JoAL
Journal of Applied Linguistics

JOURNAL OF APPLIED LINGUISTICS
Vol. 6 No. 1, July 2026
e- ISSN 2809 - 1922 , p-ISSN 2809 - 1914
Available online at:
https://journal.eltaorganization.org/index.php/joal/index

Morphophonemic Allomorphy of the Prefix *Um-* In Toba Batak: An Extended Generative Morphology Analysis

Esron Ambarita[1*], Edi Suprayetno[2], Atabong Alemanjoh Mariana Nkea [3], Emmanuel Kyei[4]
[1]Fakultas Sastra Universitas Methodist Indonesia, Indonesia
[2]Sekolah Tinggi Keguruan dan Ilmu Pendidikan Al Maksum Langkat, Indonesia
[3]University of Yaoundé I, Cameroon
[4]Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development, Ghana

Correspondence Email : *esronambarita@gmail.com

ARTICLE INFO

Article History:
Received : April 3, 2026
Reviewed : May 10, 2026
Revised : May 12, 2026
Accepted : June 17, 2026
Available online : July 1, 2026

Keywords:

Morphophonemic allomorphy; Toba Batak language; Generative morphology; Prefix *um-*; Austronesian linguistics

Abstract

This study investigates the morphophonemic allomorphy of the prefix *um-* in Toba Batak and evaluates the applicability of an extended generative morphology framework for analyzing Austronesian morphological systems. The data were collected from twelve native speakers residing in Samosir and Toba Samosir Regencies through elicitation, observation, and semi-structured interviews. The study employs an expanded version of generative morphology derived from Halle's model, incorporating the components List of Morphemes, Word Formation Rules, FILTER, DICTIONARY, Orthographic Rules, and Phonological Rules. This extended framework was designed to accommodate the distinctive morphological and phonological characteristics of Toba Batak, which are insufficiently represented in models originally developed for Indo-European languages. The findings identify four morphophonemic allomorphs of the prefix *um-*, namely *umm-*, *um-*, *un-*, and *uŋ-*, each conditioned by systematic phonological environments and morphophonemic processes. The analysis demonstrates that the prefix exhibits high morphological productivity in forming grammatical and lexical meanings through predictable sound alternations. The study contributes to generative morphology by extending Hallean analysis to an underrepresented regional language while also supporting language documentation, revitalization, and preservation efforts for Toba Batak. Furthermore, the research provides empirical evidence that extended generative morphology can effectively explain morphophonemic variation in minority languages, offering a replicable analytical approach for future studies on indigenous languages across Austronesian linguistic communities.

INTRODUCTION

Language functions as the primary medium for human interaction and meaning transmission across social and cultural contexts. Morphology studies how morphemes combine to form words and encode grammatical and lexical meanings (Nida, 1949; Katamba, 1993; Malmkjaer, 1995). Among

Published by English Lecturers and Teachers Association (ELTA)

DOI: 10.52622/joal.v6i1.694

morphological phenomena, morphophonemic allomorphy is theoretically significant because it reveals systematic interactions between morphological processes and phonological environments. Prefix alternations are especially important in understanding how languages regulate sound changes in word formation, particularly in agglutinative and Austronesian languages. However, in Toba Batak, morphophonemic behavior remains largely descriptive and has not been formally explained within a generative morphology framework.

Batak, an Austronesian language of North Sumatra, has a productive verbal affixation system. Previous studies mainly address cultural traditions, oral literature, kinship systems, and general grammar (Sinaga, 2002). While some works mention Toba Batak morphology and Austronesian morphophonemics (Blust, 2023; Luthfiani et al., 2020), none has systematically analyzed verb prefix alternation using a formal generative approach. As a result, phonological conditioning in verbal prefix variation, particularly of *um-*, remains under-explained.

The prefix *um-* is central in Toba Batak verb formation due to its productivity and systematic surface variations, including *um-*, *umm-*, *un-*, and *uŋ-*. These alternations provide a clear domain for analyzing interactions between assimilation, nasal adjustment, and phonological strengthening, making *um-* a key entry point for morphophonemic investigation.

Theoretically, this study extends Halle's (1973) generative morphology model by incorporating Orthographic Rules and Phonological Rules alongside morpheme lists, Word Formation Rules, FILTER, and DICTIONARY components. This extension is required to capture the full derivational process in Toba Batak verb formation. This study aims to analyze the morphophonemic allomorphy of *um-* and explain how phonological environments determine its surface realizations within an extended generative framework. It focuses on rule-governed alternations among *um-*, *umm-*, *un-*, and *uŋ-*.

This research contributes in three ways: (1) providing a formal generative account of *um-* allomorphy, (2) operationalizing an extended Hallean model for Austronesian morphophonemics, and (3) presenting empirical data from speakers in Samosir and Toba Samosir, confirming that the alternations are phonologically conditioned.

Beyond theoretical contribution, this study also supports documentation of Toba Batak, which faces language shift pressures toward Indonesian (Lubis & Bowo, 2022; Purba et al., 2002), by formalizing its morphophonemic system.

REVIEW OF RELATED LITERATURE

Halle's (1973) generative morphology framework explains word formation via rule-based processes and filtering mechanisms generating actual and potential forms (Nida, 1949; Chomsky & Halle, 1968; Haspelmath et al., 2010). It highlights morphology's productivity, deriving multiple words from a single root. Language competence allows intuitive manipulation of word forms, central to generative grammar (Diver, 1969). Toba Batak has been studied extensively in this framework (Ambarita et al., 2026, 2025a, 2025b, 2025c, 2025d, 2025e; Ambarita, 2024; 2023a; 2023b; 2022a; 2022b; 2021a; 2021b; 2020; 2018a, 2018b, 2018c, 2018d, 2018e; 2017a; 2017b; Panggabean et al., 2025). Generative rules capture transformations in language systems (Sigurd, 1966), and Halle's model applies broadly to morpheme combinations (Aronoff, 1976; Scalise, 1984). Contemporary views treat it as a computational system with morphemes as fundamental units (Manzini, 2021).

1. List of Morphemes

List of Morphemes (LM), the building blocks of words, include free and bound forms and are divided into lexical bases and affixes. In some languages, affixes encode multiple meanings and selectional relations among lexical items are semantically motivated. This study focuses on verbs as core lexical units. Unlike agglutinative languages, verb segmentation in Kiranti languages is complex

(Driem, 2011), and support-verb constructions illustrate verb-noun interactions (Fendel, 2025). Affixes are categorized into derivational and inflectional types (Anderson, 1982; Stump, 2001).

2. Morphophonemic Allomorphs in Word Formation Rules

Following LM, Word Formation Rules (WFR) govern linguistic category construction (Crystal, 2008), ensuring morpheme combinations yield grammatical forms (Aronoff, 1976; Scalise, 1984). Affixation often triggers phonological alternations—nasal and vowel changes, phoneme insertion/deletion—producing morphophonemic allomorphs (Schane, 1973). Sound changes like assimilation, epenthesis, and haplology are widely attested. For example, the negative morpheme *in-* becomes *im-* in *impossible* due to bilabial influence but remains unchanged in *incongruous*, reflecting assimilation to adjacent phonemes (Nida, 1949).

3. Filter

The FILTER functions as a selection mechanism that determines whether a given word form is acceptable according to phonological, syntactic, and semantic criteria. Essentially, it evaluates whether the form conforms to the general rules of the language or if it violates them. Forms that fail to meet these criteria are either discarded or flagged as exceptional. Importantly, the filter also handles idiosyncratic deviations that cannot be accounted for by general grammatical or phonological rules. These include irregular verb forms, rare derivational patterns, or semantic anomalies. By managing such exceptions, the filter plays a dual role: it screens out ill-formed or unacceptable forms while preserving information about irregular but legitimate lexical items.

In a generative morphology framework, the filter operates after morphophonemic and syntactic processes have applied, ensuring that only forms consistent with the language's structural and functional constraints are allowed to progress. This step is crucial for maintaining the systematicity of word formation while also accommodating the richness and variability of natural language.

4. Dictionary

The DICTIONARY represents the final stage in the lexical processing pipeline. It serves as a repository or inventory of all word forms that have successfully passed through earlier stages, including the WFR and the filter. This lexicon contains both base forms (root words) and derived forms (words generated through morphological processes), provided they meet the phonological, syntactic, and semantic criteria. Each dictionary entry includes not only the form itself but also its meaning, grammatical properties, and any unique characteristics. This ensures that the lexicon is both comprehensive and precise, allowing the system to retrieve word forms efficiently for production, comprehension, or further morphological operations.

METHODS

This study adopts a qualitative phenomenological case study design to examine the morphophonemic variation of the prefix *um-* in Toba Batak verbs. The approach is designed to capture naturally occurring linguistic patterns as they are realized in everyday speech and ritual discourse.

Data were collected from six subdistricts in the Samosir and Toba Samosir regions, namely Pangururan, Ronggur Nihuta, Harian Boho, Ajibata, Balige, and Borbor. These locations were selected due to their strong maintenance of indigenous Toba Batak linguistic practices. Twelve native speakers participated in semi-structured interviews lasting approximately 25–30 minutes. To

strengthen data reliability, supplementary data were obtained from video recordings of traditional ceremonial events. All participants were anonymized to ensure ethical compliance.

Analytical Framework

The study employs an extended generative morphology framework originally developed by Halle for analyzing English as a language within the Indo-European family. However, in this research, the framework is systematically modified to accommodate the structural characteristics of Toba Batak, a language belonging to the Austronesian family. The modification is necessary because Halle's original model does not explicitly account for orthographic variation and certain phonologically conditioned alternations that are prominent in Austronesian morphophonemic systems. Therefore, two additional components are introduced:

- Orthographic Rules (OR)
- Phonological Rules (PR)

These components are integrated into the analytical architecture as follows:

List of Morpheme (LM) → Word Formation Rules (WFR) → Filter → Orthographic Rules (OR) → Phonological Rules (PR) → Dictionary

Derivation of OR and PR

OR and PR were not pre-assumed but were derived empirically from the dataset. The process began with systematic identification of surface realizations of the prefix *um-* across different phonological environments. Orthographic patterns were established from consistent written forms found in interview transcriptions and ceremonial discourse recordings. Phonological Rules were subsequently derived by analyzing recurrent sound alternations triggered by the initial phoneme of the base verb. These alternations were observed to follow predictable patterns that could not be fully explained by the original Halle framework, thus justifying the inclusion of PR as an explicit component in the modified model.

Inductive Rule Formation

The formulation of rules in this study follows an inductive procedure. Rather than applying pre-established theoretical rules, generalizations were extracted from repeated morphological and phonological patterns found in the data. These patterns were then formalized into WFR, OR, and PR, ensuring that the analysis remains data-driven and empirically grounded.

Validation of Forms

The validity of derived forms and rules was ensured through a multi-layered validation process. First, cross-speaker comparison was conducted across the twelve participants to confirm consistency of morphological patterns. Second, triangulation with recorded ceremonial discourse was used to verify the natural occurrence of the identified forms in authentic communicative settings. Third, internal rule consistency was tested by evaluating whether the proposed WFR, OR, and PR could systematically generate attested surface forms without producing contradictory outputs. Only rules that passed all validation stages were retained in the final extended generative morphology model.

FINDINGS

The prefix *um-* in Toba Batak exhibits systematic allomorphic variation conditioned by the initial segment of the base lexeme. Each of the four allomorphs undergoes a uniform three-stage

derivational process: place assimilation, nasal gemination in homorganic clusters, and compensatory gemination. The four allomorphs are:

1. *um-* →*umm-* (before vowels [a], [e], [i], [o], [u]). Vowel-initial bases trigger gemination (*umm-*), reflecting homorganicity and ensuring morphological stability while optimizing phonetic output.
2. *um-*remains *um-* (before bilabial consonants [b], [p], [m]). The nasal remains /m/, interacting with the base consonant through gemination and feature redistribution to produce stable phonetic forms.
3. *um-* → *un-* (before alveolar consonants [d], [t], [n], [j], [l], [r]). The nasal shifts /m/ → /n/, followed by homorganic cluster strengthening and compensatory gemination, maintaining articulatory efficiency and morphophonemic cohesion.
4. *um-* → *uŋ-* (before velar consonants [g], [h], [ŋ]; with /h/ → [k]). The nasal adjusts /m/ → /ŋ/, with subsequent gemination and compensatory processes producing consistent and predictable surface forms.

Across allomorphs, OR and PR operate sequentially, ensuring compliance with phonological, syntactic, and semantic constraints before entries are stored in the DICTIONARY. These findings show that Toba Batak morphology–phonology interaction is highly structured, systematic, and adaptive, with allomorphic variation arising from the dynamic interplay of morphological identity and phonological optimization in active verb formation as discussed in more details in the following parts.

1. [um-] → [umm-] with Vowel-Initial Bases

The derivational pathway from base lexeme to phonetic realization can be illustrated in a pipeline format:

Table 1. Morphophonemic Derivation of the Prefix *um-* with Vowel-Initial Bases

Prefix + Base	Orthographic Rule	Phonological Rule	Meaning
um- + adong	/ummadong/	[ummadoŋ]	more suitable
um- + ellep	/ummellep/	[ummellep]	less clear
um- + ila	/ummila/	[ummila]	shier
um- + olo	/ummolo/	[ummolo]	more willing

Observations show that OR ensures consistent intermediate forms, allowing the prefix to interact with the base without disrupting morphological structure. PR executes phonological realization, adjusting segments to the phonetic context. Here, the bilabial nasal in *um-* remains intact, reflecting efficient articulatory structure.The pipeline demonstrates that complex forms emerge from dynamic interactions between morphological and phonological components. OR determines the derivational path, while PR ensures alignment with articulatory constraints, providing a conceptual framework applicable to other Toba Batak lexemes, including cases of place assimilation, nasal reduction, and compensatory gemination.

a. Place of Articulation Assimilation of the Prefix *um-* with Vowel-Initial Bases

Place assimilation transforms the bilabial nasal /m/ in *um-* into an alveolar nasal /n/ when followed by base-initial alveolar consonants /d/, /t/, /l/, or /r/ (Halle, 1993), reflecting homorganicity for efficient speech.However, when the base begins with a vowel, /m/ remains unchanged as displayed below:

Table 2. Place Assimilation of the Prefix *um-* with Vowel-Initial Bases

Prefix + Base	Orthographic Rule	Meaning
um- + adong	/ummadong/	more suitable

um- + ellep	/ummellep/	less clear
um- + ila	/ummila/	shier
um- + olo	/ummolo/	more willing

These data indicate that assimilation is context-dependent. OR regulates morphophonemic operations, producing stable intermediate forms foundational for subsequent phonological processes, such as nasal reduction and compensatory gemination. The /m/ → /n/ shift occurs only with compatible alveolar bases, showing systematic, predictable patterns.

b. Nasal Gemination in Homorganic Clusters of the Prefix *um-* with Vowel-Initial Bases

Following place assimilation, nasal gemination occurs in homorganic clusters. The nasal /m/ doubles before vowels, particularly in rapid speech, to maintain morphological integrity and facilitate smooth articulation. PR governs this compensatory process: the geminated nasal strengthens the cluster and preserves morphological cohesion. Formally:→ /m/ → m / __ mm (in homorganic clusters)

Table 3. Nasal Gemination in Homorganic Clusters of the Prefix *um-* with Vowel-Initial Bases

Prefix + Base	Orthographic Rule	Phonological Rule	Meaning
um- + adong	/ummadong/	[ummadoŋ]	more suitable
um- + ellep	/ummellep/	[ummellep]	less clear
um- + ila	/ummila/	[ummila]	shier
um- + olo	/ummolo/	[ummolo]	more willing

The gemination occurs only in vowel-initial bases, with no phoneme reduction. OR sequences ensure correct morphophonemic operations, while PR manages articulatory adjustments. This strengthens the prefix's morphemic identity and prepares for compensatory gemination.

c. Compensatory Gemination of the Prefix *um-* with Vowel-Initial Bases

Compensatory gemination occurs after nasal gemination, redistributing articulatory features of the doubled nasal to the following segment. This preserves morphological cohesion and ensures articulatory fluency. Formally, this transformation transfers [+nasal, +continuant] features from pre-vocalic nasal segments to the subsequent segment, producing lengthened consonants:

Table 4. Compensatory Gemination of the Prefix *um-* with Vowel-Initial Bases

Nasal + Vowel Cluster	Phonological Rule	Output
/m/ + /a/	[mma]	ummadong → [ummadoŋ]
/m/ + /e/	[mme]	ummellep → [ummellep]
/m/ + /i/	[mmi]	ummila → [ummila]
/m/ + /o/	[mmo]	ummolo → [ummolo]

This process systematically strengthens segments while preserving the [+nasal] feature. The final phonetic forms remain aligned with their morphological representations, optimizing articulatory efficiency. Within the generative framework, *um-* variation—whether /um-/ or geminate /umm-/—constitutes an allomorphic manifestation produced by sequential interactions between morphology and ordered phonological rules.Compensatory gemination confirms the prefix's grammatical function

as an active verb marker while demonstrating Toba Batak's phonological flexibility. It ensures both morphological integrity and articulatory efficiency.

2. [um-] remains [um-] with Bilabial-Initial Bases

In bilabial environments ([b], [p], and [m]), the prefix *um-* exhibits allomorphic variation conditioned by the initial segment of the base lexeme. A prominent pattern involves consonant strengthening via gemination, resulting from interactions between the nasal segment of *um-* and the base-initial consonant. Formally, this can be expressed as:→ /m/ + C → [CC] (within homorganic or articulatorily compatible environments). The derivational pathway is illustrated in a pipeline format as follows:

Table 5. Morphophonemic Derivation of the Prefix *um-* with Bilabial-Initial Bases

Prefix + Base Lexeme	Orthographic Rule	Phonological Rule	Meaning
/um- + bahen/	/umbahen/	[ubbahen]	to make
/um- + puhut/	/umpuhut/	[uppuhut]	more diligent
/um- + mora/	/ummora/	[ummora]	wealthier

OR generates intermediate forms (/umbahen/, /umpuhut/, /ummora/) where the prefix *um-* is morphologically preserved without changing its place of articulation. Unlike alveolar or velar contexts, progressive assimilation does not occur in bilabial environments, maintaining morphological stability. PR applies segmental strengthening via gemination. In /umbahen/ → [ubbahen] and /umpuhut/ → [uppuhut], the nasal /m/ transfers its articulatory features to the following consonant, producing geminates [bb] and [pp]. In /ummora/ → [ummora], /m/ is fully retained as [mm]. This demonstrates that gemination arises either directly from morphophonemic interaction or via nasal reduction followed by compensatory strengthening.

This process does not involve explicit nasal deletion as in alveolar or velar clusters; instead, articulatory features such as [+nasal] and [+closure] are redistributed, yielding stable and efficient forms. The pipeline demonstrates the dynamic interaction between OR and PR: OR establishes structural configurations, and PR optimizes phonetic realization through gemination. These patterns are systematic, predictable, and adaptive to bilabial contexts, illustrating the prefix *um-*'s morphological and phonological integration.

a. Place of Articulation Assimilation of the Prefix *um-* with Bilabial-Initial Bases

In bilabial environments, the /m/ in *um-* does not shift to a velar articulation as in /g/ or /k/ contexts. It retains bilabial articulation while triggering strengthening of the following consonant (/b/ or /p/). This partial assimilation preserves bilabial homorganicity and produces intermediate forms like /umbahen/, /umpuhut/, and /ummora/.

Table 6. Place of Articulation Assimilation of the Prefix *um-* with Bilabial-Initial Bases

Prefix + Base Lexeme	Orthographic Rule	Phonological Rule	Meaning
/um- + bahen/	/umbahen/	[ubbahen]	to make
/um- + puhut/	/umpuhut/	[uppuhut]	more diligent
/um- + mora/	/ummora/	[ummora]	richer

No segmental substitution occurs; articulatory consistency at the bilabial level is maintained. Place assimilation does not entail radical changes but stabilizes features, supporting subsequent phonological processes such as gemination. This demonstrates a tightly coupled morphology–phonology system: the prefix retains its identity while engaging the phonological system to produce optimal phonetic outputs.

b. Nasal Reduction in Homorganic Clusters of the Prefix *um-* with Bilabial-Initial Bases

PR governs nasal reduction within homorganic clusters (/mb/, /mp/, /mm/). In these clusters, /m/ is not always fully realized as an independent segment. Instead, it undergoes phonetic reduction while its articulatory features are transferred to the following consonant. Formally:→ /m/ → Ø / __ C [+bilabial]. Here, Ø indicates the absence of explicit phonetic realization, not absolute deletion; [+nasal] features remain distributed within the segmental structure.

Table 7. Nasal Reduction in Homorganic Clusters of the Prefix *um-* with Bilabial-Initial Bases

Prefix + Base Lexeme	Orthographic Rule	Phonological Rule	Meaning
/um- + bahen/	/umbahen/	[ubbahen]	to make
/um- + puhut/	/umpuhut/	[uppuhut]	more diligent
/um- + mora/	/ummora/	[ummora]	richer

In /mb/ and /mp/ clusters, the nasal strengthens the following consonant ([bb], [pp]). In /mm/, the nasal remains as a stable geminate. Nasal reduction is context-sensitive and preserves segmental features while optimizing articulatory efficiency.

c. Compensatory Gemination of the Prefix *um-* with Bilabial-Initial Bases

Compensatory gemination transforms nasal + bilabial consonant clusters into geminate consonants. The bilabial nasal /m/ weakens but does not disappear; its features ([+nasal], [+closure]) are redistributed to the following consonant.

Table 8. Compensatory Gemination of the Prefix *um-* with Bilabial-Initial Bases

Nasal + Consonant Cluster	Phonological Rule	Transformation
/m/ + /b/	[bb]	/umbahen/ → [ubbahen]
/m/ + /p/	[pp]	/umpuhut/ → [uppuhut]
/m/ + /m/	[mm]	/ummora/ → [ummora]

Gemination occurs systematically before bilabial consonants. Nasal reduction in /mb/ and /mp/ strengthens the following consonant, whereas /mm/ preserves the geminate. Features are reorganized rather than deleted, confirming a transformational (not eliminative) process. [+nasal] features are preserved through increased duration or intensity, demonstrating systematic and predictable phonological patterns.

The prefix *um-* in Toba Batak exhibits stable bilabial retention in bilabial environments. Phonological processes involve nasal reduction and compensatory gemination rather than place assimilation. These processes redistribute features within the segmental structure, ensuring both morphological cohesion and articulatory efficiency. The interaction of OR and PR in this extended generative morphology framework produces predictable allomorphic outcomes in bilabial contexts, demonstrating the dynamic and adaptive nature of Toba Batak phonology. The systematic gemination

patterns highlight how the language maintains morphological integrity while optimizing articulation, providing a clear model of morphology–phonology interaction.

3. [um-] → [un-] with Alveolar-Initial Bases

The bilabial nasal /m/ of the prefix *um-* assimilates to an alveolar nasal /n/ in the context of alveolar-initial base forms ([d], [t], [n], [j], [l], [r]). Within the extended generative morphology framework adapted from Halle (1973) and extended with OR and PR, complex forms in Toba Batak unfold through systematically ordered derivational stages. OR governs morphophonemic sequencing, while PR realizes phonological output. The final surface form emerges from the interaction of morphemic structure and ordered phonological rules. The following pipeline illustrates these stages:

Table 9. Morphophonemic Derivation of the Prefix *um-* with Alveolar-Initial Bases

Prefix + Base	Orthographic Rule	Phonological Rule	Meaning
um- + dangol	/undangol/	[uddaŋol]	more miserable
um- + tigor	/untigor/	[uttigor]	straighter
um- + neang	/unneang/	[unneaŋ]	lighter
um- + jotjot	/unjotjot/	[ujjoʔjot]	more frequent
um- + lambas	/unlambas/	[ullabbas]	wider
um- + roa	/unroa/	[urroa]	worse

a. Place of Articulation Assimilation of the Prefix *um-* with Alveolar-Initial Bases

At this stage, the bilabial nasal /m/ of *um-* assimilates to the alveolar nasal /n/ before alveolar-initial bases, producing forms such as /undangol/, /untigor/, /unneang/, /unjotjot/, /unlambas/, and /unroa/.

Table 10. Place of Articulation Assimilation of the Prefix *um-* with Alveolar-Initial Bases

Prefix + Base	Orthographic Rule	Meaning
um- + dangol	undangol	more difficult
um- + tigor	untigor	straighter
um- + neang	unneang	lighter
um- + jotjot	unjotjot	more frequent
um- + lambas	unlambas	wider
um- + roa	unroa	worse

This reflects homorganicity, where segments align articulatorily for efficiency. OR establishes structural conditions, formally: → /um-/ → /un-/ / __ [+alveolar] /m/ → /n/ occurs consistently before alveolar consonants, preserving morphological integrity while aligning the phonological structure.The shift is not purely phonetic but a morphophonemic strategy mediating between morphological stability and phonological harmony. This stage prepares the form for subsequent nasal reduction and compensatory gemination.

b. Nasal Reduction in Homorganic Clusters of the Prefix *um-* with Alveolar-Initial Bases

PR governs nasal reduction in homorganic clusters such as /nd/, /nt/, /nn/, /nj/, /nl/, and /nr/. Here, the alveolar nasal /n/ from the previous assimilation is phonetically weakened but not

eliminated; its articulatory features are transferred to the following consonant. Formally: → /n/ → Ø / __ C (within homorganic clusters)

Table 11. Nasal Reduction in Homorganic Clusters of the Prefix *um-* with Alveolar-Initial Bases

Prefix + Base	Orthographic Rule	Phonological Rule	Meaning
um- + dangol	undangol	[uddaŋol]	more difficult
um- + tigor	untigor	[uttigor]	straighter
um- + neang	unneang	[unneaŋ]	lighter
um- + jotjot	unjotjot	[ujjoʔjot]	more frequent
um- + lambas	unlambas	[ullabbas]	wider
um- + roa	unroa	[urroa]	worse

/n/ is not fully realized phonetically but contributes to strengthening the following consonant, producing [dd], [tt], [nn], [jj], [ll], and [rr]. Nasal reduction functions as a transitional stage linking homorganic cluster formation with compensatory gemination.

c. Compensatory Gemination of the Prefix *um-* with Alveolar-Initial Bases

Compensatory gemination transforms reduced nasal + consonant clusters into geminate consonants. The features of the weakened /n/ are reassigned to the following consonant, producing segmental strengthening or lengthening.

Table 12. Compensatory Gemination of the Prefix *um-* with Alveolar-Initial Bases

Nasal + Consonant Cluster	Phonological Rule	Transformation
/n/ + /d/	[dd]	undangol → [uddaŋol]
/n/ + /t/	[tt]	untigor → [uttigor]
/n/ + /n/	[nn]	unneang → [unneaŋ]
/n/ + /j/	[jj]	unjotjot → [ujjoʔjot]
/n/ + /l/	[ll]	unlambas → [ullabbas]
/n/ + /r/	[rr]	unroa → [urroa]

Gemination occurs systematically whenever a homorganic nasal is weakened. [+nasal, +closure] features are redistributed rather than deleted, ensuring phonological cohesion and articulatory fluency. OR and PR interact dynamically to maintain morphological and phonological integrity. Apparent segment deletion is better understood as feature redistribution, not elimination.The alternation /um-/ → /un-/ in Toba Batak before alveolar-initial bases illustrates a layered morphophonemic process.

Within the extended generative framework incorporating OR and PR, these interactions produce systematic and predictable allomorphic outcomes without altering the grammatical function of the prefix as a marker of active verb formation. The alternation exemplifies Toba Batak's dynamic interplay between morphology and phonology, balancing articulatory efficiency, structural cohesion, and morphological identity.

4. [um-] → [uŋ-] with Velar-Initial Bases

The bilabial nasal /m/ in the prefix *um-* undergoes place assimilation, surfacing as the velar nasal /ŋ/ when attached to velar-initial bases such as /g/, /ŋ/, and historically /h/-initial roots. This alternation reflects a systematic application of homorganic assimilation, in which nasal segments adjust their place of articulation to match the following consonant in order to optimize articulatory efficiency and maintain phonological cohesion. Within the Extended Generative Morphology framework (Halle, 1973), these derivations are regulated through ordered interactions between OR and PR. OR determines the structural configuration of the derived form, while PR executes phonological adjustments that yield the surface realization of the allomorph [uŋ-].

A critical refinement in this environment concerns the historical and phonological behavior of /h/-initial bases. The observed /h/ → [k] alternation is not an arbitrary phonetic shift but reflects a phonological strengthening process conditioned by velar adjacency. Specifically, in Toba Batak phonotactics, /h/ in morpheme-initial position before affixation may undergo fortition to [k] in derived environments, particularly when followed by nasal assimilation processes that favor velar convergence. This indicates that the process is phonetically motivated rather than random substitution. The transformation aligns with a broader tendency in Austronesian languages toward articulatory reinforcement in weak consonantal positions, especially in morphologically complex constructions.

Regarding productivity, the /h/ → [k] alternation is best characterized as lexically semi-restricted rather than fully productive. It does not apply uniformly across all /h/-initial lexical items but is conditioned by specific morphological environments in derivational contexts involving the prefix *um-*. This suggests that while nasal place assimilation (m → ŋ) is fully productive and rule-governed, the /h/ strengthening process operates at the interface of phonological rule application and lexical specification.

The distinction is important for understanding the morphophonemic architecture of Toba Batak: nasal assimilation represents a general phonological rule, whereas /h/ strengthening reflects a restricted rule sensitive to lexical and morphological conditioning.

Concerning the forms marked with an asterisk (*), these are used to indicate unattested, marginal, or theoretically derived forms that are not directly observed in the elicited corpus but are predicted by the rule system. Their inclusion serves a diagnostic function within the generative framework, demonstrating the predicted scope of the phonological rules even in the absence of attested usage. These forms are therefore not treated as erroneous, but as analytically necessary representations for testing rule consistency within the OR–PR derivational model.

Overall, the interaction between OR and PR in velar environments confirms that morphophonemic output in Toba Batak is not arbitrary but structurally constrained. The consistent emergence of [uŋ-], together with conditioned velar strengthening and rule-governed predictive forms, supports the view that Toba Batak phonology operates through highly organized and hierarchically structured derivational mechanisms.

Table 13. Morphophonemic Derivation of the Prefix *um-* with Velar-Initial Bases

Prefix + Base	Orthographic Rule	Phonological Rule	Meaning
um- + gurgur	/unggurgur/	[uŋgurgur]	*more boiling
um- + holong	/ungkolong/	[ukkolong]	*more loving
um- + ngalut	/ungngalut/	[uŋŋalut]	more difficult

a. Place of Articulation Assimilation of the Prefix *um-* with Velar-Initial Bases

At this stage, the bilabial nasal /m/ in *um-* assimilates to the velar nasal /ŋ/ when preceding velar-initial bases such as /g/, /k/ (from /h/), or /ŋ/. Intermediate forms like /unggurgur/, /ungkolong/, and /ungngalut/ illustrate this alignment. This process is motivated by homorganicity, where segments align their place of articulation with adjacent segments for articulatory efficiency.

Table 14. Place of Articulation Assimilation of the Prefix *um-* with Velar-Initial Bases

Prefix + Base	Orthographic Rule	Meaning
um- + gurgur	unggurgur	*more boiling
um- + holong	ungkolong	*more loving
um- + ngalut	ungngalut	more difficult

Direct production of /m/ before velars is articulatorily costly; converting /m/ → /ŋ/ resolves this mismatch while preserving morphological integrity.Within the OR–PR framework, OR ensures that /m/ → /ŋ/ occurs systematically before phonological realization. The assimilation rule is formally:→ /um-/ → /uŋ-/ / __ [+velar]

The shift occurs consistently in velar environments, independent of lexical idiosyncrasies. Place assimilation mediates between morphology and phonology, ensuring cohesion between structural and phonetic representation.

b. Nasal Reduction in Homorganic Clusters of the Prefix *um-* with Velar-Initial Bases

Following assimilation, PR governs nasal reduction within velar homorganic clusters /ŋg/, /ŋk/, and /ŋŋ/. In these clusters, /ŋ/ is often partially weakened rather than fully articulated, especially in /ŋg/ and /ŋk/. However, this reduction is not eliminative; articulatory features remain implicitly active, contributing to the phonological structure. Formally:→ /ŋ/ → Ø / __ C (within velar homorganic clusters)

Table 15. Nasal Reduction in Homorganic Clusters of the Prefix *um-* with Velar-Initial Bases

Prefix + Base	Orthographic Rule	Phonological Rule	Meaning
um- + gurgur	unggurgur	[uŋgurgur]	*more boiling
um- + holong	ungkolong	[ukkoloŋ]	*more loving
um- + ngalut	ungngalut	[uŋŋalut]	more difficult

/ŋg/ and /ŋk/, the nasal contributes indirectly through strengthening of the following consonant. In /ŋŋ/, the nasal is retained as a geminate, showing that reduction is cluster- and context-dependent.Nasal reduction mediates between assimilation and compensatory gemination, balancing articulatory efficiency and phonological cohesion.

c. Compensatory Gemination of the Prefix *um-* with Velar-Initial Bases

Compensatory gemination transforms weakened nasal + consonant clusters into geminate consonants, transferring articulatory features of the nasal to the following consonant. In velar contexts, clusters /ŋg/, /ŋk/, and /ŋŋ/ yield predictable patterns:

Table 16. Compensatory Gemination of the Prefix *um-*

with Velar-Initial Bases

Nasal + Consonant Cluster	Phonological Rule	Transformation
/ŋ/ + /g/	[gg]	unggurgur → [uŋgurgur]
/ŋ/ + /k/	[kk]	ungkolong → [ukkoloŋ]
/ŋ/ + /ŋ/	[ŋŋ]	ungngalut → [uŋŋalut]

Gemination occurs systematically whenever /ŋ/ is weakened pre-consonantally. Features [+nasal, +closure] are preserved and reassigned to the following consonant. Stable geminates (e.g., /ŋŋ/) demonstrate that not all homorganic clusters undergo reduction. From a theoretical perspective, OR generates intermediate forms with homorganic clusters. PR implements feature redistribution, preserving segmental integrity. Apparent deletion is in fact systematic feature reallocation, maintaining morphological and phonological cohesion.

Within the OR–PR framework, these stages reveal a predictable, systematic, and adaptive interaction between morphology and phonology. The prefix *um-* is not deleted but undergoes representational transformation across derivational stages, yielding allomorphic realizations *um-*, *uŋ-*, and their geminated forms. These findings demonstrate that Toba Batak active verb formation is shaped not only by morphological structure but also by dynamic, structured, and highly predictable phonological processes. Gemination functions as a structural mechanism that ensures both articulatory efficiency and segmental preservation, highlighting the intricate integration of morphology and phonology in the language.

DISCUSSION

The findings of this study demonstrate that the allomorphic variation of the prefix *um-* in Toba Batak is not arbitrary but governed by a systematic interaction between morphological structure and phonological processes. Within the extended generative morphology framework incorporating OR and PR, the alternations *um-*, *umm-*, *un-*, and *uŋ-* emerge as rule-governed outputs conditioned by phonological environments.

1. Theoretical Significance of the Patterns

The observed patterns matter theoretically because they provide empirical evidence that morphophonemic variation is best explained through feature-based interaction rather than surface-level rule listing. In particular, the data demonstrate that assimilation, gemination, and feature redistribution operate as a single integrated mechanism. This supports a shift from segment-based explanations toward feature geometry and constraint-based interpretations in generative morphology. Thus, the study extends Halle's model by showing that OR and PR function as essential mediators in Austronesian morphophonemics.

2. Relation to Previous Linguistic Theories

The results refine and extend classical generative morphology (Halle 1973) by demonstrating that rule ordering alone is insufficient to explain Austronesian morphophonemic behavior without explicit phonological optimization layers. While earlier models emphasize linear derivation, the present findings show that feature redistribution and phonological strengthening must be incorporated to account for surface variation. This aligns partially with post-lexical phonology and autosegmental approaches, but goes further by integrating orthographic representation as an active component in derivation (OR), which is not explicitly addressed in earlier Indo-European-based models.

3. Implications for Toba Batak Morphology and Phonology

The study reveals that Toba Batak morphology and phonology are tightly integrated systems in which morphological identity is preserved despite extensive phonetic restructuring. The prefix *um-* consistently maintains its grammatical function as an active verb marker across all environments. Morphophonemic variation is therefore not morphological alternation but phonological adaptation within a stable morphological framework. This indicates that Toba Batak exhibits a highly organized system where phonological processes are systematically constrained by morphological structure.

4. Comparison with Other Austronesian Languages

Compared to other Austronesian languages, such as Tagalog and Malay, where nasal substitution and prefix variation often show partial lexical conditioning, Toba Batak displays a more deterministic and fully environment-driven system. In Tagalog, for instance, nasal assimilation may be lexically variable, whereas in Toba Batak allomorph selection is fully predictable based on place of articulation alone. This suggests that Toba Batak represents a more strictly rule-governed morphophonemic system within the Austronesian family, with fewer lexical exceptions and stronger phonological conditioning.

5. Broader Theoretical Implications

The broader implication of these findings is that morphophonemic processes should be interpreted as feature reorganization systems rather than segment deletion or replacement mechanisms. The consistent evidence of feature preservation (e.g., [+nasal], [+closure]) through compensatory gemination supports a non-eliminative model of phonology in which surface reduction reflects structural redistribution. Furthermore, the integration of OR and PR demonstrates that morphological and phonological modules are not strictly separable but function as interacting layers in real-time derivation. This has implications for modeling Austronesian morphophonology and potentially for cross-linguistic theories of affix behavior.

Overall, the prefix *um-* in Toba Batak illustrates how a single morphological unit can generate multiple phonetic outputs through a systematic, feature-driven, and environment-sensitive morphophonemic mechanism. The findings contribute not only to the description of Toba Batak but also to broader theoretical debates in generative morphology regarding the interaction between structure, feature geometry, and phonological optimization.

Theoretical Contribution

This study contributes to generative morphology by demonstrating that morphophonemic variation is best accounted for through an integrated interaction between morphological structure and feature-based phonological optimization, rather than through independent or sequential rule application alone. By extending Halle's framework with Orthographic Rules (OR) and Phonological Rules (PR), the analysis shows that orthographic representation is not merely a surface transcriptional device but an active intermediary in derivation.

More importantly, the findings advance theoretical understanding of allomorphy by showing that assimilation, nasal adjustment, and compensatory gemination function as manifestations of a single unified feature-reorganization system. This challenges segment-deletion and rule-listing approaches, and instead supports a non-eliminative model in which phonological processes preserve underlying features (e.g., [+nasal], [+closure]) through redistribution rather than loss.

In addition, the study provides evidence that Austronesian morphophonemic systems, as exemplified by Toba Batak, may exhibit stricter phonological conditioning than previously described in related languages, thereby refining cross-linguistic assumptions about prefix behavior within the Austronesian family. Overall, the proposed model contributes a more integrated and constraint-

sensitive perspective to the theory of morphophonology, where morphological identity and phonological optimization operate as co-dependent components of a single generative system.

CONCLUSION

This study examined the morphophonemic allomorphy of the prefix *um-* in Toba Batak within an extended generative morphology framework. The analysis demonstrates that the prefix exhibits four systematic surface realizations—*um-*, *umm-*, *un-*, and *uŋ-*—which are not random variants but contextually conditioned outcomes of phonologically governed derivation. Across vowel-initial and consonant-initial environments (bilabial, alveolar, and velar), the alternations are shown to follow a predictable and rule-governed pattern. These variations arise through a sequential derivational process in which OR establish intermediate morphological structures, while PR apply context-sensitive operations such as place assimilation, nasal adjustment, and compensatory gemination. This layered interaction ensures that morphological identity is maintained throughout derivation while allowing phonetic form to adapt to articulatory constraints.

A key conclusion of this study is that the observed morphophonemic processes do not involve segment loss but rather systematic feature redistribution within the phonological structure. Apparent reductions or alterations of nasal segments are better explained as reorganization of phonological features that preserve grammatical function and enhance articulatory efficiency. In this sense, allomorphic variation in Toba Batak reflects a balance between morphological stability and phonological optimization.

The findings further confirm that the prefix *um-* plays a consistent role in active verb formation, maintaining its grammatical function across all phonological environments despite surface variation. This demonstrates a high degree of morphological productivity governed by strict phonological conditioning rather than lexical irregularity.

From a theoretical perspective, the study supports and extends generative morphology by showing that Halle's model, when enriched with OR and PR, can effectively account for morphophonemic complexity in Austronesian languages. The results also reinforce the view that morphology and phonology operate as an integrated system, where surface forms emerge from structured interactions rather than independent rule application.

Overall, this research contributes to a deeper understanding of Austronesian morphophonemics by providing empirical evidence that Toba Batak employs a highly systematic, feature-based, and context-sensitive mechanism of allomorphy. It also demonstrates the explanatory adequacy of an extended generative morphology framework for analyzing underrepresented languages, thereby offering a replicable model for future linguistic studies in similar typological contexts.

REFERENCES

Ambarita, E., & Sibarani, R. (2017a). Daftar morfem dalam sistem morfologi generatif bahasa Batak Toba. *TIFA, 9*(1), 191–207. https://www.researchgate.net/publication/332875475

Ambarita, E. (2017b). Denominal verbs in Toba Batak. *The Episteme: Journal of Linguistics and Literature, 3*, 13–24. https://www.researchgate.net/publication/332875637

Ambarita, E., Sibarani, R., Widayati, D., & Setia, E. (2018a). Nominal word formations in Toba Batak language: A study of generative morphology. *KnE Social Sciences*, 306–320. https://doi.org/10.18502/kss.v3i4.1942

Ambarita, E. (2018b). Deverbal nominals in Toba Batak language: A generative transformational study. *International Journal of Research and Review*, *5*(9), 184–192. https://www.ijrrjournal.com

Ambarita, E. (2018c). Adverbial derivation in Toba Batak language: A generative transformational study. *Language Literacy: Journal of Linguistics, Literature, and Language Teaching*, *2*(2), 103–110. https://doi.org/10.30743/ll.v2i2.689

Ambarita, E. (2018d). Deverbal nominals in Toba Batak language: A generative transformational study. *International Journal of Research and Review*, *5*(9), 183–192. https://www.ijrrjournal.com

Ambarita, E. (2018e). *Sistem morfologi bahasa Batak Toba: Kajian transformasi generatif* (Doctoral dissertation). Universitas Sumatera Utara. https://repositori.usu.ac.id/handle/123456789/12184

Ambarita, E. (2020). *Adjektiva komparatif menggunakan sufiks –an dalam bahasa Batak Toba*. Mega Press Nusantara. https://books.google.co.id

Ambarita, E. (2021a). *Adjektiva komparatif menggunakan infiks -um- dalam bahasa Batak Toba*. Mega Press Nusantara.

Ambarita, E. (2021b). Superlative forms of adjectives by article na + prefix *um-* in Toba Batak. *The Explora*, *7*(5), 14–22. https://ejournal.uhn.ac.id

Ambarita, E. (2022a). Comparative degree of adjectives: Patterns of the suffix *-an* in Toba Batak. Mega Press Nusantara.

Ambarita, E. (2022b). Adverbial affixations in Toba Batak: A study of generative morphology. *Quest Journals Journal of Research in Humanities and Social Science*, *10*(12), 114–127. https://www.questjournals.org

Ambarita, E. (2023a). Denominal nominals in Toba Batak: A generative transformational study. *The Explora*, *9*(2), 101–114. https://ejournal.uhn.ac.id

Ambarita, E. (2023b). Comparative degree of adjectives: Patterns of the prefix *um-* in Toba Batak. AA Rizky.

Ambarita, E. (2024). Comparative degree of adjectives: Patterns of the infix *-um-* in Toba Batak. Mega Press Nusantara.

Ambarita, E., Br. Sembiring, E. M., & Simanjuntak, F. M. P. (2025a). Verb formations with prefix *pa-* in Toba Batak: A study of generative morphology. *Atlantis Press*, 179–202. https://doi.org/10.2991/978-2-38476-501-0_18

Ambarita, E., Br. Sembiring, E. M., & Simanjuntak, F. M. P. (2025b). Prefiks infleksional pembentuk verba bahasa Batak Toba. Mega Press Nusantara.

Ambarita, E., Br. Sembiring, E. M., & Silalahi, V. (2025c). Afiks infleksional pembentuk verba bahasa Batak Toba. Dira Media Kreasindo. https://books.google.co.id

Ambarita, E., Panggabean, H., & Br. Sembiring, E. M. (2025d). Adjektiva komparatif dengan prefiks *um-* dalam bahasa Batak Toba. Mega Press Nusantara.

Ambarita, E. (2025e). The patterns of ablaut reduplication in Toba Batak: A construction-based approach. *REILA: Journal of Research and Innovation in Language*. https://doi.org/10.31849/reila.v7i1.17024

Ambarita, E., Silalahi, C. D. M., Br. Sembiring, E. M., & Silalahi, V. (2026). Morphophonemic allomorphs of the prefix *sa-* in Toba Batak verbs: A generative and pedagogical perspective. *Journal of Classroom Action Research, 5*(1), 39–59. https://journal.eltaorganization.org

Anderson, S. R. (1982). Where's morphology? *Linguistic Inquiry, 13*(4), 571–612. https://babel.ucsc.edu

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

Blust, R. (2023). Repair and drift in Austronesian languages. *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 & Row.

Damanik, S., & Mulyadi, M. (2020). Ergativity case marking in Toba Batak language. *BIRLE Journal, 3*(1). https://doi.org/10.33258/birle.v3i1.755

Diver, W. (1969). The system of relevance of the Homeric verb. *Acta Linguistica Hafniensia*, *12*(1), 45–68.

Driem, G. van. (2011). Proto-Kiranti verbal morphology. *Acta Linguistica Hafniensia*, *22*(1), 27–48.

Fendel, V. B. (2025). Celebrating diversity: The origins and pathways of three support-verb constructions. *Lexis*. https://journals.openedition.org/lexis/8217

Gerritsen, N. (2011). Russian reflexive affix *-sja*. *Acta Linguistica Hafniensia*, *25*(1), 29–37.

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

Haspelmath, M., & Sims, A. D. (2002). *Understanding morphology*. Oxford University Press.

Katamba, F. (1993). *Morphology*. Macmillan.

King, J. T. (2002). Marked transitive scenarios and archaic biactinal agreement morphology in Dhimal. *Acta Linguistica Hafniensia*, *34*(1), 7–28. https://doi.org/10.1080/03740463.2002.10414608

Lubis, D. F., & Bowo, T. A. (2022). Toba Batak script preservation. *NST Proceedings*, 28–34. http://nstproceeding.com

Luthfiani, E., Wagiran, W., & Rustono, R. (2020). Halle deadjectival verbs. *Seloka*, *9*(1), 45–52. https://doi.org/10.15294/seloka.v9i1.36496

Nida, E. (1949). *Morphology*. University of Michigan Press.

Panggabean, H., Ambarita, E., & Silalahi, C. D. M. (2025). Kata sifat komparatif dengan akhiran –*an* dalam bahasa Batak Toba. UMSU Press.

Purba, D. F., et al. (2022). Contrastive analysis English–Batak Toba. *Journal of Humanities and Social Sciences and Business*, 1(4), 121–128. https://doi.org/10.55047/jhssb.v1i4.288

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

Saragih, D. (2021). An analysis of metaphor in Toba Batak 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

Schane, S. A. (1973). *Generative phonology*. Prentice-Hall.

Sigurd, B. (1966). Generative grammar and historical linguistics. *Acta Linguistica Hafniensia*, *10*(1), 1–12. https://doi.org/10.1080/03740463.1967.10411454

Sinaga, A. B. (2002). *Tata bahasa Batak Toba*. Bina Media Perintis.

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

Stump, G. T. (2001). *Inflectional morphology*. Cambridge University Press.

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