

SEMANTIC ROLES OF *BATAK TOBA* EMOTION VERBS: A ROLE AND REFERENCE GRAMMAR PERSPECTIVE

Febrika Dwi Lestari¹, Mulyadi², Mahriyuni³, Erikson Saragih⁴

Universitas Sumatera Utara, Medan, Indonesia

Corresponding author: febrikadwilestari@uhs.ac.id

Abstract

This study investigates the semantic roles of emotion verbs in *Batak Toba*, an Austronesian language spoken in North Sumatra, Indonesia, from a Role and Reference Grammar (RRG) perspective. Drawing on a lexical dataset of 100 emotion verbs categorized across six primary emotion domains — *happiness*, *sadness*, *fear*, *shame*, *anger*, and *love/admiration* — the study analyzes how thematic relations such as Actor, Undergoer, Experiencer, Stimulus, and Theme are encoded in emotion verb constructions. The data reveal that *Batak Toba* emotion verbs exhibit a rich system of semantic role assignment shaped by cultural values, bodily experience, and the concept of *roha* (the seat of emotion). The analysis employs RRG's Logical Structure (LS) representation to account for the predicate-argument structure of emotion predicates. Results show that most *Batak Toba* emotion verbs follow an Experiencer-Stimulus configuration, with a significant number also encoding Actor-initiated affective states. Additionally, morphological derivational patterns (such as *pa-*, *mar-*, and *tar-* prefixes) systematically modify the semantic role configurations of base emotion verbs. This study contributes to the documentation of *Batak Toba* as an endangered language and provides cross-linguistic insight into the semantics of emotion across Austronesian languages.

Keywords: *Batak Toba* language; emotion verbs; semantic roles; Role and Reference Grammar; Austronesian linguistics; thematic relations

1. Introduction

The Language reflects the inner emotional life of its speakers, and the ways different languages lexicalize emotion reveal profound cultural and cognitive underpinnings. *Batak Toba* is an Austronesian language spoken by the *Batak Toba* ethnic group, primarily inhabiting the Lake Toba region in North Sumatra, Indonesia. As a language with a rich oral literary tradition, a distinctive script (Surat Batak), and deep-rooted cultural norms (*adat*), *Batak Toba* offers a unique lens through which to examine the cross-linguistic encoding of emotional experience.

Emotion verbs constitute one of the most semantically complex domains in any language. Unlike prototypical action verbs, emotion verbs do not denote physical events in the external world but instead express internal psychological states. From a semantic standpoint, the predicates in this domain often involve non-canonical argument structures: participants may simultaneously be interpreted as an Experiencer (the one who undergoes an emotional state) and an entity affected by an external Stimulus. This duality presents

challenges to standard semantic role analysis and makes emotion verbs an especially productive site for theoretical investigation.

Role and Reference Grammar (RRG), developed by Van Valin, R. D., Jr., & LaPolla (1997) and subsequently refined by Van Valin (2005), provides a robust framework for analyzing the syntax-semantics interface through the concept of Logical Structure (LS) representations and a hierarchically organized set of thematic relations. Unlike case-based theories, RRG employs a macro-role system that subsumes multiple micro-roles under two overarching labels — Actor and Undergoer — making it particularly well-suited for cross-linguistic analysis of predicates with complex argument structures, such as emotion verbs.

Batak Toba belongs to the Batak subgroup of the Malayo-Polynesian branch of the Austronesian language family (Adelaar 2005). It is spoken by approximately two to three million speakers, with significant diaspora communities in Jakarta, Bandung, and other Indonesian urban centers. The language is characterized by a Verb-Subject-Object (VSO) basic word order in formal registers, though Subject-Verb-Object constructions also occur in colloquial speech. Nominal morphology is extensive, with a complex system of affixes modifying argument structure and semantic role assignment.

Despite the theoretical and typological interest of *Batak Toba*, the language remains significantly under-documented in the linguistic literature. Existing studies have focused primarily on morphophonology and nominal morphology, while the semantic structure of the verbal lexicon — particularly the emotion verb domain — has received little systematic attention. This study addresses this gap by analyzing a comprehensive corpus of 100 *Batak Toba* emotion verbs across six emotional domains: happiness (*sonang*, *las*, *martua*), sadness (*lungun*, *hansit*, *susa*), fear (*mabiar*, *holsoan*, *lomos*), shame (*maila*, *manolsoli*, *marsala*), anger (*muruk*, *sogo*, *marsak*), and love/admiration (*holong*, *masihol*, *longang*).

One of the most culturally and grammatically significant concepts in *Batak Toba* is *roha*, a term that roughly translates as 'heart, mind, or locus of feeling.' The *roha* functions in *Batak Toba* as the seat of all emotional, cognitive, and volitional processes is similar in some respects to the concept of *nawa* in Javanese or the Malay *hati*. Crucially for the purposes of this study, *roha* appears frequently as a nominal argument in emotion verb constructions, and its presence or absence systematically affects the semantic role interpretation of the construction. Sentences such as *Las do roha* (lit. 'The heart is happy') employ *roha* as the grammatical subject but function semantically as Experiencer-state predications about the entire person.

Previous scholarship on *Batak Toba* linguistics includes Nababan (1981) foundational grammar, Kozok (2009) study of the script tradition, and Tuuk (1971) classical dictionary. More recent work by Siahaan (2008) has examined metaphorical constructions in *Batak Toba*, noting the centrality of the heart-metaphor in emotional expression. However, no study to date has systematically applied RRG's analytical apparatus to the full range of *Batak Toba* emotion verbs, and the present study fills this gap.

The primary objectives of this study are to identify and classify the semantic roles encoded in *Batak Toba* emotion verb constructions; and to analyze the predicate-argument structures of these verbs using RRG's Logical Structure representation.

The findings of this study have implications for several areas of inquiry: the documentation and revitalization of *Batak Toba* as an endangered language; cross-linguistic typology of emotion predicates; and the capacity of RRG's theoretical apparatus to account for the semantic complexity of emotion in non-European languages.

2. Literature Review

Role and Reference Grammar (RRG) is a monostratal, non-transformational syntactic-semantic theory developed by Robert D. Van Valin Jr. and William A. Foley (1980) and substantially elaborated in Van Valin and LaPolla (1997) and Van Valin (2005). Unlike Government-Binding Theory or Minimalism, RRG is typologically motivated and seeks to account for the syntactic and semantic properties of natural languages without positing abstract underlying representations or covert movement.

Central to RRG is the decomposition of predicates into Logical Structures (LS) based on Vendlerian aktionsart distinctions: states (be'), activities (do'), achievements (INGR), and accomplishments (BECOME). Each predicate type is associated with a distinct LS template, and semantic roles (thematic relations) are derived from the argument positions within these structures. Van Valin (2005) identifies a set of thematic micro-roles including: Agent (volitional instigator), Effector (non-volitional causer), Experiencer (sentient entity undergoing a psychological state), Perceiver, Stimulus (entity that triggers perception/emotion), Recipient, Theme (entity undergoing state change), Patient (affected participant), and Locative.

Crucially, RRG organizes these micro-roles into a macro-role hierarchy: Actor (the most agent-like argument) and Undergoer (the most patient-like argument). This Actor-Undergoer continuum is determined by the hierarchical ranking: Agent > Effector > Experiencer > Locative > Theme > Patient. In the domain of emotion verbs, the placement of Experiencer and Stimulus within this hierarchy is a central theoretical issue, as these roles do not map uniformly onto Actor and Undergoer across languages.

The semantic analysis of emotion verbs has attracted considerable attention in linguistics since Belletti, A. and Rizzi (1988) influential study of psych-verbs in Italian, which identified three structural classes: Class I (Stimulus-Subject verbs, e.g., worry), Class II (Experiencer-Subject verbs, e.g., fear), and Class III (Experiencer-Object verbs involving psych predicates). Dowty (1991) further contributed to the analysis through his proto-role theory, arguing that Agent and Patient are clusters of semantic entailments rather than discrete categories.

In a cross-linguistic perspective, Croft (1993) demonstrated that emotion verbs exhibit cross-linguistically diverse argument structure patterns, with the grammatical encoding of Experiencer and Stimulus showing considerable typological variation. Languages may frame the Experiencer as subject (active-experiencer framing) or as object (stative-experiencer framing), and these choices correlate with cultural and cognitive attitudes toward emotional agency.

For Austronesian languages specifically, studies by Kroeger (2004) on Philippine-type voice morphology and Klamer (2008) on Eastern Indonesian have noted that emotion predicates interact in interesting ways with voice alternation systems, focus marking, and applicative constructions. The encoding of the Experiencer as subject versus object has been shown to correlate with the degree of emotional control attributed to the Experiencer in these cultures.

3. Research Method

This study adopts a qualitative-descriptive research design rooted in the tradition of linguistic fieldwork and lexical-semantic analysis. The research integrates documentary linguistics with formal semantic theory, combining corpus-based data collection with RRG-

based analysis of predicate-argument structure. This methodological approach follows the framework recommended by Himmelmann (1998) for endangered language documentation, adapted to the specific analytical goals of semantic role investigation.

The primary data for this study consists of a purpose-built lexical database of *Batak Toba* emotion verbs, compiled through systematic elicitation sessions with native speakers and cross-referenced against existing lexicographic resources, including Tuuk's (1971) Batak-Dutch dictionary and Kozok's (2009) Batak lexicon. The corpus comprises 100 emotion verbs organized into six semantic domains: (1) happiness (16 verbs); (2) sadness (20 verbs); (3) fear (20 verbs); (4) shame (7 verbs); (5) anger (28 verbs); and (6) love/admiration (9 verbs). A seventh micro-domain of surprise/wonder (6 verbs) was also identified and included in the analysis.

For each verb, the database records: the *Batak Toba* lexical form; morphological class (with derivational affixes noted); the approximate Indonesian gloss; a characterization of the emotion's semantic nuance (e.g., whether the emotion involves physical bodily expression, whether it is triggered by an external stimulus, whether it is associated with the *roha*); and one or more clause-level example sentences drawn from native speaker elicitation.

4. Result and Discussion

4.1 Overview of the Batak Toba Emotion Verb Lexicon

The analysis of the full corpus reveals that *Batak Toba* possesses a semantically dense and typologically interesting emotion verb lexicon. A total of 100 distinct lexical forms were identified, with a high degree of semantic differentiation within each emotional domain. This lexical richness mirrors what Wierzbicka (1999) has called the 'emotional vocabulary' of a culture, whereby fine-grained emotional distinctions encode culturally significant differences in the way emotions are experienced, expressed, and evaluated.

Table 1 provides a summary of the corpus distribution across the six primary emotional domains:

Table 1. Distribution of Batak Toba Emotion Verbs Across Semantic Domains

Domain	Representative Verbs	Key Semantic Feature	No. of Verbs
Happiness	<i>sonang, las, martua, marolopolop, sombu, minar</i>	Positive affect; bodily/facial expression	16
Sadness	<i>lungun, marsak, hansit, susa, porsuk, bonsa</i>	Negative affect; <i>roha</i> -internal	20
Fear	<i>mabiar, holsoan, lomos, busisaon, gamang</i>	Anticipation of threat; psychosomatic	20
Shame	<i>maila, tanggak, haduk, manolsoli, marsala</i>	Social/moral evaluation; hierarchy	7
Anger	<i>muruk, sogo, marsak, hosom, pamaloson</i>	Negative arousal; <i>roha</i> -internal/external	28
Love/Admiration	<i>holong, masihol, longang, iburu, elat</i>	Positive orientation toward other	9
Surprise/Wonder	<i>longang, tarmali, tarsonggot, tarhatotong</i>	Sudden cognitive reorientation	6
TOTAL			100

4.2 Semantic Role Configurations in RRG Analysis

Following Van Valin's (2005) RRG framework, *Batak Toba* emotion verbs can be classified into four principal thematic configurations based on their Logical Structure representations:

- a. Configuration I — Experiencer-State: The predicate denotes an internal state of the Experiencer with no encoded external Stimulus. The Experiencer surfaces as the grammatical subject (or as the possessor of *roha* in the *roha* construction). The macro-role of the Experiencer is analyzed as Undergoer, given the stative nature of the predicate. Examples include *sonang* 'happy,' *lungun* 'sad/longing,' and *mabiar* 'afraid.' The Logical Structure for this configuration is: be' (x, [Pred']) where x = Experiencer/Undergoer.
- b. Configuration II — Experiencer-Stimulus: The predicate encodes both an Experiencer and an external Stimulus that triggers the emotional state. The Experiencer is the Undergoer; the Stimulus is an oblique or secondary argument. This configuration accounts for a large class of *Batak Toba* emotion verbs including *holsoan* 'worried' (about something), *masihol* 'longing (for someone),' and *holong* 'loving (someone).' LS: be' (x, [Pred' (y)]) where x = Experiencer/Undergoer and y = Stimulus.
- c. Configuration III — Causative-Experiencer: These are transitive emotion verbs formed through the causative prefix pa-, in which an external cause (Actor) brings about an emotional state in the Experiencer/Undergoer. Examples include *pasonang* 'to make happy,' *palashon* 'to gladden,' and *pasusahon* 'to cause trouble/sadness.' LS: [do' (x, Ø)] CAUSE [BECOME be' (y, [Pred'])] where x = Actor/Causer, y = Experiencer/Undergoer.
- d. Configuration IV — Bodily-Enactive: These verbs encode emotional states that are constitutively expressed through bodily movement or physical gesture. The Experiencer performs a bodily enactment of the emotion, warranting Actor-like properties. Examples include *marolopolop* and *mariaia* (joyful bodily movement during festivities), *mirjakirjak* (joyful physical restlessness), and *girgir* (excited physical movement). LS: do' (x, [Pred']) where x = Experiencer/Actor.

4.3 Emotion Domain-Specific Semantic Role Patterns

Analysis of the six emotional domains reveals both universal tendencies and language-specific configurations in semantic role assignment.

In the happiness domain, *Batak Toba* exhibits a particularly high degree of semantic differentiation, with 16 distinct verbs encoding different nuances of positive affect. Semantically, these verbs can be stratified by: the presence or absence of bodily expression (e.g., *marolopolop* and *mariaia* involve physical movement, whereas *sombu* is purely internal); the type of trigger (*sombu* is specifically triggered by reunion or the completion of a burden, while *lambok* and *lambas* encode relief-joy); and the domain of the *roha* (some happiness verbs exclusively occur with *roha* as argument, e.g., *marlas ni roha*, while others take the whole-person Experiencer).

The anger domain, with 28 verb entries, is the most lexically dense in the *Batak Toba* emotion lexicon. This richness may be culturally significant: in Batak society, where social hierarchy (*daliha na tolu*) and public honor are central values, fine-grained distinctions in the type, intensity, and social context of anger are communicatively important. The data reveal a semantic continuum from mild, internalized displeasure (*jut* 'slight annoyance; *roha-*

internal') through increasing intensity (*sogo* 'disgust/dislike; *roha*; directed at repeated behaviors,' *marsak* 'resentful disappointment,' *muruk* 'general anger,' *piripirian* 'intense anger with physical expression') to extreme loss of control (*mangamuk* 'to burst anger,' *marsigorgor* 'to rage internally').

Within RRG terms, the most notable finding in the anger domain is the systematic distinction between verbs encoding *roha*-internal anger states (Undergoer-prominent) and verbs encoding Actor-prominent anger expressions that entail behavioral consequences. *Muruk* follows a stative LS: be'(x,[*muruk*']), whereas *mangamuk* follows an activity LS: do'(x,[*mangamuk*']), placing the Experiencer in an Actor macro-role as an Agent-like, physically enacting angry subject.

The fear domain displays a prominent psychosomatic dimension: many fear-domain verbs explicitly encode the physical manifestation of the fear state. *Hatahutan* 'to feel terrified with physical reaction,' *gamang* 'to feel dizzy-with-fear,' *hitirhitir* 'to tremble with nervousness,' and *gumagang* 'to reach about unsteadily from fear' all incorporate somatic entailments into their semantic structure. In RRG terms, these verbs carry additional lexical material in their LS predicate node that specifies the manner of physical response, yielding complex LS representations such as: be'(x, [*hitirhitir*']) & do'(x, [*tremble*']).

In the love and admiration domain, semantic role assignment is complicated by the orientation of the emotion toward an external Other. Verbs such as *iburu* 'to be jealous' and *elat* 'to be envious' involve a three-participant structure: Experiencer, Stimulus (the envied/jealous-making entity), and in some readings, a Rival or Possessor whose advantage triggers the emotion. This configuration calls for a three-argument LS representation, where the Stimulus is encoded as an embedded predicate argument.

4.3 The *Roha* Construction and Its Semantic Role Implications

One of the most distinctive features of *Batak Toba* emotion expression is the frequent use of the *roha* construction, in which the abstract emotional locus *roha* appears as the grammatical subject of a stative predication. This construction is typologically significant because it creates a complex double-Experiencer structure: the *roha* is the surface grammatical subject, while the whole person (the true psychological Experiencer) is either implicit or encoded as a possessor nominal.

Consider the following examples from the corpus:

(1) *Las do roha marnida panortoron ni angka inanta soripada i.*

'The heart is joyful seeing the beautiful display of our mothers.'

(2) *Lungun do roha molo marsirang na maraleale.*

'The heart is sad when those who are close part from one another.'

(3) *Sogo do roha ni lae i ala sai olo donganna i manangko.*

'The brother-in-law's heart is angry because his friend keeps stealing.'

In these constructions, the *roha* functions grammatically as the Undergoer of the stative emotion predicate, but culturally and pragmatically it is understood as the emotional locus of a specific person whose identity is provided by the possessor (*ni lae i* 'of the brother-in-law' in example 3) or by the discourse context. Within RRG's macro-role system, the *roha* is assigned the Undergoer macro-role in its Logical Structure position, yielding LS representations such as: be' (*roha*, [*Las*']) = 'the heart-entity is in a state of joy.'

Crucially, the choice between the *roha* construction and a construction with the whole-person Experiencer as subject is not semantically vacuous. Corpus analysis reveals

that the *roha* construction is preferentially used when the emotion is experienced as involuntary, deeply felt, or culturally recognized as pertaining to the core inner self. By contrast, constructions without *roha*, where the Experiencer-person appears directly as the grammatical subject (e.g., *Muruk do guru* 'The teacher is angry'), tend to encode emotions that are more overt, publicly expressed, or volitionally enacted. This distinction parallels the difference between the 'passive' and 'active' Experiencer types identified by Kailuweit (2005) in Romance languages.

5. Conclusion

This study has investigated the semantic roles of *Batak Toba* emotion verbs from a Role and Reference Grammar (RRG) perspective, drawing on a corpus of 100 lexical items distributed across six primary emotional domains. Several significant findings emerge from the analysis.

First, *Batak Toba* emotion verbs exhibit four principal semantic role configurations within RRG's Logical Structure framework: (I) Experiencer-State, in which the Experiencer is the Undergoer of a stative predicate with no encoded Stimulus; (II) Experiencer-Stimulus, in which a Stimulus argument is present as a secondary or oblique participant; (III) Causative-Experiencer, derived through the *pa-* prefix and introducing an Actor-Causer argument; and (IV) Bodily-Enactive, in which the Experiencer assumes Actor-like macro-role properties due to the constitutive physical expression of the emotion.

Second, morphological derivation — particularly the prefixes *pa-*, *mar-*, and *tar-* — systematically modifies the semantic role configurations of base emotion predicates, creating transitive causatives, inherent-state predicates, and inchoative involuntary-onset emotion predicates respectively. These derivational patterns demonstrate that the syntax-semantics interface in *Batak Toba* emotion verbs is closely regulated by the morphological system.

Third, the concept of *roha* constitutes a grammatically and semantically productive element in *Batak Toba* emotion expression. The *roha* construction, in which *roha* serves as the grammatical Undergoer of an emotion predicate, encodes emotions experienced as involuntary, deeply felt, and culturally significant, thereby distinguishing a culturally recognized 'inner emotional self' from the social, behaviorally-oriented Experiencer. This two-tiered Experiencer system has implications for the cross-linguistic typology of emotional expression and deserves further investigation within the broader Austronesian language family.

The findings of this study contribute to several areas of linguistic inquiry. From a theoretical standpoint, they provide evidence for the cross-linguistic applicability of RRG's macro-role system to the semantically complex domain of emotion verbs in a non-European Austronesian language. From a typological standpoint, they add to the growing body of literature on Experiencer constructions and psych-predicates across languages. From a documentary standpoint, they represent a contribution to the documentation of *Batak Toba* as an endangered language, providing a systematic lexical-semantic record of a domain whose cultural richness has hitherto been under described.

Future research directions include: (1) a syntactic investigation of the RRG-theoretic operator projections in *Batak Toba* emotion verb clauses; (2) a comparative study of emotion verb semantics across the Batak language cluster (including Karo Batak, Pakpak Dairi, and Angkola Mandailing); (3) a corpus-based analysis of emotion verb usage in *Batak*

Toba oral literature (*umpasa* and *andung*); and (4) an experimental psycholinguistic investigation of the cognitive representation of *roha* in *Batak Toba* speakers' emotional self-understanding.

References

- Adelaar, A. (2005). The Austronesian languages of Asia and Madagascar: A historical perspective. In A. Adelaar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 1–42). Routledge.
- Belletti, A., & Rizzi, L. (1988). Psych-verbs and θ -theory. *Natural Language and Linguistic Theory*, 6(3), 291–352. <https://doi.org/10.1007/BF00133904>
- Croft, W. (1993). Case marking and the semantics of mental verbs. In J. Pustejovsky (Ed.), *Semantics and the lexicon* (pp. 55–72). Kluwer Academic Publishers.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3), 547–619. <https://doi.org/10.1353/lan.1991.0021>
- Himmelmann, N. P. (1998). Documentary and descriptive linguistics. *Linguistics*, 36(1), 161–195. <https://doi.org/10.1515/ling.1998.36.1.161>
- Kailuweit, R. (2005). *Linking: Syntax und Semantik französischer und italienischer Gefühlsverben* [Linking: Syntax and semantics of French and Italian emotion verbs]. Niemeyer.
- Klamer, M. (2008). *Alorese: A language of Eastern Indonesia*. LINCOM Europa.
- Kozok, U. (2009). A 14th century Batak manuscript. *Bijdragen tot de Taal-, Land- en Volkenkunde*, 165(2), 278–305. <https://doi.org/10.1163/22134379-90003639>
- Kroeger, P. (2004). *Analyzing syntax: A lexical-functional approach*. Cambridge University Press.
- Nababan, P. W. J. (1981). *A grammar of Toba Batak* (Pacific Linguistics Series B, No. 82). Research School of Pacific Studies, Australian National University.
- Siahaan, P. (2008). Head and heart in Batak. In F. Sharifian, R. Dirven, N. Yu, & S. Niemeier (Eds.), *Culture, body, and language: Conceptualizations of internal body organs across cultures and languages* (pp. 37–67). Mouton de Gruyter.
- Tuuk, H. N. van der. (1971). *Tobasche spraakkunst* [Toba Batak grammar] (Reprint of 1864–1867 ed.). Koninklijk Instituut voor Taal-, Land- en Volkenkunde.
- Van Valin, R. D., Jr., & LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function*. Cambridge University Press.
- Van Valin, R. D., Jr. (2005). *Exploring the syntax-semantics interface*. Cambridge University Press.
- Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press.