimate syllable of the original word, it prescriptively shifts to match the ultimate syllable in the reduplicated form, underscoring the dynamic interaction between phonological structures and morphological constraints. These findings suggest that ablaut reduplication is not merely a lexical feature but a reflection of systematic phonetic and phonological principles that shape pronunciation and comprehension. By demonstrating how vowel quality in specific syllabic positions affects word recognition and segmentation, this research enhances our understanding of linguistic rhythm and flow in spoken language (Wivell et al., 2024; Minkova, 2002).

In contrast, in the penultimate, it becomes a *u* (which can be an *o*), if the repetition of the word stands first. This finding is consistent with Guardiano (2021), who found that the orthographic rules are relatively abstract constructs, adapted to better reflect word formation sequences rather than recent debatable word similarities. Evidence from consonant-vowel-consonant interactions in TB's ablaut reduplication supports this, as the rule dictates the order of vowels and consonants in repeated words. Repeated words cannot be considered natural phenomena since language, as a dynamic communication tool, evolves diachronically and synchronically. Through neologisms, new words, diverse speakers, and shifting contexts, language continuously changes, requiring users to be more selective in their diction based on temporal and situational contexts.

Language experts over the years have continuously try to address the need to write the unwritten grammar rule of ablaut reduplication. This study is another step forward to realize this mission, but it can only be accomplished through executing more studies on ablaut reduplication in other (preferably) regional languages that will not only uncover unique insights but also contribute to the linguistic treasures as one way to ensure the survival of ethnic and endangered languages.

6. Conclusions

This study has made significant contributions to the understanding of ablaut reduplication in Toba Batak by identifying three clear categories: verbal ablaut reduplication, nominal ablaut reduplication, and adjectival ablaut reduplication. Each category exhibits systematic vowel alternations that operate without affixation, which means they bypass the filter component in Halle's generative morphology model. The findings reveal that Toba Batak's reduplication patterns differ markedly from those in Indo-European languages, as they are unproductive and derive meaning solely through phonological changes rather than grammatical markers.

The adaptation of Halle's model to include orthographic and phonological rules represents an important methodological innovation for analyzing non-concatenative morphology in Austronesian languages. This approach successfully addresses a notable gap in linguistic typology while providing a framework that could be applied to other understudied languages. Beyond theoretical implications, these findings have practical value for language preservation efforts, particularly for endangered languages like Toba Batak, by documenting previously unrecorded grammatical structures. For future research, it would be valuable to apply this construction-based approach to other regional languages within the Austronesian family. Additionally, incorporating sociolinguistic and experimental methods could help explore how these reduplicative patterns evolve over time, vary across dialects, and are processed cognitively by speakers. Such investigations would not only deepen our understanding of morphological diversity but also strengthen global efforts to preserve linguistic heritage.

References

Ambarita, E. (2018). Morphological analysis of adjective reduplications in Toba Batak language. *Jurnal Penelitian Pendidikan dan Bahasa Sastra*, *2*(1), 130–137. https://doi.org/10.32696/ojs.v2i1.157

Ambarita, E. (2023). Base full reduplication in Toba Batak: A study of generative morphology. *Majalah Ilmiah Methoda*, *13*(1), 72–81. https://doi.org/10.46880/methoda.Vol13No1.pp72-81

Arbi, A. P., Sulistyaningsih, S., Syakur, A., & Lestariningsih, L. (2022). Derivational and inflectional morphemes on the thesis abstracts. *Jurnal Basicedu*, *6*(3), 3895–3907. https://doi.org/10.31004/basicedu.v6i3.2659

Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press.

Bauer, L. (2006). Compounds and minor word-formation types. In B. Aarts & A. McMahon (Eds.), *The handbook of English linguistics* (pp. 483–506). Blackwell. https://doi.org/10.1002/9781119540618.ch23

Blust, R. (2022). The Neogrammarian hypothesis and pandemic irregularity: Take two. *Journal of Historical Linguistics*, *12*(2), 167–193. https://doi.org/10.1075/jhl.20027.blu

Blust, R. (2023). Repair and drift in Austronesian languages: Avoidance of dissimilar labials as the onsets of successive syllables. *Oceanic Linguistics*, *62*(2), 304–323. https://doi.org/10.1353/ol.2023.a913563

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper and Row.

Creswell, J. W., & Poth, C. N. (2017). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage.

Cristy, S., Lubis, F., & Chunliu, L. (2023). The comparison of verb affixations contrastive between Batak Toba and Indonesia language. *Tradition and Modernity of Humanity*, *3*(1), 1–11. https://doi.org/10.32734/tmh.v3i1.11448

Crystal, D. (1993). *An encyclopedic dictionary of language and languages*. Blackwell.

Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Blackwell.

Curto, J. (2024, May 7). Why you can't stop saying 'That's that me, espresso'. *Vulture*.

Damanik, S., & Mulyadi, M. (2020). Ergativity case-marking in Batak Toba language. *Budapest International Research and Critics in Linguistics and Education Journal*, *3*(1), 80–87. https://doi.org/10.33258/birle.v3i1.755

Dance, R. (2019). *Words derived from Old Norse in Sir Gawain and the Green Knight: An etymological survey*. Blackwell.

Dardjowidjojo, S. (1983). Beberapa Masalah dalam Teori Morfologi Jeneratif: Suatu Kasus dalam Pembentukan Kata Kerja in Beberapa Aspek Linguistik Indonesia. *Seri ILDEP, Jakarta: Penerbit Djambatan*.

Dardjowidjojo, S. (1988). *PELLBA I: pertemuan Linguistik Lembaga Bahasa Atma Jaya, pertama*. Lembaga Bahasa Unika Atma Jaya.

Djajasudarma, T. (1993). *Metode linguistik: Ancangan metode penelitian dan kajian*. Eresco.

Downing, L. J., & Inkelas, S. (2015). What is reduplication? Typology and analysis part 2/2: The analysis of reduplication. *Language and Linguistics Compass*, *9*(12), 516–528. https://doi.org/10.1111/lnc3.12150

Durrel, M. (1975). Reduplication and ablaut in the Germanic strong verb. *German Life and Letters*, *29*(1), 48–59. https://doi.org/10.1111/j.1468-0483.1975.tb00146.x

Giannakis, G. (1992). Reduplication as a morphological marker in the Indo-European languages: Reduplicated presents. *WORD*, *43*(2), 161–196. https://doi.org/10.1080/00437956.1992.12098298

Ginting, F. S., Malau, J. R. T., Miranti, M., & Silaban, R. E. (2023). Morphological analysis in Batak Toba language phrases: Death ceremony system. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, *11*(2), 1620–1625. https://doi.org/10.62568/ella.v2i1.122

Green, V. V. (2016). *Native and non-native intuitions on the phonology of binomial locutions* (Doctoral dissertation). University of Texas.

Guardiano, C., Longobardi, G., Cordoni, G., & Crisma, P. (2021). Formal syntax as a phylogenetic method. In R. D. Janda, B. D. Joseph, & B. S. Vance (Eds.), *The handbook of historical linguistics* (Vol. II, pp. 141–172). Wiley. https://doi.org/10.1002/9781118732168.ch7

Halle, M. (1973). Prolegomena to a theory of word formation. *Linguistic Inquiry*, *4*(1), 3–16.

Haspelmath, M., & Sims, A. D. (2010). *Understanding morphology* (2nd ed.). Oxford University Press.

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science & Medicine*, *292*, 114523. https://doi.org/10.1016/j.socscimed.2021.114523

Hickey, R. (2014). *A dictionary of varieties of English*. Wiley. https://doi.org/10.1002/9781118602607

Hornby, A. S. (1987). *Oxford advanced learner's dictionary of current English* (3rd ed.). Oxford University Press.

Jackendoff, R. (1975). Morphological and semantic regularities in the lexicon. *Language*, *51*(3), 639–672. https://www.jstor.org/stable/412891

Jasanoff, J. H. (2007). From reduplication to ablaut: The class VII strong verbs of Northwest Germanic. *Historische Sprachforschung*, *120*, 241–284.

Jensen, J. T. (1995). *Morfologi struktur kata dalam nahu generatif*. Dewan Bahasa dan Pustaka.

Jespersen, O. (1965). *Modern English grammar on historical principles: Part VI, morphology*. George Allen & Unwin.

Keraf, G. (1984). *Tata bahasa Indonesia*. Nusa Indah.

Körtvélyessy, L. (2016). Reduplication from a cross-linguistic perspective. *British and American Studies*, *22*, 153–164.

Körtvélyessy, L. (2020). Onomatopoeia unique species? *Studia Linguistica*, *74*(2), 506–551. https://doi.org/10.1111/stul.12133

Lǐ, Y., & Ponsford, D. (2018). Predicative reduplication: Functions, their relationships and iconicities. *Linguistic Typology*, *22*(1), 51–117. https://doi.org/10.1515/lingty-2018-0003

Lipka, L. (1975). Prolegomena to a theory of word formation: A reply to Morris Halle. In E. Koerner (Ed.), *The transformational-generative paradigm and modern linguistic theory* (pp. 157–180). John Benjamins. https://doi.org/10.1075/cilt.1.09lip

Lubis, D. F., & Bowo, T. A. (2022). Batak Toba script, preserving its authenticity in globalization stream. *Nusantara Science and Technology Proceedings*, 28–34. http://nstproceeding.com/index.php/nuscientech/article/view/564

Luthfiani, E., Wagiran, W., & Rustono, R. (2020). Meaning patterns and potential forms of Halle deadjektival verbs model in generative Indonesian morphology. *Seloka: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, *9*(1), 45–52. https://doi.org/10.15294/seloka.v9i1.36496

Manzini, M. R. (2021). Morphology as syntax, features as categories, functional cascades. *Isogloss*, *7*(1), 1–8. https://doi.org/10.5565/rev/isogloss.191

Marchand, H. (1975). Motivation by linguistic form: English ablaut and rime combinations and their relevancy to word formation. *Studia Neophilologica*, *47*(1), 54–66. https://doi.org/10.1080/00393275708587147

Mathews, P. (1974). *Morphology: An introduction to the theory of word structure*. Cambridge University Press. https://doi.org/10.1017/CBO9781139166486

Mathiassen, T. (1969). On the problem of lengthened ablaut degree and the Slavic sigmatic aorist. *Scando-Slavica*, *15*(1), 201–214. https://doi.org/10.1080/00806766908600522

Mattes, V. (2017). Iconicity in the lexicon: The semantic categories of lexical reduplication. *Studies in Language*, *41*(4), 813–842. https://doi.org/10.1075/sl.16024.mat

McCarthy, A. C. (1992). *Current morphology*. Routledge. https://doi.org/10.4324/9780203206720

McCarthy, J. J. (2004). *Optimality theory in phonology: A reader*. Blackwell. https://doi.org/10.1002/9780470756171

McLendon, L. (2020). Ablaut reduplication is tip top. *ACES*.

Minkova, D. (2002). Ablaut reduplication in English: The criss-crossing of prosody and verbal art. *English Language and Linguistics*, *6*(1), 133–169. https://doi.org/10.1017/S1360674302000206

Muhadjir, N. (1989). *Metode penelitian kualitatif*. Rake Sarasin.

Nadeak, K. L. (2023). The types of clitics in Toba Batak: A study on book of Poda in the Bible. *METHOLANGUE: Language Teaching and Literature, Linguistics and Literature*, *8*(2), 35–58.

Nida, E. (1949). *Morphology: The descriptive analysis of words*. University of Michigan Press.

Padgett, J. (2011). Consonant–vowel place feature interactions. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (Vol. 3, pp. 1437–1465). Wiley. https://doi.org/10.1002/9781444335262.wbctp0075

Penney, J. H. (1977). The treatment of Indo-European vowels in Tocharian. *Transactions of the Philological Society*, *75*(1), 66–91. https://doi.org/10.1111/j.1467-968X.1977.tb00352.x

Purba, D. F., Malau, F., Siahaan, M. L., & Napitupulu, S. (2022). A contrastive analysis between English and Batak Toba language in verbal affixes. *Journal of Humanities, Social Sciences and Business*, *1*(4), 121–128. https://doi.org/10.55047/jhssb.v1i4.288

Puspani, I. A. M., & Indrawati, N. L. K. M. (2021). Challenges in translating Indonesian reduplication into English. *Journal of Language and Linguistic Studies*, *17*(4), 1973–1983.

Rajagukguk, R., Malems, T., & Ginting, M. (2022). Preserving linguistic diversity: Strategies for language preservation and cultural heritage safeguarding. *Jurnal Ilmu Pendidikan dan Humaniora*, *11*(3), 221–236. https://doi.org/10.35335/jiph.v11i3.25

Regjer, T. (2022). Reduplication and the arbitrariness of the sign. In *Proceedings of the twentieth annual conference of the Cognitive Science Society* (pp. 887–892). Routledge. https://doi.org/10.4324/9781315782416-161

Samarin, W. J. (1988). *Ilmu bahasa lapangan* [Field linguistics]. Penerbit Kanisius.

Samsuri. (1982). *Analisis bahasa: Memahami bahasa secara ilmiah* [Language analysis: Understanding language scientifically]. Erlangga. https://opac.perpusnas.go.id/DetailOpac.aspx?id=38298

Saragih, D. (2021). An analysis of metaphor in Batak Toba song lyrics (mother love's theme). *Budapest International Research and Critics Institute Journal, 4*(2), 2013–2021. https://doi.org/10.33258/birci.v4i2.1889

Scalise, S. (1984). *Generative morphology*. Foris. https://doi.org/10.1515/9783112328033

Seliger, H. W., Shohamy, E. G., & Shohamy, E. (1989). *Second language research methods*. Oxford University Press.

Shariq, M. (2020). Tools and techniques used in the language of advertisements: The linguistic perspective. *Media Watch, 11*(3), 565–580. https://doi.org/10.15655/mw/2020/11092020

Siahaan, J., & Barus, A. (2022). Form, content of Umpama and Umpasa in Batak Toba language: Study of local wisdom. *Talenta Conference Series: Local Wisdom, Social, and Arts, 5*(3), 51–61.

Simaremare, J. A., Padang, S., Sinaga, A. K., & Sagala, N. E. (2023). Analisis kategori kelas kata pada cerita rakyat "Sitagan Bulu" yang berasal dari Batak Toba [Analysis of word class categories in the folklore "Sitagan Bulu" from Toba Batak]. *IdeBahasa, 5*(2), 310–318. https://doi.org/10.37296/idebahasa.v5i2.149

Sinaga, A. B. (2002). *Tata bahasa Batak Toba: Meresapkan jiwa dan darah Batak* [Toba Batak grammar: Imbuing the soul and blood of Batak]. Bina Media Perintis. https://books.google.co.id/books/about/Tata_bahasa_Batak_Toba.html

Sirulhaq, A., Sukri, S., Jafar, S., & Burhanuddin, B. (2022). Potential words in Indonesian language: A study of generative morphology. *Lingua Cultura, 16*(2), 231–240. https://doi.org/10.21512/lc.v16i2.8674

Sitanggang, A., Murni, S., & Sibarani, B. (2024). Textual meaning on "umpasa" Batak Toba in wedding ceremony. *International Journal of Religion, 5*(8), 513–523. https://doi.org/10.61707/7aqmtd83

Sitorus, N., & Lubis, T. (2023). The lexicon on natural knowledge of Umpasa in Batak Toba. *Jurnal Teknik Kimia USU, 3*(1), 28–34. https://doi.org/10.32734/tmh.v3i1.11451

Sudaryanto. (1986). *Metode linguistik* [Linguistics methodology]. Duta Wacana University Press.

Suhadi, J. (2018). The adequacy of grammatical metaphor to account for world languages. *The 1st Annual International Conference on Language and Literature*, 1–14.

Tampubolon, M., Rakasiwi, R., Silalahi, M. M. R., & Siahaan, R. A. E. (2024). How do figurative languages reflect the cultural values in Toba Batak Saur Matua ceremony. *Cendikia: Media Jurnal Ilmiah Pendidikan, 14*(3), 364–369.

Templeton, S. (2012). Teaching and learning morphology: A reflection on generative vocabulary instruction. *The Journal of Education, 192*(2–3), 101–107.

ten Hacken, P. (2020). Classical generative morphology. In *Oxford research encyclopedia of linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.575

Thun, N. (1968). *Reduplicative words in English: A study of formations of the types tick-tick, hurly-burly and shilly-shally*. ProQuest.

Tuuk, V. H. (1864). *A grammar of Toba Batak*. Martinus Nijhoff. https://doi.org/10.1007/978-94-017-6778-1

Vertegaal, A. (2020). Syllable weight gradation in the Luwic languages. *Transactions of the Philological Society, 118*(2), 211–236. https://doi.org/10.1111/1467-968X.12186

Wallace, T. (2019). *A study in flim-flam: Reduplicative ablaut ideophones in English*. Lund University.

Weijer, J., Weiyun, W., Yumeng, W., Ren, G., & Ran, Y. (2020). Words are constructions, too: A construction-based approach to English ablaut reduplication. *Linguistics, 58*(6), 1701–1735. https://doi.org/10.1515/ling-2020-0169

Weiss, M. (2010). *A companion to the Ancient Greek language*. (E. J. Bakker, Ed.). Wiley. https://doi.org/10.1002/9781444317398

Wekker, H., & Haegeman, L. (1996). *A modern course in English syntax*. Routledge. https://doi.org/10.4324/9780203133439

Wivell, G. B., Miatto, V., Ji, J., Karakas, A., Kostyszyn, K., & Repetti, L. (2024). All about ablaut: A typology of reduplicative vowel change. *Department of Linguistics Stony Brook*.

Yip, M. (2009). Repetition and its avoidance: The case in Javanese. *Arizona Phonology Conference, 5*, 238–262.