Vol. 11, Issue I 2026

ISSN: 2426-2149

SEL

Society for Endangered and lesser known Languages

वाक् मंथन

सोसाइटी फॉर इंडेन्जर्ड एंड लेसर नोन लैंग्वेज़ेज की शोध पत्रिका

वाक् मंथन
ISSN: 2426-2149
https://selindia.org/

Editorial

We feel privileged to grab the opportunity of this special issue of Vaak Manthan to publish the manuscripts obtained from the two-day National Seminar on '**India as a Linguistic Area: Exploring Shared Features Across Language Families**' (ILAESFALF -2025), organised by the **Centre for Linguistics, SLL&CS, Jawaharlal Nehru University**.

The Biannual, bilingual peer-reviewed E-Journal is an offshoot of the Society for Endangered and Lesser-Known Languages. Thus, we bring to this issue a selected list of manuscripts dealing with convergence, language contact, and more significantly the typological features of lesser-known languages.

We see that in a multilingual country where a region or a district is a habitat of diverse communities, language convergence or contact-induced features are adapted by the speakers of varied languages resulting in structural similarities. Rastogi provides us with a broad linguistic landscape of Uttarakhand, and throws light on the broader patterns and implications of language contact in marginalized and multilingual settings of two Tibeto-Burman languages - Jad and Raji; and aims in contributing to how the nature and extent of contact-induced convergence has significantly affected the linguistic identities of its communities, and at the same time showing different sociolinguistic trajectories.

Padhan attempts a sociolinguistic study by exploring an evolving morphosyntactic structure of Sundargarh Sadri in the context of the usage of enclitic marker *-har* in inalienable possession in two different age groups - adult and children's speech, and its interaction with the neighbouring languages. The study contributes to the language variation and contact owing to the influence of the regional languages – Odia, Sambalpuri, and Hindi. Both the Sadri varieties Sundargarh and Jharkhand show deviation from the original nature, simplifying and reducing the distinction of alienable-inalienable possessive system, to be at par with the neighbouring regional languages. However, the usage of *-har* varies in different regions. The study further investigates phonological and morphological variations which states deeper patterns of contact-induced change.

The language families show convergence in a couple of morpho-syntactic features yet showing specificity in certain functions. Jakhar exhibits phonological processes in pluralization of Bagri that shows variation with Standard Hindi and other Indo-Aryan languages. The author elaborates with plural examples of vowel lengthening, tonal variation, epenthesis, gemination, degemination, voicing of voiceless consonants, reduplication, among many others.

Reduplication is regarded as an essential grammatical criterion found across all language families -Indo Aryan, Tibeto Burman, Dravidian, Austro Asiatic, and Tai Kadai. Kopparapu and Sankar contributes to the evaluation of Reduplicative meanings which manifests intensification, distribution, and iteration, capturing cross linguistic patterns in Indo Aryan, Dravidian, Munda, and Tibeto Burman languages. This is done by positing a R-D semantic model comprising Reduplicator (R), Iterator (I), and Distributor (D) and grounding their effects in binary time/space features of morpho-syntactic categories – nominals, verbs, adpositions,

and modifiers. Lal and Ray demonstrate Reduplication in Bagri adopting Goldsmith's Auto segmental framework which reveals that Complete, Echo, and Numeral reduplication is governed by a systematic phonological phenomenon showing distinct manipulation of independent phonological tiers. Naha illustrates Reduplication in Kokborok with examples from Expressives, Echo formation, Complete, Partial, Discontinuous, Semantic Reduplication, and Reduplication as iconic form-meaning mapping.

The characteristic typological features of languages manifest the interesting grammatical patterns and the nature of human languages. This also brings to the understanding of the formation of the grammatical (or phonological) concepts, functions and markers, and their intermediacy to geographical and pragmatic factors. Thus, in Trans-Himalayan languages, topographical deixis is an essential grammatical function; Tibeto-Burman languages exhibit restricted temporal functions; Dravidian languages lack in lexical adjectives, the Austroasiatic languages show predominance of sesquisyllables; Tai Kadai languages are tonal in nature. In this volume, Nikhilesh demonstrates a typological morphosyntactic feature of Dravidian languages in general, and Telangana Telugu in particular. The paper elaborately focuses on how these languages lacking in lexical adjectives compensates with the formation of Property Concept expressions derived from nouns and verbs through copula and relative clause morphology, and operates in attributive modification positions. Ronald investigates the structure and phonological function of sesquisyllables in Khasi, making a departure from classical Sonority Sequencing Principle (SSP), and evaluating the rich onset clusters using the Beats and Binding (B&B) theory and Baroni's Net Auditory Distance (NAD) model. Doley discusses in detail the case marking system in Pagro Mising, representing the Eastern Tani branch of the Tibeto-Burman family. The study makes note of the fact that ablative marker is derived from the merging of locative and genitive cases in Tani languages. Majee's work entails interesting patterns of negative particles, negative clitic and negative copula in Kumaoni. The Present Tense form of verbs brings to the surface the usage of negative copula and negative clitics which reasonably hints to pragmatic significance of the discourse. Sen provides formal syntactic analysis to investigate how verbs in Kumaoni license arguments syntactically and interacts with the morpho-syntactic processes, using the theoretical models.

A language can be categorized on the basis of the property of word formation. Some languages like Mandarin or Tai Khamti are mono-morphemic where each morpheme is a content word and a free word form. Other languages differ in varying degrees in the affixation processes or bound forms eg., in Nyishi, *biŋ- co- ja- mu- tarine* (speak-first-wait-let-FUT) 'let you speak first'. While in certain languages morphological affixations are identified, in other languages affixations are opaque. In addition, languages coin new words to keep pace with the evolution of mankind, development of new ideas, objects, or social phenomena. Teronpi, Yimchunger, and Som explores Neologisms across the North-East Indian languages - Yimkhiung, Bodo, Mizo, Lotha, Karbi, Ao, Simtee and Khasi through the process of Compounding, hybridization, blending, loan words, calque, coinage, and metaphorical extension. The work examines the conceptualisation of neologism, the word structures and semantic functions. Deka and Nath contributes to a documentation of the word formation processes of Zyphe or Zophei, language which belongs to the Maraic branch of the Kuki-Chin group of the Tibeto-Burman Language Family, data collected from Siata and Iana, two remote villages of the Saiha district of Mizoram.

This issue also contributes to the problems of the introduction of the indigenous languages in the school curriculum. Talukdar and Nath explore the hurdles and challenges faced in translating and creating effective school materials for the Rabha and Tiwa languages under Assam's Mother Tongue-Based Multilingual Education (MTB-MLE) program, gathering primary data from the two major regions: Goalpara district (for Rabha) and Morigaon district (for Tiwa).

Singhania and Ray illustrates that the sociolinguistic profile of Magahi, originated from Magadhi Prakrit. Despite its appreciable demographic size and rich oral tradition which can trace back to the administrative language of Emperor Ashoka, there is an urgency for the policy interventions and community efforts to preserve and revitalize the language.

Finally, Ray subsumes the present volume with an exhaustive account of the notions of 'language', 'standard language', 'dialect' and the cultural hegemony of the ruling elites. He provides rational understanding of how the minorities are severely sabotaged by the ruling elites with their prescribed notions of 'standardization', and their cultural and religious ideologies.

The special issue contributes to data-based analysis attempted on diverse languages, addressing a wide variety of research questions, hurdles & challenges of translating education materials, theoretical models, linguistic peculiarities and impregnable fluidity of languages which converges into India as a linguistic area.

Dr. Bishakha Sarma

Independent Researcher
&
SEL Fellow

Language Contact and Convergence in Tibeto-Burman Languages of Uttarakhand: A Layered Perspective

Kavita Rastogi
krastogi1960@gmail.com

Abstract

Languages naturally interact and influence one another when they come into contact, often leading to structural and lexical convergence. Historically, every language has undergone a degree of transformation due to such interactions, with some changes being subtle and others resulting in the emergence of entirely new linguistic forms. The Tibeto-Burman (TB) languages of Uttarakhand, spoken primarily in the high-altitude border regions, exhibit varying degrees of linguistic convergence due to their prolonged contact with Indo-Aryan languages such as Kumauni, Garhwali, and Hindi.

This study examines the impact of sustained language contact on the structure of TB languages in Uttarakhand, conceptualizing their evolution as a *layered cake*—where different linguistic elements have been gained, lost, or restructured over time. The research aims to identify and analyse these accumulated *tiers* of naturalized foreign elements, focusing on lexical borrowing, phonetic adaptation, syntactic influence, and grammatical restructuring. Through a comparative analysis of specific TB languages—such as Jad and Raji—this study seeks to discern patterns of convergence and divergence, shedding light on the mechanisms that drive language change in multilingual settings.

By drawing on historical linguistic data, fieldwork observations, and sociolinguistic analysis, this research contributes to a broader understanding of language contact phenomena in the Himalayan region. The findings will offer insights into the adaptive strategies of minority languages in response to linguistic dominance and cultural integration, furthering discussions on language preservation and shift in contact zones.

Key words: Language Contact, Language Convergence, Raji, Jad, Uttarakhand.

1.Introduction

On November 9, 2000, the state of Uttaranchal, the 27th state of India, was carved out of Uttar Pradesh. In January 2007, the state was renamed Uttarakhand, meaning "northern region," which reflects the area's traditional designation. Following its formation, Hindi and Sanskrit were established as the official languages of the state. The state of Uttarakhand is divided into two socio-cultural regions:

- Kumaun, and
- Garhwal.

In the Kumaun region, several dialects are spoken under the cover term *Kumauni*. Similarly, *Garhwali*, with its own dialectal variations, is spoken in the Garhwal region. In addition to these major linguistic groups, the state is home to five Scheduled Tribes—

- Jaunsari,
- Bhotia,
- Buksha,
- Tharu, and
- Raji.

Smaller groups such as the Jad (Rongpa), Darma, Chaudansh, and Byansh also inhabit the region and are typically classified under the broader label *Bhotia*. Each of these communities speaks a distinct language. While Kumauni, Garhwali, and Hindi belong to the Indo-Aryan language family, the languages spoken by the Scheduled Tribes and smaller ethnic groups predominantly belong to the Tibeto-Burman language family.

Language contact in Uttarakhand has had a profound impact on the linguistic identities of its communities. Prolonged interaction with dominant languages has led to a noticeable shift among many communities toward these more widely spoken tongues. In UNESCO's *Atlas of the World's Languages in Danger* (2010)[1], three major regional languages—Rongpa, Garhwali, and Kumauni—are classified as *vulnerable*. In addition, four languages—Byangsi, Darma, Jad, and Jaunsari—are listed as *definitely endangered*. Bangani and Raji are categorized as *critically endangered* and *severely endangered*, respectively, while Tolcha is considered *extinct.*

Hindi currently serves as the lingua franca across the region, functioning as the principal medium for intergroup communication. In most social domains, speakers of Garhwali and Kumauni predominantly use Hindi or a hybridized form of their native language that incorporates significant Hindi lexical influence. Other smaller Indo- Aryan languages spoken in the region include Bangani, Jaunpuri, Marcha, Bawari, and Rawalti.

This paper aims to examine the processes of linguistic convergence in two Tibeto-Burman languages of Uttarakhand—*Jad* and *Raji*—with a particular focus on the outcomes of sustained contact with dominant Indo-Aryan languages, especially Kumauni, Hindi and Garhwali. It explores key mechanisms of contact-induced change, including code-mixing, borrowing, relexification, and grammatical restructuring. Through a comparative analysis of structural and sociolinguistic features, the study investigates how prolonged interaction has influenced the lexicon, syntax, phonology, and language use patterns within these endangered speech communities. By comparing these two distinct linguistic ecologies, the research highlights broader patterns and implications of language contact in marginalized and multilingual settings.

2. Jad Community: An Introduction

Several tribal communities inhabit the north-western region of Uttarakhand, among which the Jad community holds a distinct place. The term *Jad* is an exonym—used by outsiders to refer both to the community and their language—while the community members refer to themselves as *Rongba* or *Rongpa*. In official records and common discourse, they are generally categorized under the broader label *Bhotia*. As per *The Constitution (Scheduled Tribes) Order, 1967* and *Act 29 of 2000*, the Jads are recognized as a Scheduled Tribe.

The Bhotias constitute one of the major ethnic groups of the central Himalayan region. They are a transhumant community of semi-Mongoloid people of Tibetan origin (Fuchs, 1988). Owing to their close racial and cultural affinity with Tibetans, the Bhotia-inhabited regions are often referred to as *Bod* or *Bhot*. Within this broader category, the Jads are one of the most prominent sub-groups.

Historically, the Jads inhabited the Nilang and Jadong valleys along the banks of the Jad Ganga River—a tributary of the Bhagirathi—near the Tibetan border, and it is likely that their name derives from this river. Following the Indo-China War of 1962, the Indian Army relocated them to Bagori and Veerpur

[1] https://unesdoc.unesco.org/ark

Dunda villages in Uttarkashi district for security reasons (Naithini, 1986). Today, they practice seasonal migration, living in Bagori during the summer (May–September) and moving to Dunda in winter (September–April). Some families have also migrated to Chorpani near Rishikesh to safeguard livestock, access fodder, and trade woollen goods.

Although not enumerated separately in the Census of India and instead grouped under the broader Bhotia category, local records indicate a much larger population than earlier estimates. Ethnologue (2019), citing Roland Breton (1977/1997), placed their number at about 300, but by 2016, Bagori alone had around 1,009 registered voters and a total population of about 2,500 across 400 families, according to village head Bhawan Singh Rana[2].

Traditionally nomadic cattle herders and shepherds (Chatterjee, 1976), the Jads gradually established permanent settlements and took up agriculture in their original valleys. Until the 1962 border closure, they maintained active trans-Himalayan trade with Tibet through high-altitude routes such as Thang-La (5,050 m) and Tang-Choke-La (5,400 m), exchanging cotton, grain, oilseeds, metals, and sugar for salt, wool, and borax.

Today, the Jads continue to graze sheep and goats in the upper Jahnvi (Jad Ganga) valley during the summer, when alpine vegetation is abundant. While their livelihood has diversified—spanning agriculture, trade, business, and labour—their primary economic activity remains animal husbandry. Yak rearing, once common, is now limited to a few households. Wool from sheep and goats is still used to make rugs and other woollen garments, a skill particularly mastered by Jad women. In Bagori, some households also own apple orchards. However, these traditional occupations have largely declined among the younger generation, many of whom are engaged in menial jobs or have entered government service, indicating a shift towards mainstream livelihoods.

Religiously, the Jads were originally followers of Tibetan Buddhism. However, due to acculturation and proximity to predominantly Hindu populations, they now exhibit practices from both religions. They celebrate the Buddhist New Year (Losar) and display prayer flags, while also worshipping Hindu deities and engaging in Hindu rituals. Despite these influences, they maintain that they are ethnically distinct from Tibetans. Their *kul devta* (family deity) legends often resemble those found in the Hindu epic Mahabharata. This syncretism is also reflected in their medicinal traditions, with herbs and flowers believed to carry mythological associations from both Buddhist and Hindu origins. Members of the community possess skills in medicine, and other traditional practices. In times of bereavement, Buddhist priests are consulted, and funeral timings are determined using the Buddhist calendar. Marriage rituals also reflect a blend of both religious traditions. Inter-group marriages occur among Jads in Uttarkashi, Chamoli, and Kinnaur (Himachal Pradesh). Occasionally, they marry into Garhwali Rajput, Nepali, or other local communities. Internally, the Jad community is divided into two groups—one considers itself socially superior and does not intermarry with the group it perceives as lower. Notably, while all Jads are classified as Scheduled Tribes in government records, during this study, it was observed that members of the self-perceived ‘higher’ group tend to reserve the label of ‘Scheduled Tribe’ for those they consider socially inferior.

[2] As told to Ajay Kumar Singh,RP,SPPEL -during the field work.

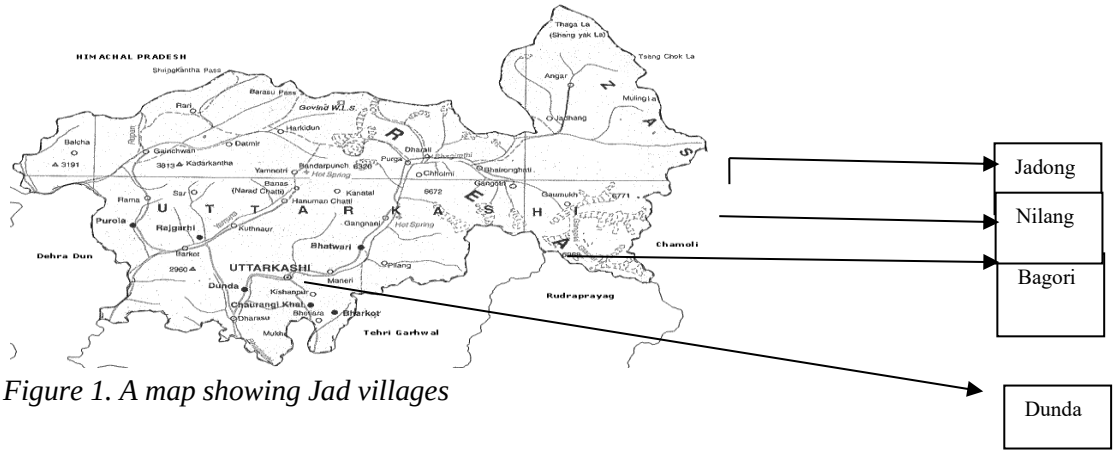


Figure 1. A map showing Jad villages

Source: http://www.eindiatourism.com/uttaranchal-india-maps/uttarkashi-uttaranchal-india-maps1.html (markings by the field researcher)[3]

2.1 Jad Language and its contact background

As mentioned earlier, the language spoken by the Jad community is commonly referred to as Jad or Jaar (ISO 639-3: *jda*), and it belongs to the Tibeto-Burman branch of the Sino-Tibetan language family. The language is known by several alternate names, including Jadh, Rongpa, Rongba, Rongma, and Bhotia, terms that are also used interchangeably to describe the community itself. This multiplicity of names reflects both external nomenclature (exonyms) and internal variation, as well as broader ethnolinguistic classifications in the Himalayan region.

Traditionally, the community possessed a rich oral literature, including songs sung during weddings and religious ceremonies, as well as narrative folktales and ritual chants. However, much of this oral heritage has been lost in recent decades. Elders in the community report that many traditional songs are no longer remembered or performed. While older speakers continue to use Jad fluently—especially in private and intra-generational conversations—the speech of younger community members is typically a heavily code-mixed variety, interspersed with Garhwali and Hindi lexical and grammatical elements. This points to an ongoing process of language shift toward the dominant regional languages.

This geographic mobility, combined with prolonged contact with speakers of other languages, has further contributed to the erosion of linguistic vitality. Without urgent intervention and community-driven revitalization efforts, the Jad language faces the possibility of further decline in both fluency and cultural function. Currently, the Jad language is in a state of endangerment, primarily due to intense contact with dominant regional languages such as Garhwali and Hindi, the latter being widely used in education, administration, and media. The Jad language is oral and lacks a writing system, which further limits its transmission, especially among the younger generation.

3. Raji: Ethnic and Linguistic Overview

The Raji tribal community, also known as Banrawats, Banmanus or Banrajis, first came into documented attention in 1823 through G.W. Traill, then Commissioner of Kumaun. Ethnically, they are believed to be descendants of the prehistoric Kiratas, early Himalayan settlers predating the Nagas and Khasas (Majumdar, 1944), though no informant in the present study confirmed knowledge of this origin. Physically, Rajis show mixed Aryan, Dravidian, and Mongoloid traits, with Driem (2002) classifying

[3] Ajay Kumar Singh

them as indigenous South Asian Mongoloids. A culturally affiliated Raji-Raute group also exists across the border in Nepal, divided into four regional clusters (Reinhard, 1974).

Traditionally animist, the Rajis today identify as Hindu, yet maintain a belief system centred around spirits and ancestors (*ama/bubu*), often viewed with fear rather than reverence. Ritual life is minimal, with only birth, marriage, and death formally marked. Marriage customs involve bride-price and a shift toward nuclear family structures, while gender roles, though formally patriarchal, see women as central to economic survival. Women primarily sustain the household through daily wage labour such as stone quarrying and grass cutting, as male participation is severely impacted by alcohol dependency.

Until recently, the Rajis led a nomadic, food-gathering life, living in caves and relying on hunting and fishing (Majumdar, 1944; Bora, 1988). Sedentarization has now set in, with communities settled in ten remote hamlets across Pithoragarh and Champawat districts. Access to education remains limited—no informant was found to have completed higher education—and economic mobility is constrained by poor infrastructure, isolation, and lack of viable skill development. Customs like leaving the dead in forests, though once practiced, are now largely abandoned.

Linguistically, Raji is a highly endangered Tibeto-Burman language with only 732 speakers reported in the 2011 Census. It is spoken in isolated hamlets such as Kimkhola, Bhogtirua, Chipltara, and others. Grierson (1909) called Raji *janggali*, and classified it under Tibeto-Burman family, a view echoed by Chatterji (1926). However, scholars like S.R. Sharma and D.D. Sharma suggest a substratum of Munda elements. Rastogi (2002, 2012) notes strong Indo-Aryan influence—particularly from Kumauni and Hindi—on Raji's grammar and vocabulary. Processes such as borrowing, code-mixing, and relexification have led to extensive lexical and structural convergence, threatening the native speech form.

3.1 Contact History of the Kumaun Rajis

The linguistic and cultural profile of the Rajis cannot be understood in isolation from their long history of contact and migration. In my view, the first major movement of Rajis into present-day Uttarakhand occurred well before the eighteenth century, when a group of hunter-gatherers crossed over from Nepal through dense forest routes. In Nepal, Raji communities continue to live in districts such as Surkhet, Baitadi, Kanchanpur, Kailali, and Bardia. Unlike their Indian counterparts, some Raute groups in Nepal still maintain a nomadic lifestyle. At the time of their initial migration, Rajis lived in caves and subsisted entirely on forest-based practices.

Historical and linguistic evidence also suggests earlier contact with Austro-Asiatic groups. Scholars such as Rev. Sten Konow and S.R. Sharma have highlighted the presence of Munda elements in Himalayan languages, including Raji. Though no Austro-Asiatic languages are currently spoken in this region, certain linguistic features point to a Munda substratum, likely absorbed in an earlier period.

Following Indian independence, the government made efforts to settle the Rajis and classify them as a ‘primitive tribe.’ In 1982, about twenty Raji individuals moved to Khirdwari in search of food and livelihood. During my fieldwork in 2022, Laxmi Rajbaar from Khirdwari recalled how they were provided with farming land and limited financial assistance to build homes. However, due to the inaccessibility of the hamlet, several families later relocated to Chakarpur, marking a second wave of migration. This movement has had lasting effects on the Raji language and cultural practices, intensifying their contact with Indo-Aryan languages and accelerating language shift.

The geographic proximity of Uttarakhand to Nepal, separated only by the Kali River, has contributed to sustained cross-border interaction. Nepali, an Indo-Aryan language, is widely spoken in this region, and many Rajis are fluent in it. Intermarriage between Indian Rajis and Nepali Raji-Raute women is also common. Thus, the region represents a unique contact zone where Indo-Aryan and Tibeto-Burman languages coexist, layered over remnants of an earlier Austro-Asiatic presence.

Even in terms of material culture, elements like stone bead jewellery and wooden utensils reflect ongoing ties with the Nepali Raji-Raute population. These markers of cultural continuity underscore a shared history that traverses both political and linguistic borders. Meanwhile, the mixed physical features of the community—ranging from Aryan to Mongoloid traits—further reinforce the idea of prolonged contact and inter-group interaction over centuries.

4. Methodology

This study is based on fieldwork conducted in Uttarkashi district (for Jad) and Pithoragarh and Champawat districts (for Raji) in Uttarakhand. Data was collected using a combination of participant observation, semi-structured interviews, and elicitation sessions with native speakers. In the Jad-speaking region, the primary field sites included Bagori and Veerpur Dunda, with supplementary information gathered in Chorpani, where some families migrate seasonally. For Raji, fieldwork was carried out in Kimkhola, Bhogtirua, Madanbori, and a recently settled area near Chakarpur. Respondents ranged from elders with fluency in traditional speech to younger speakers exhibiting significant contact-induced variation.

The study focused on:

- Lexical borrowing from Hindi, Garhwali, and Kumauni
- Code-mixed utterances across speech domains (e.g., home, marketplace, schools)
- Grammatical restructuring or simplification influenced by dominant languages

Audio recordings were made with informed consent and later transcribed and analysed using a combination of qualitative linguistic analysis and comparative lexical and morphosyntactic comparison.

5. Contact Induced Changes

5.1Contact-Induced Gain

This occurs when a language adopts new forms or patterns from another language. Such changes can include borrowing words (loanwords), grammatical structures, or phonetic elements. No aspect of a language's structure is entirely immune to linguistic influence from neighbouring speakers, particularly those from politically dominant communities. However, the lexicon is especially susceptible to influence, particularly in cases where speakers lack words for concepts that do not exist in their cultural experience. For instance, the Raji people historically had no need for words such as *farm*, *cow*, or *plough*, as these concepts were absent from their traditional way of life. Raji has borrowed many such words from Kumauni, Nepali, Hindi, Santali, and English but along with these terms they are replacing their native words.

A comparative lexical survey of 276 core vocabulary items in Raji revealed that 119 (approximately 43%) are loanwords from Indo-Aryan languages such as Hindi, Kumauni, or Nepali. Several native

words have been replaced by Kumauni/Hindi/ Nepali loans and have resulted in the loss of native speech form. Thus, the process of 'relexification'[4] (Hill & Hill 1977) can be clearly seen in Raji. For example-

English	Raji	Kumauni
Uncle	*dukkəija*	*kəkka*
Back	*bəwi*	*pʊʈʈʰi*
Mother	*ja*	*iɟa*
Father	*ba/bubu*	*babu*

Raji has numerical expressions only up to six and beyond that it has loans with little or no phonetic modification. It has also borrowed many nominals and verb forms from the dominant language. Names of days and months are also Indo Aryan loans. Except for the red and black other colour names are borrowed from Kumauni or Hindi. Apart from these kinship terms, names of body parts, expression for flora and fauna, household artifacts, adjectives, conjunctions have also been heavily borrowed. Retroflex sounds in Raji, such as /ʈ/, /ɖ/, and /ɽ/, are largely restricted to Indo-Aryan loanwords and do not appear to contrast phonemically with native alveolar or dental stops, indicating partial phonological convergence. For example- *jaɽo, boɖo, laɖo, dʰaɽe*. La polla (2006) also found the same development in other Tibeto - Burman languages of Indo-sphere.

In contrast, Jad exhibits signs of lexical convergence without extensive structural erosion. While contact with Garhwali, Hindi, and English has led to borrowing—particularly in domains previously absent in Jad culture (e.g., names of vegetables, religious terms, cultural traditions)—these remain largely surface-level additions. For example-

English	Jad	Garhwali- Hindi
School	*ɪskuːl*	*ɪskuːl*
Book	*kɪtab*	*kɪtab*
Receipt	*rəsid*	*rəsid*
Apple	*ʃeu*	*seb*
to enter new house ceremony	*gorpʊʃi*	*gorpʊʃi*

They have expressions for certain things but instead of using their own word now they have started using Garhwali words, for example-

English	Jad	Garhwali -Hindi
Root	*patoŋ*	*jəɽ*
Four o'clock	*zi*	*cʰar*
Money	*ŋʊl*	*pɛsa*
Mistake	*tsʰoʃuŋ*	*gələti*

It is interesting to note that Adjectives perform predicating function in this language. The modifier comes after the modified nouns, but due to areal influence non-numeral quantifiers like - **maŋbo** (more), ***boŋte*** (many/much), ***tsɪdʒi*** (some) are sometimes added before the modified item. For citation-

[4] The replacement of native lexicon with the dominant language.

dɪ	*pã*	***rɪŋbo***	*duk*
this	Tree	**tall**	COP

This is a tall tree/ This tree is tall.

Whereas-

maŋbo	*ləp*	*ma*	*tɔŋ*
much/ many	Talk	NEG	speak

Do not speak too much.

A common strategy for connecting two sentences in a narrative is known as **tail-head linkage**. Jad speakers use the word **/tɛ/** for this purpose, which is a borrowed form of Hindi/ **təb/.** See the following example-

hei	*ro*	*lep*	*ʃa*	*ŋa:*	*rɛ*	*hɔləŋ*	*na*	*lakəpa*	*ʧu:*
there	LOC	reach	CVB	1SG	POSS	face	and	hand	wash

tɛ	*ŋa:*	*ɖaŋsa*	*naŋ du*	*put*	***tɛ***	*ŋa:*	*təkt^{h}a*	*tsha*
then	1SG	house	Inside	go	then	1SG	flour	sieved

urã	*hei*	*ʃe*	*kolkol*	*dʒo*
And	3SG	GEN	Balls	made

'After reaching there I washed my hands and face then went inside the room and then I sieved the flour and made balls with it.'

5.2 Contact-Induced Loss

This phenomenon occurs when a language loses specific forms or patterns under the influence of another language. It often takes place during language shift, as speakers gradually abandon features of their native language, resulting in linguistic attrition or simplification.

In Raji, the loss of nasalization and glottalization can be observed in present-day usage. There are seven vowel sounds in Raji. All vowels have nasal counterparts irrespective of their proximity to nasal consonants but except in /hã/ nasalization is not phonemic in present speech form. The use of glottal sounds was prominent in the hamlet of Altodi speakers (1998, 2001) but it was rarely heard at other hamlets during later field work. In his article D.D. Sharma (p. 147) had also mentioned about this feature which is probably lost with times.

5.3 Loss or Gain of a Pattern

This refers to structural changes in syntax or grammar due to linguistic influence or internal evolution. Traditionally, Raji speakers predominantly used the Verb-Object-Subject (VOS) word order. However, contemporary usage has become more flexible, now allowing multiple word orders such as VOS, Subject-Object-Verb (SOV), and Object-Verb-Subject (OVS). The introduction of SOV and OVS orders alongside the traditionally dominant VOS may reflect influence from neighbouring Indo-Aryan languages, where SOV is the default word order, as well as increasing bilingualism among Raji speakers. Additionally, in certain cases, the subject (*S*) is dropped altogether, reflecting a shift in syntactic patterns influenced by language contact and evolving speech habits.

5.4 Hybridization

In Raji a loan pronominal phrase '***əi ke bad'*** meaning ***'after this'*** is used at the beginning of a sentence to introduce a new event. Subsequently, the Raji particle '***pəi***' marks continuity in the discourse. For example-

əi	***ke***	***bad***	***bʰitəRi***	***kəmra***	***ya***	***bi***	***ye***	***pəi***	***na***
this	GEN	after	inside	Room	LOC	come	PART	SEQ	1SG

'After this, I went inside the room.'

Jad has created several hybrid expressions, such as –

English	**Jad**	
Mustard Oil	*sərso ʃe mergʊ*	Hindi + Jad
To drink tea	cɑ *tʰʊgən*	Garhwali+ Jad

6.Key Comparative Features

Feature	**Jad Language**	**Raji Language**
Dominant Contact Language(s)	Hindi, Garhwali	Hindi, Kumaoni
Lexical Borrowing	Moderate (ritual, modern terms)	High (daily vocabulary)
Phonological Change	Some evidence	Significant loss of TB features
Syntactic Influence	Present (e.g., auxiliaries)	High (word order, case changes)
Code-Mixing (Youth)	Frequent	Very Frequent
Literacy in Mother Tongue	Low	Very Low
Language Vitality	Threatened	Nearly Endangered

7. Conclusion

The contact histories of the Jad and Raji communities reveal fundamentally different sociolinguistic trajectories, despite both speaking Tibeto-Burman languages within the multilingual landscape of Uttarakhand. Although both languages are endangered and situated in Indo-Aryan-dominated ecologies, the nature and extent of contact-induced convergence differ significantly.

The Raji, a historically nomadic hunter-gatherer community, experienced layered contact over time—with Austro-Asiatic, Tibeto-Burman, and Indo-Aryan groups. Comparative lexical data (e.g., 84-item lists showing 24 Austro-Asiatic, 52 Tibeto-Burman, and 50 Indo-Aryan cognates) reveal extensive borrowing and hybridization. This sustained interaction has influenced not only the lexicon but also core grammatical structures, indicating deep structural convergence. Raji thus exemplifies a case of intensive language contact, resulting in a stratified linguistic system wherein multiple layers of influence obscure straightforward genealogical classification.

In contrast, the Jad language retains a robust underlying grammatical framework and preserves much of its native vocabulary across key domains. This points to a scenario of moderate contact with limited structural borrowing—characteristic of a mixed language that maintains its core identity while incorporating lexical elements from other sources.

In sum, Raji exhibits significant structural convergence and relexification due to prolonged, multi-source contact, whereas Jad demonstrates a more contained convergence, preserving its linguistic integrity despite exposure to multiple languages. These divergent outcomes highlight how geography, social isolation, cultural practices, and intensity of contact collectively shape the course of linguistic change and convergence.

References

Aikhenvald, Alexandra Y. 2007. "Grammars in Contact: A Cross-Linguistic Perspective." In *Language Contact and Grammatical Change*, edited by Alexandra Y. Aikhenvald & R.M.W. Dixon, Oxford University Press.

Bora, H. S. 1988. *Cave Dwellers of Himalayas - Rajis a primitive tribal group of UP*. Paper presented in the Seminar 'Tribal Situation in Uttar Pradesh' at Yojana Bhawan. Lucknow.

Chatterji, S. K. 1926. *The Origin and Development of the Bengali Language*. Calcutta University Press.

Dixon, R.M.W.1997. *The Rise and Fall of Languages*. Cambridge University Press.

Graham Thurgood & Rany La Polla. 2003. The Sino-Tibetan Languages. Routledge, London.

Grierson G. A. 1909. *Linguistic Survey of India*. Vol iii Tibeto- Burman Family. Calcutta. Office of the Surveyor.

H.B. Naithani, 1986. Orchid Flora of North-West Himalaya. Print Media, Delhi, pp. 1-459. 4.

Hill & Hill, 1977. Language Death and Relexification in Tlaxcalan Nahuatl. Linguistics 191:55-69

Konow, Sten. (1909). *On the Classification of the Tibeto-Burman Languages*. WORD Volume 11, Issue 1, on pages 94-111.

Majumdar, D.N.1944. *Races and Cultures of India.* Universal Publishers. Lucknow.

Rastogi, K. 2002. b *Raji: Language of a Vanishing Himalayan Tribe*. *Linguistics of Tibeto-Burman Area*. Berkeley Area, Berkeley. USA Vol. 25.2. Fall 2002.

Rastogi, K. 2012 A Descriptive Grammar of Raji, Aviram Prakashan, New Delhi.

Rastogi, Shukla, S. & Singh A. 2017 Some Aspects of Jad: An Endangered Language, *Vaak Manthan*, vol II, Issue-I, pp 1-7

Sharma, D. D.1994. *A Comparative Grammar of Tibeto- Burman Languages of Himachal Pradesh and Uttarakhand.* Studies in T-B languages iv. New Delhi, Mittal Publication.

Sharma, S.R. 1987. *Madhya Himalaya Ki Raaji Boli*, Pahaar-2, Nainital.

Thomason, Sarah G., & Kaufman, Terrence.1988. *Language Contact, Creolization, and Genetic Linguistics*. University of California Press.

Winford, Donald. 2003. *An Introduction to Contact Linguistics*. Blackwell Publishing.

वाक् मंथन
ISSN: 2426-2149

A Sociolinguistic Study of Sundargarh Sadri

Asmita Padhan
Jawaharlal Nehru University, Delhi
asmita94_llh@jnu.ac.in

Abstract

This study explores the sociolinguistic dynamics of Sadri, a lingua franca among diverse tribal populations in Sundargarh, Odisha. Sadri, an Indo-Aryan language, coexists with Sambalpuri (Indo-Aryan), Oraon and Kurukh (Dravidian), and Mundari (Austro-Asiatic), forming a complex multilingual setting that fosters language contact phenomena such as code-switching, lexical borrowing, and structural convergence.

The paper focuses on the alienable/inalienable possession distinction in Sadri, particularly examining the usage of the enclitic marker -*har*. Preliminary findings indicate that its use is inconsistent among adult speakers and almost absent in children's speech. This decline may point to a broader grammatical shift toward simplifying the possessive system, potentially reducing the alienable/inalienable distinction. Such a shift could lead to a more neutral possessive structure, aligning Sadri with other regional languages that do not maintain this distinction.

Phonological and morphological convergence also emerges as a significant feature in Sundargarh Sadri. Vowel shifts, such as [ʌ] > [ɑ] (e.g., *lʌgin* → *lagin*), and consonantal modifications in borrowed lexemes (e.g., rɔk → ɔʈik) demonstrate phonological adaptation.

This study highlights the need for more extensive, comparative research to map linguistic variation across Sadri-speaking regions and to identify factors driving these emerging structural changes.

Keywords: Language Contact, Language Variation, Multilingualism, Alienable/Inalienable Possession, Sadri.

1. Introduction

India boasts one of the most diverse linguistic landscapes in the world, with 456 languages spread across six different language families (Ethnologue, 2024). Given this linguistic richness, it is unsurprising that most Indians are multilingual. In such a context, the use of lingua francas i.e., link languages that facilitate communication, becomes quite common and often region-specific. For example, Hindi functions as a lingua franca for many states in North India.

This study focuses on a particular lingua franca, which, although not as widespread as the official language, plays a significant role in its region. Sadri is an East Indo-Aryan language predominantly spoken in the states of Jharkhand, Odisha, and Chhattisgarh. It serves as a lingua franca for the large tribal populations residing in these areas. Beyond these states, many ethnic communities in Assam, West Bengal, the Andaman and Nicobar Islands, Bangladesh, Nepal, and Bhutan also use Sadri. Outside India, the language is often called 'Sadani'.

1.1 Linguistic Background

According to Ethnologue (Eberhard et al., 2019), Sadri has about 12.13 million speakers in India, with 5.13 million L1 speakers and 7 million L2 speakers. Over 58% are L2 speakers, leading to significant linguistic variation influenced by native languages. Its role as an inter-group communication tool has caused contact-induced changes, especially where Sadri coexists with Sambalpuri, Hindi, and tribal languages (Ravindranath, 2015). Studies (Peterson & Baraik, 2022) note phonological, morphosyntactic, and lexical differences among Sadri ethnic groups, shaped by regional, social, and historical factors.

Despite its widespread use, Sadri is not officially recognized in Jharkhand, Odisha, or Chhattisgarh. While Jharkhand law permits its use in early education, implementation has been lacking (Odisha Review, 2010). Nevertheless, Sadri continues to flourish through literature, folk traditions, music, and digital content, helping maintain its linguistic resilience despite Hindi's dominance.

1.2 Geographical Background

Sundargarh, located in northwestern Odisha, is the state's second-largest district by area and shares borders with Jharkhand and Chhattisgarh. Over 50% of its population comprises Adivasi communities, making it one of Odisha's most tribal-dominated regions. These communities speak various tribal languages, including Oraon (Kurukh) and Kisan from the Dravidian family, and Mundari and Kharia from the Austroasiatic family.

In this linguistically diverse setting, Sadri has emerged as the primary lingua franca, facilitating communication among different tribal and non-tribal groups. Its widespread use is attributed to historical migration, economic exchanges, and social interactions. Over time, Sadri has incorporated elements from regional languages, giving it a distinct identity in Sundargarh. In urban centres like Rourkela and Sundargarh town, linguistic interactions are dynamic, with frequent mixing between Sadri, Sambalpuri, and Hindi. Conversely, in

rural and tribal areas, Sadri is used more consistently in its traditional form for inter-community communication.

2. Aims and Objectives

The study aims to document the structural features of inalienable possession in Sundargarh Sadri, including its morphosyntactic form, distribution, and function in natural discourse. A comparative approach will be employed to analyze whether the inalienable possession marker observed in Sundargarh Sadri is an innovation or retention when compared to other Sadri dialects and related Indo-Aryan languages. It also seeks to investigate the linguistic variations within Sadri as spoken in Sundargarh, examining phonological, morphosyntactic, and lexical aspects that deviate from standardized forms.

The research further explores sociolinguistic variation in the use of this marker across different age groups (younger vs. older speakers). It will also investigate key factors driving linguistic change, particularly the role of language contact with other Indo-Aryan languages like Odia, Sambalpuri, and Hindi. The study aims to contribute to broader theoretical discussions in possession typology, language shift, and the concept of India as a Linguistic Area.

2.1 Research Questions

a. How does Sundargarh Sadri differ from the Jharkhand variety, particularly in its use of inalienable possession, and what are the motivations behind these variations?

b. How does the use of inalienable possession markers vary across social groups in Sundargarh, and what sociolinguistic factors drive this variation?

c. What are the phonological, morphosyntactic, and lexical variations in Sundargarh Sadri, and how do these features deviate from the standardized variety?

2.2 Research Methodology

The most effective and appropriate method for studying language variation due to contact is through a diachronic analysis, drawing on historical texts. However, in most cases, such data are either unavailable or insufficient for conclusive reconstruction. Although some historical data on Sadri exist, they primarily focus on the Jharkhand variety, which is considered the standard form. Literary works by Sadri speakers from Odisha, particularly from Sundargarh, are scarce. As a result, this study will adopt a synchronic approach, relying on contemporary linguistic evidence to examine linguistic transfer. While lexical borrowing is relatively easy to identify, the replication of grammatical meaning is more complex and often subject to debate.

2.2.1 Sampling Method

A purposive sampling strategy was employed to ensure that all participants were native speakers of Sadri. Selection criteria included long-term residence in Sundargarh and self-identification as Sadri speakers. This method ensured linguistic authenticity and contextual relevance in data collection.

2.2.2 Participant Selection and Grouping

A total of 40 participants were selected for the study, divided into two age-based groups to examine sociolinguistic variation across generations

- 20 children (under 15 years of age)
- 20 adults (25 years and above)
  All participants were permanent residents of Sundargarh town.

2.2.3 Data Collection

Data were collected using a pre-designed questionnaire specifically constructed to elicit responses containing the inalienable possession marker *-har*. The questionnaire also included items targeting variation in phonology, morphosyntax, and lexicon. Given practical constraints and limited access to naturalistic speech settings, data collection took place primarily through

- Targeted individual interviews
- Telephonic interviews
  These methods, while limiting the opportunity for spontaneous language use, allowed for controlled and consistent elicitation of relevant linguistic structures.

2.2.4 Supplementary Sources

In instances where direct input from speakers of the Chotanagpuri Sadri variety (Jharkhand) was unavailable, the study consulted authoritative secondary sources. Notably, *A Grammar of Chotanagpuri Sadri An Indo-Aryan Lingua Franca of Eastern Central India* by John Peterson and Sunil Baraik (2022) was used for comparative analysis.

While lexical borrowing could be relatively easily identified in the collected data, tracing grammatical replication (such as the spread of inalienable possession constructions) required more nuanced analysis. The study acknowledges that the replication of grammatical meaning remains contentious and is often harder to prove conclusively than lexical transfer. Nevertheless, the synchronic approach, supported by carefully selected participants and methodical data collection, provided a reliable foundation for examining current patterns of linguistic variation in Sundargarh Sadri.

3. Literature Review

Language contact significantly influences linguistic structures, often resulting in changes to lexicon, phonology, syntax, and discourse. Thomason, (2001: 62) defines contact-induced change as any shift more likely to occur due to language contact, categorizing it as either *system-preserving* (e.g., lexical borrowing) or *system-altering* (e.g., grammatical restructuring). Winford (2003) identifies three contact scenarios—language maintenance, shift, and creation—with *maintenance* particularly relevant where a minority language adapts under the influence of a dominant one.

Contact-induced language change occurs when linguistic shifts arise as a result of interaction between languages. Particularly, when changes would have been unlikely outside the contact situation. Contact can also accelerate or support changes that might otherwise emerge independently. To establish whether a given linguistic feature has arisen due to contact requires evaluating the historical relationship between languages and the extent of exposure among speaker communities.

The two most important factors that scholars must take into consideration while studying contact and variation are identifying evidence of transfer and assessing whether the change could have occurred independently. Since historical linguistic data is often unavailable, comparative evidence is crucial. If a language exhibits structural properties absent in its genetic relatives but present in a neighboring contact language, transfer becomes a plausible explanation. However, certain grammatical features, like the emergence of articles, occur naturally in many languages without contact influence, necessitating additional evidence to establish causality.

Sociolinguistic factors also play a role in grammatical replication. Language contact is often more pronounced among urban speakers, younger generations, and male speakers, as they tend to have greater exposure to linguistic influences. Patterns of variation across demographic groups may thus indicate stages of contact-induced change.

Ultimately, the strength of any claim about contact-induced grammatical replication depends on the accumulation of linguistic, historical, and social evidence. The more

converging indicators there are, the more convincingly contact influence can be established.

Sadri was historically classified as a dialect of Bhojpuri (Grierson 1903; Tiwari 1960; Jordan-Horstmann 1969). However, recent scholarship disputes this view. Linguistic evidence shows that Sadri shares a closer genetic relationship with Khortha, Kurmali, and Panchparganiya. They form a distinct subgroup known as the Sadani languages (Yadav 2012; Peterson & Baraik, in press). These languages exhibit internal structural similarities and diverge significantly from Bhojpuri, Magahi, and Maithili.

Sadri serves as a lingua franca in western and central Jharkhand, especially among tribal groups from Munda (e.g., Kharia, Mundari, Bhumij) and Dravidian (e.g., Kurukh, Gondi) language families. Many members of these groups now speak Sadri as their first language,

having shifted away from their heritage languages (Peterson & Baraik, in press). As a result, second-language speakers now comprise approximately 58% of the total Sadri-speaking population in India and abroad.

This extensive use as a lingua franca has led to noticeable linguistic influences. Sadri/Nagpuri has been significantly shaped by contact with Munda and Dravidian languages, especially in phonology, syntax, and lexicon. Despite its role as a recipient of influence, Sadri has also provided loanwords to neighboring languages, often in core semantic domains such as kinship, body parts, and numerals (Abbi, 1997).

Morphologically, Sadri/Nagpuri differs from other Sadani languages by lacking ergativity. Except for a few experiential predicates, it follows a nominative-accusative alignment, where the subject of both transitive and intransitive verbs appears in the unmarked (direct) case, while the object takes the oblique marker =ke when definite or human. This contrasts with the ergative constructions still preserved in Khortha, Kurmali, and Panchparganiya, suggesting that Sadri may have lost ergativity under the influence of non-ergative Munda and Dravidian languages.A particularly unique feature of Sadri/Nagpuri is its morphosyntactic distinction between alienable and inalienable possession—a trait rare in Indo-Aryan languages but common in Munda languages.

4. Alienable and Inalienable possession

4.1 Inalienable Possession

Inalienable possession refers to relationships between a possessor and a possessum that are considered intrinsic, inherent, or permanent. These possessums are typically not subject to transfer or separation. Common semantic domains for inalienable possession include

- Kinship terms (e.g., mother, brother, daughter)
- Body parts (e.g., hand, heart, eye)
- Part-whole relationships (e.g., roof of a house)
- Personal attributes or names

In many languages, inalienably possessed nouns are obligatorily marked for a possessor, often using special morphological markers. These constructions may lack a genitive construction entirely and often rely on a close syntactic bond between the noun and the possessor (e.g., no use of a possessive marker, or special agreement morphology). In some languages, these nouns cannot stand alone and must appear with a possessor.

For example, in some Oceanic and American indigenous languages, body parts and kin terms require direct possessive marking (e.g., his eye, not just eye), as they are viewed as fundamentally “owned.”

4.2 Alienable Possession

Alienable possession, by contrast, refers to non-intrinsic relationships between a possessor and possessum—ones that are more contingent, temporary, or socially constructed. This includes

- Objects (e.g., book, car, clothes)
- Property (e.g., house, money)
- Animals
- Abstract possessions (e.g., rights, ideas)

These nouns are usually not required to appear with a possessor and can often be referred to independently. In languages that distinguish between alienable and inalienable possession, alienable possession typically involves more explicit or "fuller" possessive constructions—such as possessive pronouns or genitive case markers—because the relationship must be overtly stated.

For instance, while my hand (inalienable) might be expressed with a possessive suffix, my book (alienable) may require a genitive phrase or a different construction altogether.

4.3 Structural Marking Differences

Haspelmath emphasizes that many languages around the world mark this distinction grammatically—often through

- Different possessive pronouns or affixes
- Different constructions (e.g., juxtaposition vs. genitive particles)
- Obligatoriness of possessive marking for in-alienables

However, not all languages mark this distinction. In languages like English, for example, the same possessive construction (my book, my hand) is used for both types, though the semantic difference is still intuitively understood.

4.4 Possession in Sadri

Sadri overtly marks inalienable possession through *-har,* especially with kinship terms and body parts (Nowrangi 1956 161; *Language Dynamics and Change* 12, 2022 240). In such constructions, the possessum (head noun) bears the *-har* marker, while the possessor may optionally appear as a genitive-marked NP. For example,

ghʌr jʌnana-har puch-l-ʌk
'He returned home and his wife asked …'
hã̃ pʌrdip = kʌr aba-har
'Yes, Pradip's dad.' (Peterson, 2022, p. 240)

In contrast, alienable possession uses only a genitive-marked possessor, and the possessum reflects only number and case, without *-har*

u=mʌn=ʌk jʌmin jaydad=ke bãʈ-ek=hʌe
'One should divide up their wealth (lit. land [and] belongings of those should be divided).' (Peterson, 2022, p. 240)

4.4.1 Origin and Typological Comparison

Definiteness—or more precisely, specificity—has also been marked in many eastern Indo-Aryan languages through classifiers attached directly to nouns, typically a feature of numeral modification (Neukom & Patnaik 2003 24ff.). Sadri shows this mechanism as well. For instance,

u=kʌr bad dhan=mʌn=ke lʌʈki=go
'After that, the girl boiled the rice paddy, dried [it] …' (Neukom & Patnaik, 2003, p. 24)

Here, the classifier =go on *lʌʈki* ('girl') marks her as previously mentioned and referential.

The enclitic *-har* in Sadri likely derives from such a classifier, initially used to signal specificity or definiteness, and later grammaticalized into a marker of inalienable possession. This diachronic pathway is further supported by parallels in Chhattisgarhi, where a cognate form of *-har* appears to function as a definiteness marker (Tiwari, 1960), suggesting a broader areal pattern across the eastern Indo-Aryan region.

Notably, first- and second-person possessors in Sadri already use genitive constructions that convey definiteness inherently, making *-har* functionally redundant in these contexts. Thus, *-har* became restricted to third-person possessors, where such disambiguation remained salient.

4.5 Possession in Sundargarh Sadri

The primary objective of this preliminary study is to investigate whether Sadri speakers in Sundargarh maintain the alienable possession feature. A previous pilot study indicated that certain speakers refrain from using the inalienable possession marker, which sparked the motivation for this research. Detailed data, findings, and analysis are presented in the following section.

5. Findings and Analysis

5.1 **Adult Participants** (Ages 25 and above)

The distribution of the enclitic *-har* among the participants shows variation. This indicates that its use is not uniform across speakers. Of the twenty adult participants in the study, eleven (55%) did not employ *-har* at all. The remaining nine participants (45%) made use of *-har* to varying degrees. Among these, six participants (30%) used it exclusively in

contexts involving human possessors, pointing to a possible semantic restriction wherein *-har* is associated with human referents. In contrast, three participants (15%) applied *-har* in both human and non-human possessive constructions, indicating a broader and a more general usage. This distributional pattern suggests that the use of *-har* is subject to variation not only in terms of frequency but also in terms of semantic scope. Examples-

a. Group 1 – Used enclitic for +human possessors only

i. u-kar dʒani-har iskul-mɛ kam kɔrɛla
he.3S-GEN wife.3S-POSS school-LOC work work.HAB.3S
His wife works at the school.

ii. rahul-ker goɽ-har ʈuiʈ dʒahe
rahul.3S-GEN leg.3S-POSS break be.PRF.3S
Rahul's leg is broken.

iii. kukur-ker pũtʃʰri-∅ keriya ahe
dog.3S-GEN tail.3S-POSS black be.PRES.3S
The dog's tail is black.
(No enclitic used for non-human possessor)

b. Group 2 – Used enclitic for both +human and -human possessors

iv. u-man-ker saŋ-har dɪlli-mɛ rɔhɛla
they-3Pl-GEN friend.3S-POSS delhi.LOC live.HAB.3S
Their friend lives in Delhi.

v. kukur-ker pũtʃʰri-har keriya ahe
dog.3S-GEN tail.3S-POSS black be.PRES.3S
The dog's tail is black.

c. Group 3 – Did not use enclitic at all

vi. u-kar kẽs-∅ ləmba ɔr keriya ahe
she.3S-GEN Hair long and black be.PRES.3S
Her hair is long and black.

vii. prədip-ker aba-∅ bemar ahe
pradip.3S-GEN father.3S-POSS sick be.PRES.3S
Pradip's father is sick.

Table 1: Summary of Adult Participants

Group	Number of Participants	Percentage
Did not use *-har* at all	11 / 20	55%
Used *-har* only for +human possessors	6 / 20	30%
Used *-har* for +human & -human	3 / 20	15%

5.2 **Child Participants** (Under 15 Years)

In the parallel group consisting of child participants, the use of the enclitic *-har* was further limited. Out of twenty children, only four (20%) used the marker in their speech, a much lower rate compared to the adult group. Notably, none of the children used *-har* in contexts involving non-human possessors. This suggests a more restricted usage of the form. Examples-

a. Group 1 – Did not use the enclitic at all

viii. u-kɑr pətni-∅ iskul-mɛ kɑm kɔrɛlɑ
he.3S-GEN wife.3S-POSS school-LOC work work.HAB.3S
His wife works at the school.

ix. rɑhul-ker goɽ-∅ ʈʊiʈ dʒɑhe
rahul.3S-GEN leg.3S-POSS break be.PRF.3S
Rahul's leg is broken.

x. kʊkʊr-ker keriyɑ pũʈʃʰri-∅ ɑhe
dog.3S-GEN Black tail.3S-POSS be.PRES.3S
The dog's tail is black.

b. Group 2 – Used the enclitic for only +human possessors

xi. pərdip-ker ɑbɑ-hɑr bemɑr ɑhe
pradip.3S-GEN father.3S-POSS sick be.PRES.3S
Pradip's father is sick.

xii. kʊkʊr-ker pũʈʃʰri-∅ keriyɑ ɑhe
dog.3S-GEN tail.3S-POSS black be.PRES.3S
The dog's tail is black.

xiii. rahul-ker goɽ-har ʈʊiʈ dʒahe
rahul.3S-GEN leg.3S-POSS break be.PRF.3S
Rahul's leg is broken.

Table 2: Summary of Child Participants

Group	Number of Participants	Percentage
Did not use *-har* at all	16 / 20	80%
Used *-har* only for +human possessors	4 / 20	20%
Used *-har* for -human possessors	0 / 20	0%

5.3 Findings & Discussion

The use of the enclitic marker among adults is inconsistent, while it is almost entirely absent in children speech. This pattern points toward a potential generational decline in the productive use of *-har*, particularly in its extended functions. This may be indicative of an ongoing grammatical shift or reanalysis in the language. Sundargarh Sadri may be undergoing a simplification in its possessive system, possibly alleviating the alienable/inalienable distinction.

One possible explanation for the inconsistent use of the alienable possession marker *-har* in Sadri is the influence of contact languages like Sambalpuri and Odia, which do not distinguish between alienable and inalienable possession. This lack of distinction may contribute to the marker's variable use among Sadri speakers in Odisha. However, a contact-based explanation alone may be insufficient. Sadri in Jharkhand is also in contact with Hindi and Khortha, languages that similarly lack this distinction, yet *-har* usage still varies across regions. This suggests additional sociolinguistic or internal grammatical factors may be at play.

Nevertheless, this study is based on a small, regionally limited sample. While the trends are significant, they cannot be generalized to all Sadri speakers. Broader, comparative, and longitudinal research is needed to map the extent of variation, identify consistent patterns, and determine the sociolinguistic or grammatical factors behind this shift. This preliminary study highlights the need for further investigation into Sadri's evolving structure and its interaction with neighboring languages.

6. Further Findings

Beyond the use of the possessive enclitic, this study also identified several other linguistic variations in Sadri that can be attributed to language contact. The practise of code-switching and code-mixing is common, which is unsurprising considering the extent of

multilingualism. Lexical variation is also expected under such conditions. However, notable phonological and morphological variations point to deeper patterns of contact-induced change.

One prominent area of variation is lexical borrowing, particularly of verbs. Borrowed verbs are frequently adapted to align with Sadri's phonological rules. For instance, the English verb *stop* undergoes cluster reduction in Sadri. In the Sundargarh variety, this results in the form *ɔʈik*, where an epenthetic vowel *i* is inserted to break the consonant cluster, followed by apocope (deletion) of the final vowel. Other verbs show signs of metathesis, where phonemes are reordered without addition or deletion, reflecting a strategy for phonological compatibility. Importantly, Sadri generally avoids word-final vowels. This pattern suggests a case of phonological integration, wherein loanwords undergo adaptation to conform to the accepted phonological patterns of the recipient language. In such instances, the word-final vowel 'i' shifts to the word-medial position, goes through vowel shortening, and forms a diphthong.

Table 3: Phonological Integration in Sundargarh Sadri

English Gloss	Jharkhand Sadri	Sundargarh Sadri	Sambalpuri
stop	rɔk	ɔʈik	ɔʈkɑ
lɑn	lɑn	ɑɪn	ɑni
fill	bʰin	bʰɔɪr	bʰɔri
enter	hɛl	ɖʰuɪk	ɖʰuki

More examples of phonological variations

Table 4: Vowel Shift (e -> i)

English Gloss	Jharkhand Sadri	Sundargarh Sadri	Sambalpuri
1st person pronoun	mõe	mõi	muĩ
2nd person pronoun	tõe	tõi	tui

The close-mid front unrounded vowel [e] changes into the close front unrounded vowel [i] as exemplified by the 1st and 2nd person pronouns.

Table 5: Vowel Shift (ʌ -> ɑ)

English Gloss	Jharkhand Sadri	Sundargarh Sadri	Sambalpuri
for	lʌgin	lɑgin	lɑgi

and	ʌur	ɑur	ɑu
you (hon)	rʌure	rɑure	tɔmɛ

The open-mid back unrounded vowel [ʌ] changes into [ɑ], the open back unrounded vowel. In the first two instances, the phonological changes can plausibly be linked to Sambalpuri influence. However, in the case of the second-person honorific pronoun, the observed variation cannot be directly associated with Sambalpuri. This particular instance might serve as evidence of synchronic variation, indicating a consistent sound change occurring within the language. An analogous phonological variation can be identified in the case of the third-person singular past morpheme. A similar shift is evident in the third-person singular past morpheme, which appears as /ʌk/ in Jharkhand Sadri and as /ɑk/ in Sundargarh Sadri, as seen in the following sentence pairs

xiv. u gir ge-l-ʌk
s/he-3SG-NOM fall-V1 go-V2-PST-3SG
He fell down.
(Jharkhand Sadri)

xv. u gir ge-l-ɑk
s/he-3SG-NOM fall-V1 go-V2-PST-3SG
He fell down.
(Sundargarh Sadri)

Together, these phonological and morphological changes indicate a complex interplay of language contact and internal evolution within Sadri.

7. Conclusion

This sociolinguistic investigation of Sundargarh offers valuable insights into the complex and dynamic linguistic landscape of a region that remains relatively underexplored in scholarly research. By examining variables such as possessive markers, the study reveals significant intra-linguistic variation within comparable linguistic ecologies. Data collected from diverse social groups demonstrate that while Sadri continues to serve as a vital marker of cultural identity, it is increasingly subject to the influence of dominant regional languages, resulting in patterns of code-switching, lexical borrowing, and gradual language shift.

The findings emphasize the urgency of documenting and analyzing underrepresented languages like Sadri, both to enrich the understanding of linguistic diversity and to safeguard intangible cultural heritage. As language change unfolds under sociolinguistic pressures, particularly in multilingual settings, ongoing research is essential to uncover

how grammatical features such as possession evolve and what these shifts reveal about speaker identity and community dynamics.

In conclusion, this study contributes meaningfully to the broader field of sociolinguistics by foregrounding the lived realities of linguistic variation and contact. Acknowledging the inherent value of Sadri can inform future efforts toward its preservation and revitalization, ensuring that its rich linguistic and cultural traditions endure within an ever-changing socio-linguistic landscape.

8. List of Abbreviations

1SG – First person singular
2SG – Second person singular
3SG – Third person singular
3PL – Third person plural
GEN – Genitive
LOC – Locative
NOM – Nominative
POSS – Possessive
PST – Past
PRES – Present
PRF – Perfect
HAB – Habitual
IPA – International Phonetic Alphabet
L1 – First language
L2 – Second language

References

Ethnologue. 2024. *Languages of India*. Dallas: SIL International.

Haspelmath, Martin. 2008. Syntactic universals and usage frequency. Leipzig Spring School on Linguistic Diversity, March.

Odisha Review. 2010. *Linguistic diversity in Odisha*. Bhubaneswar: Governments of Odisha.

Winford, Don. 2003. *An introduction to contact linguistics*. Oxford: Blackwell.

Winford, Don. 2010. Contact and borrowing. In Raymond Hickey (ed.), *The handbook of language contact*. Chichester: Wiley-Blackwell, 170–187.

Plural Formation in Bagri: An Areal Perspective of Phonological Patterns

Sanya Jakhar &
Dr. V. Madhava Subramanya Sharma
University of Delhi, Delhi
sanya.jakhar45@gmail.com
vmssharma@gmail.com

Abstract

Bagri is a dialect spoken in North-western India and travels across 3 states: Haryana , Rajasthan and Punjab. It Covers around 14 districts, which is known as the Bagar region. Bagri is a rich dialect showing unique ways of forming plurals that are influenced by the linguistic environment of North Western India. Instead of using separate plural markers, Bagri often relies on phonological changes such as vowel lengthening, consonant shifts, and tonal variations to indicate plurality. These patterns are similar to those found in neighboring languages like Haryanvi, Punjabi, and Rajasthani, suggesting that Bagri's plural formation has been shaped by language contact and interaction.

This study examines how phonological processes influence Bagri's plural formation system, and also how these processes represent larger patterns in Indian language convergence. The study analyzes spoken data from native Bagri speakers to identify major sound changes that occur when nouns shift from singular to plural. The findings show that Bagri's pluralization strategies are not only inherited from its Indo-Aryan roots but are also influenced by the languages spoken around it. This supports the idea that languages in India do not develop in isolation but rather influence each other through interaction over the course of time.

The paper presents a detailed analysis of these phonological processes, supported by field data, and demonstrates how Bagri's pluralization system reflects both genetic inheritance and areal influence. By examining Bagri's plural formation from an areal linguistic perspective, this paper highlights how phonological adaptation plays a role in shaping languages within shared geographical space. The study contributes to our understanding of India as a linguistic area, with languages from numerous regions and backgrounds sharing patterns as the outcome of centuries of interaction. These findings help us better understand the complexities of linguistic diversity and convergence in India.

Keywords: Bagri language, Plural formation, Phonological processes, Indo-Aryan dialects, Areal linguistics, Epenthesis, Tonal morphology, Morphophonemic variation.

1. Introduction

Bagri is a dialect spoken in North-western India and travels across 3 states : Haryana, Rajasthan and Punjab. It covers around 14 districts , which is known as the Bagar region. This dialect exhibits unique phonological processes in pluralization that makes it different from Standard Hindi and other Indo-Aryan languages. Instead of standing heavily on affixation, semantic or syntactic processes including dedicated plural morphemes, Bagri frequently employs phonological modifications, including vowel lengthening, consonant alternations, nasalization, and tonal variations, to indicate plurality.

Plural formation is a critical area of phonological research because it reveals underlying patterns of sound change, morphophonemic interactions, and areal linguistic influence. Bagri belongs to the Indo-Aryan language family which shows a diverse range of pluralization strategies. However, what makes Bagri's system more interesting is its contact and influence from the neighboring languages such as Haryanvi, Punjabi, and Rajasthani, which helped in shaping its phonological structure.

This paper aims to analyze the phonological processes involved in plural formation in Bagri through data and theoretical explanations. By examining vowel alternation, consonant shifts, nasalization, syllabic changes, and areal influences, we can understand how Bagri pluralization functions within the larger Indo-Aryan linguistic landscape.

1.1 phonetic and phonological features of Bagri

Phonemic Inventory

Bagri shares most of its phonemic inventory with Hindi but differs in certain vowel and consonantal processes that are involved in pluralization

VOWELS:

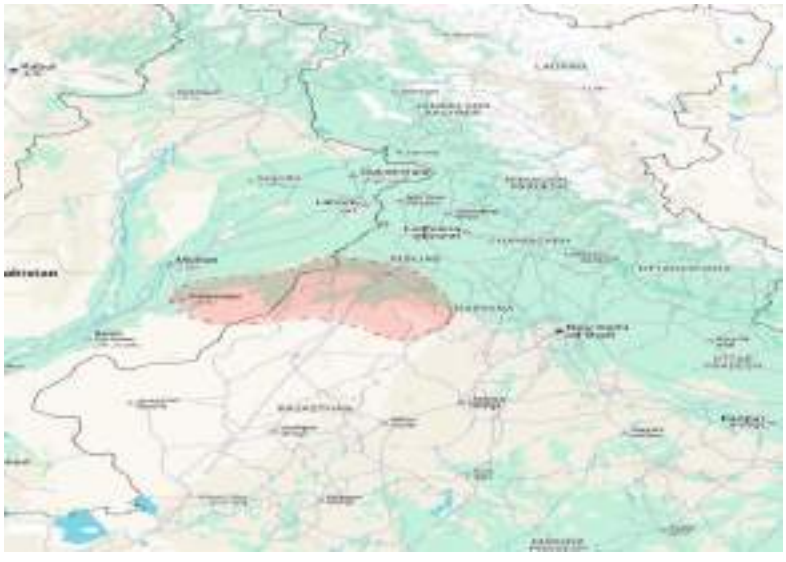
- Short vowels: /ɪ, ɛ, ə, ʊ, ɔ/
- Long vowels: /i, e, a, o, u, æ/
- Nasalized vowels: /ã, ĩ, ũ,æ̃/

CONSONANTS

- Stops: /p, b, t, d, ʈ, ɖ, k, g/
- Aspirated stops: /pʰ, bʱ, θ, ð, ʈʰ, ɖʱ, kʰ, gʱ/
- Fricatives: /s, ʃ, v, f, h/

- Affricates: /tʃ, dʒ/
- Liquids: /ɽ, r, ɭ, l/
- Nasals: /m, n, ɳ, ŋ/
- Glides: /j/

*Nasalization plays a crucial role in forming plurals.



*Map 1 of India and Pakistan showing the highlighted regions where Bagri is spoken

Rajasthan: Anupgarh, Sri Ganganagar, Jhunjhunu, Hanumangarh, Khajuwala, Chattargarh, Loonkaransar, Pugal, Sri Dungargarh, Bikaner, Taranagar, Sardarsahar, Rajgarh, Sidhmukh, Bhanipura, Ratangarh, Churu
Haryana: Sirsa, Fatehabad, Barwala, Adampur, Hisar, Siwani, Loharu, Bhiwani, Bhadra
Punjab: Abohar, Fazilka, Mukhtsar
Pakistan: Bhawalpur, Bahwalnagar

2.1 Aims and Objectives

This study seeks to explore and document the plural formation strategies in the Bagri dialect, focusing specifically on the phonological processes involved and the impact of areal linguistic influences. Bagri, a Northwestern Indo-Aryan dialect spoken across parts of Rajasthan, Haryana, and Punjab, offers a unique system of pluralization that differs significantly from other Indo-Aryan languages such as Standard Hindi. Rather than relying heavily on suffixation or inflectional morphemes, Bagri frequently employs phonological modifications—such as vowel lengthening, tonal shifts, consonant changes, and

epenthesis—as primary strategies for marking grammatical number. This marks a departure from more morphologically driven plural systems and places Bagri within a broader typological category of languages that rely on phonological, rather than morphological, cues for plurality.

The aim of this research is to provide a systematic and descriptive account of these pluralization strategies in Bagri, examining the various phonological mechanisms at play, and to investigate how these strategies have evolved through language contact with neighboring dialects such as Punjabi, Haryanvi, and Rajasthani. The research positions Bagri as a critical linguistic variety within the South Asian context, illustrating how languages in geographical proximity often converge on similar structural solutions despite genetic differences.

To achieve this aim, the study pursues the following objectives:

- To document the phonological processes involved in plural formation in Bagri, with special focus on vowel alternation, consonantal fortition and lenition, epenthesis, tonal modification, reduplication, deletion, and metathesis. Each of these strategies will be analyzed in terms of their phonetic properties and functional roles in marking plural forms.

- To analyze the distribution and conditioning of these phonological processes across different noun classes, syllable structures, and phonological environments. This includes identifying which processes tend to co-occur, under what linguistic circumstances they are triggered, and what functional load they carry in the grammar.

- To explore the extent of areal influence from neighboring dialects and languages on Bagri's pluralization system. By comparing Bagri data with plural formation patterns in Punjabi, Haryanvi, and Western Rajasthani, the study aims to identify shared areal features and distinguish them from internally motivated phonological developments.

- To examine how Bagri plural formation aligns or diverges from broader Indo-Aryan morphological traditions. The study will evaluate to what extent Bagri's strategies are inherited features, contact-induced innovations, or typological anomalies within the Indo-Aryan family.

- To contribute to the theoretical understanding of phonologically driven plural systems, particularly in the context of linguistic areas. This includes examining how phonology can serve grammatical functions typically fulfilled by morphology, and how such systems arise and stabilize within language contact zones.

- By fulfilling these objectives, the research not only adds to the descriptive knowledge of the Bagri dialect but also informs general linguistic theory concerning plural formation, language contact, and phonological typology.

2.2 Research Questions

The current study is driven by the overarching goal of understanding how pluralization is handled in Bagri through phonological means rather than traditional morphological mechanisms. Grounded in both descriptive and theoretical interests, the study raises several specific research questions that guide its analytical direction.

What are the primary phonological processes employed by the Bagri dialect to mark plurality?

This question focuses on identifying the various strategies such as vowel lengthening, tonal modifications, consonant changes (fortition and lenition), epenthesis, deletion, metathesis, and reduplication. It seeks to explore which processes are most frequent, whether they occur in isolation or combination, and how they function to encode the singular-plural distinction.

How are these phonological processes distributed across different lexical items and phonological environments in Bagri?

This question examines the conditioning factors of plural formation. It asks whether certain syllable structures, word-final consonant clusters, or stress patterns are more likely to trigger specific phonological changes. It also looks into whether plural marking is uniform across all noun classes or if variability is linked to factors such as semantics, frequency, or phonotactics.

What evidence exists for the influence of neighboring languages such as Punjabi, Haryanvi, and Rajasthani on the pluralization patterns found in Bagri?

This research question investigates the areal dimension of Bagri's pluralization system. It asks whether the observed phonological strategies can be traced to language contact and convergence, and if similar processes are present in related dialects of the Bagar region. It also asks how Bagri maintains its distinctiveness while sharing typological features with neighboring tongues.

In what ways do Bagri's pluralization strategies align with or deviate from broader Indo-Aryan morphological norms?

This question addresses the typological implications of the data. It aims to assess how Bagri fits into the wider Indo-Aryan family regarding plural formation and whether its reliance on phonological marking is unique, marginal, or reflective of a larger regional trend.

What can the pluralization system of Bagri tell us about the relationship between phonology and morphology in contact languages?

This theoretical question investigates the morphophonemic interface. It considers how grammatical distinctions typically marked by morphology are instead realized phonologically in Bagri and what this reveals about language change, efficiency, and the influence of sociolinguistic context.

These research questions aim to uncover the nature, distribution, and motivation of phonologically driven plural formation in Bagri, thereby offering insights into not only the structure of the dialect itself but also broader linguistic phenomena such as areal diffusion, typological variation, and the interplay between sound and grammar.

2.3 Research Methodology and Theoretical Framework

This study employs a qualitative, descriptive research methodology situated within the broader field of phonological analysis and language documentation. The primary objective is to investigate plural formation in Bagri using a data-driven approach that integrates both synchronic and diachronic perspectives.

2.3.1 Fieldwork and Data Collection

Data for this study were collected through fieldwork conducted across the Bagar region, which spans areas of Rajasthan, Haryana, and Punjab where Bagri is actively spoken. Native speaker consultants were selected to ensure a representative sample of dialectal variation. Participants included speakers from different age groups, genders, and sociolinguistic backgrounds to capture a wide range of phonological behavior. Data collection involved both elicited wordlists and naturally occurring speech recordings. Elicitation tasks focused on minimal pairs, noun paradigms, and plural forms, while spontaneous conversation samples provided insight into real-life usage of pluralization strategies.

2.3.2 Sampling Criteria

Purposive sampling was used to select informants with high linguistic competence in Bagri. Special attention was paid to speakers from rural areas who are less influenced by Hindi-dominated media, as these speakers are more likely to preserve traditional phonological features. Both male and female participants were included to balance sociophonetic variation.

2.3.3 Analytical Framework

The study is grounded in the theoretical tradition of descriptive and comparative phonology, with a focus on morphophonemic processes. Analytical attention was given to identifying consistent patterns of vowel alternation, consonant modification (fortition and lenition), epenthesis, deletion, and tonal variation. These patterns were compared both within Bagri and across neighboring languages such as Punjabi, Haryanvi, and Rajasthani to explore potential areal features.

The diachronic aspect of the study situates Bagri phonology in the context of historical sound changes that have occurred in the broader Indo-Aryan language family. Processes such as tone development, nasalization, and syllable restructuring were evaluated in light of both inherited and contact-induced changes.

2.3.4 Theoretical Orientation

The research draws from areal linguistics and contact phonology frameworks to interpret the influence of neighboring dialects on Bagri's plural system. The notion of a South Asian

linguistic area (Sprachbund), as discussed by Emeneau (1956) and Masica (1991), provides the conceptual foundation for understanding cross-linguistic convergence in this region.

2.3.5 Significance

This methodology not only allows for a detailed description of Bagri's plural formation but also contributes to broader linguistic theory. The findings have implications for phonology-morphology interface theory, typology of number marking, and language change in contact zones. Furthermore, this research contributes to the documentation and preservation of a less-studied Indo-Aryan dialect that faces increasing pressure from dominant regional languages.

3. Literature Review

The phonological processes involved in plural formation have long been a subject of interest in the field of linguistics, particularly within the context of Indo-Aryan languages, which demonstrate a rich typological diversity in morphophonemic patterns. In most standard Indo-Aryan languages such as Hindi, plural marking is achieved through overt suffixation, typically morphological in nature, where morphemes like -ẽ (for feminine nouns) and -e (for masculine nouns) are used. However, this conventional model does not fully account for the phonologically driven plural strategies found in non-standard dialects such as Bagri.

The foundational works of Grierson (1908) in the Linguistic Survey of India provided the earliest systematic accounts of dialectal diversity in India, including limited data on Bagri. Although descriptive, Grierson's work primarily offered a macro-level classification of language families and dialect zones without a detailed analysis of morphophonemic behavior. Subsequent descriptive grammars, notably those by Gusain (1999, 2000), have provided more specific linguistic documentation of Bagri, including its phonological inventory, syllable structures, and general morphological patterns. While these studies briefly touch upon plural formation, they do not offer an in-depth analysis of the diverse phonological strategies Bagri employs for number marking.

Areal linguistics offers a broader lens through which Bagri's pluralization system can be understood. Emeneau's (1956) seminal article "India as a Linguistic Area" posits that languages in close geographical proximity often develop shared linguistic features regardless of genetic affiliation. Bagri, spoken in the Bagar region overlapping Rajasthan, Haryana, and Punjab, exists within a highly interactive multilingual environment. The influence of neighboring languages like Punjabi (a tonal Indo-Aryan language) and Haryanvi and Rajasthani varieties is critical in understanding the emergence of tone and epenthetic vowel use in Bagri plurals. Masica (1991) further elaborates on this view, showing how convergence phenomena such as retroflexion, echo-words, and phonological deletion are prevalent across North Indian languages due to prolonged contact and bilingualism.

Other studies have contributed theoretical models for the interaction between phonology and morphology. Shapiro and Rajagopalan (2003) discuss the phonology-morphology interface and how languages can encode grammatical categories like number without strict reliance on affixation. In similar fashion, Ohala (1993) explains how perceptual and articulatory motivations often drive sound change, which can result in plural formation

through purely phonological means such as vowel lengthening, consonantal aspiration, or tonal shift.

Despite these insights, the specific case of Bagri remains under-researched, especially in terms of how its plural system reflects both typological uniqueness and areal convergence. While some comparative studies exist for Hindi and Punjabi plural systems, few have focused on Bagri as a case study of morphophonemic pluralization in contact settings.

This literature review thus reveals a critical gap in detailed, phonologically grounded studies of Bagri's pluralization system. The present research seeks to fill this gap by offering a comprehensive analysis of plural formation in Bagri through a phonological and areal linguistic lens, contributing to the broader discourse on language contact, morphophonology, and dialectology in South Asia.

4. Syllable Structure and Phonotactic Constraints

Bagri typically follows a Consonant Vowel structure, but pluralization often modifies syllable weight and structure, leading to:

4.1 Vowel lengthening

In Bagri, vowel lengthening occurs to maintain a phonological balance and is helpful in maintaining rhythm which helps in distinguishing singulars and plurals. Epenthesis (vowel insertion) is also observed which can break up difficult consonant clusters, making plural forms easier to pronounce.

The table below shows few examples along with the syllabic structures

Structure	Word (sg.)	Meaning	Structure	Word (pl.)	Meaning
V:CCV	/a:nðʰo/	blind	V:CCV:	/a:nðʰa:/	blinds
CV:CCV	/pa:ʈjo/	torn	CV:CCV:	/pa:ʈja:/	torn(pl.)
CV:C	/me:dʒ/	table	CV:CV:	/me:dʒã:/	tables
CV:CV	/gʰa:bo/	cloth	CV:CV:	/gʰa:ba:/	clothes
V:C	/ã:kʰ/	eye	V:CV	/ã:kʰã/	eyes

Table. 1

4.2 Tonal variations

In plural formations, tonal variations involve pitch modifications and changes that distinguish singular nouns from their plurals. Bagri often uses tonal changes, unlike other Indo-Aryan languages that follow the suffixation process, either independently or in co-occurrence with other phonological processes.

The tonal system of Bagri language is likely to be developed through sound changes that have occurred historically, including loss of certain consonantal sounds. One more reason for the tonal development can be the influence of neighbouring languages like Punjabi, which itself is a tonal language and has developed a 3 tone -system over the time.

4.2.1 Examples of Tonal Variations in Plural Formation

Few examples, that has been observed in the data collected from the native speakers are:

High-falling tone:

Singular	Plural	meaning	tone
gʰoɽa	gʰoɽe	horse/es	normal tone changing to high falling tone
tʃʰoro	tʃʰora	boy/s	level tone changing to high falling tone

Table. 2

One thing that has been observed is that the tones are only changing at the vowel place and are rarely impacting the consonant place.

Low-rising tone:

Singular	Plural	Meaning	Tone
baɪ	baɪjã	sister/s	mid tone changing to low rising tone
ʈab:ər	ʈab:ər	kid/s	low rising tone changing to mid tone

Table. 3

These tonal distinctions often occur in coordination with other phonological changes such as vowel quality shifts or consonant alterations. There are several over tones that have been observed like falling tone shifting to no tone, no tone shifting to low high tone. Adding to the observation, no higher tone has been shifted to a lower tone to make a plural marker. This creates a layered system of plural marking where multiple phonological cues work together.

4.3 Areal Influence on Tonal Plural Formation

The development of tonal plural marking in Bagri shows linguistic convergence. Neighboring languages like Punjabi and Western Rajasthani varieties have influenced Bagri's phonological system through centuries of contact, shared sound changes across the northwestern Indo-Aryan dialect which still continues. This represents what linguists call an "areal feature" - a linguistic characteristic that spreads across language boundaries within a geographic region.

Bagri's use of tonal variations for grammatical distinctions highlights the unique phonological character of Indo-Aryan languages in northwestern India. This feature demonstrates how languages can develop innovative grammatical marking strategies beyond the typical affixation patterns.

The presence of tonal plural marking in Bagri also supports the broader conclusion that languages in India develop through constant interaction and not in isolation. These tonal features represent both inherited Indo-Aryan characteristics and innovations emerging from language contact situations. Understanding these tonal patterns in Bagri contributes to our knowledge of how languages adapt and change within multilingual environments. The tonal plural system shows the complex interplay between inherited grammatical structures and areal innovations that characterize India's rich linguistic landscape.

5. Epenthesis (Vowel Insertion)

The phonological process of inserting vowels into specific environments plays a significant role in the plural formation system of Bagri. This process represents one of the several phonological strategies that Bagri employs to mark plurality, alongside consonantal changes, vowel lengthening, and tonal variations. Epenthetic vowels serve both phonological and morphological functions in Bagri, helping to create distinctive plural forms while simultaneously resolving potentially complex consonant clusters.

5.1 Patterns of Epenthesis in Bagri Plural Formation

In Bagri, epenthesis typically follows several systematic patterns when forming plurals:

5.1.1 Word-Final Epenthesis

When singular nouns end in consonant clusters, Bagri often inserts vowels to break up these clusters in plural forms. This epenthesis creates a more syllabically optimal structure while simultaneously marking the plural:

Singular	Plural	Meaning
g^{h}ar	g^{h}arã	house/s

Table. 4

Here, the epenthetic (nazalized) vowel /ã/ is added to the end of the word, creating an additional syllable that serves as a plural marker while also providing a simpler syllable structure.

5.2 Medial Epenthesis

Some plural forms in Bagri involve the insertion of vowels between consonants within the word:

Singular	Plural	Meaning
bakrı	bakarı	goat/s (f)

Table. 5

In this pattern, the consonant cluster /kr/ in the singular form is broken up by the insertion of an epenthetic vowel /a/ in the plural form. This medial epenthesis both marks plurality and creates a more open syllable structure.

5.3 Epenthesis with Morphological Boundaries

When certain plural suffixes would create difficult consonant sequences at morpheme boundaries, epenthetic vowels may be inserted:

Singular	Plural	Meaning
k^het	k^hetã	field/s

Table. 6

If the plural suffix /ã/ would create a difficult consonant cluster when attached directly to the stem, an epenthetic vowel might be inserted to facilitate pronunciation while preserving the morphological distinction.

5.4 Phonological Conditioning of Epenthesis

Epenthesis in pluralization of Bagri does not occur in arbitrary fashion, as its rules are determined by definite phonological conditions:

Consonant Sequence Constraints: Certain consonant combinations that have the potential of occurring in singular forms may be less favorable in plural forms, thus inciting epenthesis.

Syllable Structure Preferences: Bagri tend to prefer CV syllable structure in their plural forms, where vowels are inserted such that the preferred structure is obtained.

Vowel Harmony: The quality of epenthetic vowels is often influenced by the adjoining vowels, and Bagri presents characteristics of vowel harmony, whereby the epenthetic vowels agree with or harmonize with the stem vowels.

Stress Patterns: The place of stress in a word can influence the place of epenthesis, wherein vowels are inserted to maintain the preferred stress patterns in plural forms.

The patterns of epenthesis in Bagri plural formation reflect a range of areal characteristics within the Northwestern Indo-Aryan language group. The same epenthetic processes are observed in some of the neighboring languages, mentioning Punjabi, Haryanvi, and Rajasthani varieties, which means these features must have developed as a result of prolonged contact between the languages.

In this context, especially, the influence of shared epenthetic vowel quality and the environments that trigger epenthesis are evident. Shared patterns across the linguistic area tell us about how languages in geographical proximity tend to develop similar phonological strategies for marking grammatical distinctions such as plurality.

5.5 Functional Load of Epenthesis in the Plural System

Epenthesis serves multiple functions within Bagri's plural marking system:

Grammatical Marking: The insertion of vowels serves as more obvious overt markers of plurality, which distinguishes between singular and plural forms.

Phonological Optimization: Epenthesis resolves potentially difficult consonant sequences to render more optimal syllable structures.

System Integration: Epenthesis interacts with other phonological processes such as consonant voicing, vowel lengthening, and tonal variations to create a comprehensive system of plural marking.

These epenthetic patterns likely developed from earlier morphological processes that have been phonologized by the time. It could be said, plural markers that started off distinct have come to be represented by vowel insertion. They represent a sophisticated grammatical mechanism that efficiently encodes plural information through minimal additions of phonological material. This economy of expression showcases the natural evolution of language toward systems that balance clarity of communication with articulatory efficiency.

6. Consonantal Changes: Fortition and Lenition

Consonantal changes are some of the most important kinds of phonological processes in Bagri plural formation, with both strengthening (fortition) and weakening (lenition) processes noted to occur at different phonological environments. These processes are not only pivotal in the morpho-phonological system of the language in marking grammatical distinctions but also to assist articulation, achieving perceptual salience in diverse speech contexts.

6.1 Fortition in Plural Formation

Fortition is the process in which plurals are formed by strengthening consonants, typically occurring at morpheme boundaries in order to contrast the singular and plural adequately. This phonological process manifests through several distinct mechanisms in Bagri:

6.1.1 Aspiration of Unaspirated Stops

An interesting feature of plurals is that, very often in Bagri, aspiration is added to unaspirated stops, especially if found in word-initial and stressed syllable positions. Through this process, the plural form avails superiority in acoustic profile:

Singular	Plural	Meaning
/k:pas/	/k:p^{h}as/	cotton (s/pl.)
/toto/	/tota/	parrot/s

Table. 7

The aspiration serves to enhance the perceptual distinction of the plural form, while the place and manner characteristics of the original consonant remains intact.

6.1.2 Gemination

Gemination, or consonant doubling, is one of the other strategies in Bagri plural formation wherein it typically occurs in intervocalic positions:

Singular	Plural	Meaning
/kɪtab/	/kɪtabbã/	book/s
/moti/	/motti/	pearl/s

Table. 8

This gemination process would have the effect of elongation and increased acoustic salient articulation of the consonants reinforcing the morphological boundary between the stem and the plural (with both their degree and length of duration and formant) values being increased.

6.2 Lenition in Plural Formation

Lenition is a weakening of consonants in plural forms, mostly intervocalically or in unstressed syllables. This reflects natural articulatory tendencies toward simplification, especially in connected speech.

6.2.1 Voicing of Voiceless Consonants

One of the very common lenition processes in Bagri pluralization is the voicing of voiceless consonants mainly in intervocalic environments:

Singular	Plural	Meaning
/ra:f/	/ra:p^{h}/	outer side/s of mouth
/ək/	/ək^{h}ja/	fed up with thing/s

Table. 9

The mechanism of voicing is due to the tendency to articulate with less effort, juxtaposed between vowels, resulting in partial or complete vocal fold vibration during consonant production.

6.2.2 Degemination

Degemination, or the simplification of geminate consonants, happens if the final attachment of the plural suffix formation formalizes certain consonant sequences that are phonologically complex:

Singular	Plural	Meaning
/bnna:/	/bna:/	groom/s
/reɪtto/	/reɪt/	sand

Table. 10

This entire process of simplification reduces articulatory complexity, especially in rapid speech contexts, where maintaining geminate articulations may become challenging.

6.3 Phonological Conditioning Factors

The distribution of fortition and lenition processes in Bagri plural formation is not random but rather determined by certain phonological conditioning factors:

Syllable position: There is no fixed position in which fortition or lenition can take place. Nevertheless, it has been noted that in comparison to other positions, there are relatively fewer instances of either process at the beginning position of words.

Stress patterns: Stressed syllables tend to resist lenition and they may cause fortition, whereas unstressed syllables may facilitate the weakening process.

Adjacent segments: Adaptive behavior of surrounding segments often influences consonantal modifications, with a process of change possibly resulting in fortition or lenition, depending on the effects.

Speech rate: In rapid speech, lenition processes become more frequent since precision of articulation decreases.

The dynamic interplay between fortition and lenition processes in plural formation in Bagri reveals a fine balance between two opposing demands: perceptual distinctiveness and articulatory efficiency. The consonantal changes reflect phonologically grammatical distinctions, which differ from the morphological instances of grammatical marking by suffixation, and contribute to a rich phonological typology for the language.

7. Deletion (Elision of Sounds)

Deletion refers to the omission of phonemes in plural forms in order to obtain articulatory convenience, often in case of a phonotactic violation occurring as a result of the use of a plural suffix. It is an important phonological mechanism within Bagri for maintaining articulatory efficiency while balancing the requirement for grammatical marking of the number.

7.1 Patterns of Deletion

In Bagri pluralization, deletion commonly affects:

Redundant phonological material when plural suffixes are added: In certain phonological environments, elements that would create redundancy or aspiration are systematically eliminated, streamlining the resulting plural form while preserving its grammatical distinctiveness.

Singular	Plural	Meaning
/sa:d^{h}u/	/sa:d^{h}uã/õ/	priest

Table. 11

Word-final segments that become unstable when plural morphology is added: Terminal phonemes of the singular form, particularly vowels and certain consonants like /h/ and glides, may be vulnerable to deletion when plural suffixes are attached, especially if they would create marked structures.

Singular	Plural	Meaning
/ra:h/	/rã:/	way

Table. 12

Zero Marking: In some cases, it has been observed that the pluralization is happening but there is no visible grammatical or phonological change/ transition is happening. This is known as zero marking, and is particularly common with certain semantic classes of nouns.

7.2 Functional Analysis

Deletion in Bagri plural formation serves multiple functions:

Maintenance of preferred syllable structures and rhythmic patterns: Bagri exhibits preferences for certain syllable types (predominantly CV) and rhythmic modifications. Deletion helps preserve these preferred patterns by eliminating elements that would disrupt the language's natural prosodic contours.

Conservation of articulatory effort: From a functionalist perspective, deletion represents an economic measure that reduces the overall phonological substance while maintaining sufficient distinctiveness for communication purposes.

Preservation of morphological transparency: In some cases, deletion targets elements that might obscure the morphological boundary between stem and suffix, therefore maintaining the analytical clarity of the plural form irrespective of phonological modification.

These deletion processes demonstrate how Bagri balances between the demand for grammatical marking and phonological well-formedness constraints; in instances when phonological simplicity was prioritized over explicit morphological marking, these represent other dimension principles of language economy and competition between grammatical and phonological demands.

7.3 Phonological Conditioning and Constraints

It is thus clear that deletion processes in Bagri pluralization are not arbitrary but are systematically conditioned by various factors:

Phonological environment: The probability of deletion and the influences that determine its implementation are also highly dependent on segments surrounding the segment in question since some contexts favor elision than others.

Morphological structure: the kinds of plural suffix involved show further possible deletion scales, some requiring more extensive phonological modifications than others.

Lexical factors: Some lexical categories or semantic classes show distinctive patterns of deletion, indicating a high degree of interaction between phonology and lexical classification in Bagri grammar.

Register and speech rate: The extent of the deletion is usually correlated with speech rate and register markings, with casual and quick speech showing more substantial elision.

Dialect variation: Different regional varieties of Bagri reveal that the criteria for deletion can vary considerably, which reflects the language's sociolinguistic diversity in its geographical spread.

8. Metathesis (Reordering of Sounds)

Metathesis involves the reordering of sounds in plural forms, usually either towards attaining a better phonological configuration or ease in alternate pronunciation. Also, an elaborate phonological process maintaining the phonemic grouping, but using a varied arrangement of the members to gain better articulatory efficiency or perceptual distinctiveness.

8.1 Patterns of Metathesis

In Bagri pluralization, metathesis commonly occurs under the following conditions:

To resolve complex consonant clusters that arise at morpheme boundaries: When the mix of the stem and plural suffix produces difficult sequences of consonants to articulate or those which are perceptually vague, metathesis may rearrange these segments into more optimal configurations.

To create more preferred syllable structures, particularly CV patterns: Like many languages, Bagri opts for open syllables with single-onset consonants (CV); metathesis may convert less optimally structured syllables into these preferred structures.

To facilitate articulation in rapid speech contexts: Connected speech may induce metathesis when non forming material becomes modulated through speakers' efforts to afford ease of articulation and fluency.

Examples of metathesis in Bagri pluralization where the nasalization from one vowel shift to the word final when pluralized:

Singular	Plural	Meaning
/bʰæ̃s/	/bʰæsã/	buffalo/s

Table. 13

9. Reduplication

Reduplication involves the repetition of all or part of a word to indicate plurality, representing the morphophonology that accompanies suffixation in Bagri's pluralization system. This iconic process produces a direct phonological reflection of semantic plurality, setting a natural mapping between form and meaning.

9.1 Types of Reduplication

In Bagri pluralization, reduplication manifests in several forms:

Full reduplication, complete repetition of the word. The pattern leads to full duplication of the lexical item, usually with a linking element placed between the two repeated forms. Full reduplication generally indicates distributive or exhaustive plurality.

Singular	Plural	meaning
/ma:ɳəs/	/ma:ɳəs/-/ma:ɳəs/	man/ men

Table. 14

Partial reduplication, where only a portion of the word is repeated: This may involve the repetition of a specific syllable (often the initial syllable) or a prosodically defined unit such as a foot. Partial reduplication frequently carries more grammaticalized plural meanings.

Singular	Plural	Meaning
/ka:m/	/ka:m/-/ka:dʒ/	work/s

Table. 15

Echo reduplication, where the repeated element undergoes systematic phonological modification: In this pattern, the second occurrence of the reduplicated form exhibits

predictable sound changes, typically affecting the initial consonant or vowel. It is often beyond simple plurality and extends to representing variety or comprehensiveness.

Singular	Plural	Meaning
/roti/	/voti/	chapati/s

Table. 16

Ablaut reduplication, where vowel alternation between the base word and its reduplicate: It includes systematic vowel movements that change between the original form and its reduplicant, follows a peculiar pattern and creates rhythmic alternations that indicate plurality.

Singular	Plural	Meaning
/ʧɪʤ/	/ʧʊʤ/	object/s

Table. 17

By these examples, reduplication creates an iconic shape representing plurality through the act of repetition, reaffirming, via phonological mention, the semantic construing of multiplicity. The multiplicity underlined by various reduplication patterns allows the negation of different aspects of plurality, from different shades of mere numerosity via distributional, variousness, and completeness.

10. Conclusion

The phonological processes involved in the formation of plurals are indicative of a complex and elaborate system that reveals Bagri's deviation from standard Hindi and other Indo-Aryan languages. As shown throughout this analysis, Bagri has a less developed habit of using affixation and gives preference to phonological mechanisms varying from vowel lengthening to consonantal alterations, nasalization, and tonal alterations, to signal plurality. The Bagri dialect extends across the Bagar region traversing Haryana, Rajasthan, and Punjab, and exhibits its unique phonological strategies to pluralize nouns. These characteristics are derived from Indo-Aryan characteristics and close contact with neighboring languages, such as the Haryanvi, Punjabi, and Rajasthani varieties. The phonological processes presented in this study reveal how Bagri pluralization functions within an overall Indo-Aryan linguistic perspective.

One of the prominent principles of forming plurals in Bagri is vowel lengthening. Some form of phonological checks and balances as well as elements of rhythm in the form of Bagri allow pluralization to distinguish between singular and plural. The syllabic structures undergo systematic modifications, as demonstrated in examples like /a:nðho/ (blind) becoming /a:nðha:/ (blinds) and /pa:ʈjo/ (torn) transforming to /pa:ʈja:/ (torn-plural). These exemplify Bagri rules regarding the tendencies toward certain types of syllables and rhythmic patterns.

In contrast with any other Indo-Aryan languages that primarily depend on suffixation, Bagri makes use of tonal alternations, either separately or in combination with other phonological processes. These tonal differentiations might have arisen and developed along the lines of historical sound change, being influenced by neighboring tonal languages like Punjabi. The patterns observed are the result of most normal tones shifting to high-falling tones, such as

in the forms of /gʰoɽa/ to /gʰoɽe/ (horse/horses), showcase Bagri's innovative grammatical marking strategies.

Epenthesis here assumes both a phonological and morphological role in pluralization. The insertion of vowels in certain environments creates particular plural forms in Bagri while ameliorating potentially complex consonant clusters. This process manifests itself by way of word-final epenthesis (/ghar/ to /gharã/ - house/houses), medial epenthesis, and epenthesis at morphological boundaries. It is phonologically conditioned and reflects wider areal features of northwestern Indo-Aryan languages.

Both fortition and lenition significantly contribute to the plural marking system of Bagri. While the fortition pattern has processes like aspiration of unaspirated stops (/k:pas/ to /k:pʰas/ - cotton) and gemination, lenition works via phonetic voicing of voiceless consonants, plus degemination. The phonetic changes reveal a compromise between perceptual distinctiveness and articulatory efficacy.

Some additional phonological processes include deletion (elision of sounds), metathesis (the order of sounds switched), and all varieties of reduplication. Deletion ensures preferable syllable structures are maintained and less articulatory effort needed, while complicated consonant clusters are cleared by metathesis. Reduplication-whether full, partial, echo, or ablaut-creates an iconic manifestation of plurality through repetition.

Bagri's various pluralization strategies are a prime example of how languages in India evolve through constant interaction with other languages rather than in isolation. The phonological processes covered in this study illustrate the complex interrelation of inherited grammatical structures and areal innovations that shape the typologically rich linguistic landscape in India. By demonstrating the pluralization strategies of Bagri amid the increasing pressures of regional dialects against standard languages, we can gain a deeper insight into the typological diversity of morphophonological processes and contact-induced shifts.

List of Abbreviations, Maps and Tables

1. Abbreviations

Abbreviation	**Full Form**
IPA	International Phonetic Alphabet
CV	Consonant-Vowel (syllable structure)
sg.	Singular
pl.	Plural
M	Masculine
F	Feminine
SOV	Subject-Object-Verb

V	Vowel
C	Consonant

2. Map

Map 1: Geographical Distribution of Bagri Language Shows Bagri-speaking districts in Rajasthan, Haryana, Punjab (India), and parts of Pakistan (Bahawalpur & Bahawalnagar).

3. List of Tables

Table No.	**Title**
Table 1	Examples of Vowel Lengthening in Bagri Plural Formation
Table 2	Tonal Variations in Plural Formation: High-Falling Tone
Table 3	Tonal Variations in Plural Formation: Low-Rising Tone
Table 4	Word-Final Epenthesis in Singular and Plural Nouns
Table 5	Medial Epenthesis in Plural Formation
Table 6	Epenthesis at Morphological Boundaries
Table 7	Aspiration of Unaspirated Stops in Plural Forms
Table 8	Gemination Patterns in Plural Formation
Table 9	Voicing of Voiceless Consonants
Table 10	Degemination in Plural Forms
Table 11	Deletion of Redundant Sounds during Pluralization
Table 12	Deletion of Word-Final Segments
Table 13	Metathesis in Singular–Plural Word Pairs
Table 14	Full Reduplication in Bagri Plurals

Table 15	Partial Reduplication Examples
Table 16	Echo Reduplication Patterns
Table 17	Ablaut Reduplication in Plural Formation

References

Grierson, George Abraham. 1908. *Linguistic Survey of India, Volume IX, Part II.* New Delhi: Motilal Banarasi das.

Gusain, Lakhan. 1994. *Reflexives in Bagri*. Unpublished M. Phil. dissertation, Jawaharlal Nehru University, New Delhi.

Gusain, Lakhan. 1999. A Descriptive Grammar of Bagri. Unpublished Ph.D. dissertation, Jawaharlal Nehru University, New Delhi.

Gusain, Lakhan. (2000). *Bagri*. Munich: Lincom Europa.

Gusain, Lakhan. 2000. Limitations of Literacy in Bagri. In Nicholas Ostler & Blair Rudes (Eds.), *Endangered Languages and Literacy*, Proceedings of the Fourth FEL Conference. University of North Carolina, Charlotte.

Masica, Colin P. 1991. *The Indo-Aryan Languages.* Cambridge: Cambridge University Press.

Shapiro, Michael C. 2003. South Asia Area Linguistics. In Annual Review of Anthropology, 32, 59–74.

Emeneau, Murray B. 1956. India as a Linguistic Area. *Language*, 32(1), 3–16.

Ohala, John J. 1993. The Phonetics of Sound Change. In *Historical Linguistics: Problems and Perspectives*, ed. C. Jones, 237–278. London: Longman.

Singh, Rajendra & Sharma, Kanavillil Rajagopalan. 2003. *The Phonology-Morphology Interface.* London: Routledge.

Reduplication and Distributed Meaning in Indian Languages

Aadithya N Kopparapu &
Aparna Sankar
EFLU, Hyderabad
aadikopparapu@gmail.com
aparnasiyer17@gmail.com

Abstract

Reduplication is a widespread and semantically rich phenomenon in Indian languages, often giving rise to meanings such as distribution, iteration, and intensification. While individual studies have documented these effects in specific languages, there remains a lack of a unified, cross-linguistic semantic framework that can account for the interaction between reduplication and grammatical category. This paper proposes a new compositional model—the R-D model—to analyse reduplication in Indian languages through the interaction of three components: R (Reduplicator), a morphosyntactic trigger that licenses reduplication; I (Iteration), a semantic operation that introduces multiplicity of events or referents; and D (Distribution), a semantic operation that spreads these instances across space or time.

Crucially, the availability of I and D is constrained by the ontological features of the reduplicated base, which are defined in terms of the binary features [±time] and [±space]. Nouns ([+time, +space]) allow both iteration and distribution, verbs ([+time, –space]) allow iteration only, prepositions ([–time, +space]) allow distribution only, and modifiers ([–time, –space]) block both, resulting in mere intensification. Data was elicited directly from native speakers across Indo-Aryan, Dravidian, Munda, and Tibeto-Burman languages using translation tasks, context-based elicitation, and interpretation checks.

The R-D model accounts for observed patterns of reduplication across categories and languages while explaining the absence or blocking of certain meanings. It also predicts inter-language variation based on the distribution of ontological features and category-specific constraints. This framework not only captures the semantic diversity of reduplication in Indian languages but also offers a formal, compositional account of how grammatical category and semantic interpretation interact. The study contributes to the typology of reduplication and opens new directions for modelling semantic operators in morphology.

Keywords: Reduplication, Distribution, Iteration, Intensification, R-D model

1. Introduction

Reduplication is a widely attested linguistic process across the world's languages, serving a range of functions from morphological derivation to pragmatic emphasis. In South Asian languages, reduplication is not only pervasive but also semantically rich, yielding effects such as intensification, distribution, superordination etc. These effects are often sensitive to the syntactic category of the reduplicated element: when nouns are reduplicated, the resulting interpretation typically involves referential distribution; reduplication of verbs often leads to iterative or extended duration readings; and adjectives or adverbs yield intensification etc. Despite this well-documented diversity, there has been little work offering a unified semantic account of how reduplication systematically interacts with grammatical category and meaning in Indian languages.

This paper seeks to fill that gap by proposing a new compositional model of reduplication-induced meaning: the R-D model, which stands for Reduplicator-Distributor. In this account, reduplication is treated not as a single semantic operation but as a triad of operators that interact with linguistic units based on their syntactic category.

This model is inspired by previous proposals such as Balusu (2006), who analysed distributive readings of Telugu numeral reduplication using a spatial/temporal operator. While Balusu's work insightfully captures the distributive meaning of reduplicated numerals, it does not extend to other categories or other semantic effects such as iteration. The R-D model expands on this idea by proposing a general framework that can account for a broader range of categories, languages, and meanings.

The empirical focus of the paper is on reduplication in Indian languages across major language families, including Indo-Aryan (e.g., Hindi, Marathi), Dravidian (e.g., Telugu, Malayalam), Munda (e.g., Ho, Santali), and Tibeto-Burman (e.g., Ao, Ladakhi). These languages exhibit a remarkable variety of reduplication patterns, both in terms of form and meaning, yet they all show systematic interactions between category and interpretation. Our data shows that across languages, the same semantic operators are active—but their availability is determined by the ontological status of the reduplicated item. This makes a strong case for a compositional semantic model grounded in category-sensitive operator licensing.

In section 2, we begin by laying out the aims, objectives, and research questions of this study. In section 3, we survey the existing literature on reduplication in typology and South Asian languages, as well as prior formal semantic accounts. Section 4 introduces the R-D model in detail, including its formal structure and theoretical motivations. Section 5 presents the cross-linguistic data and applies the model to a range of reduplication constructions across languages and categories. Finally, section 6 offers conclusions and directions for further research.

2. Research goals and methodology

2.1 Aims and Objectives

The primary aim of this paper is to unify the explanation for the reduplicative meanings of iteration, intensification and, distribution by positing a set of operators named R-D.

The objectives of this study are:

- To formulate a semantic model (R-D) in which reduplicative meanings involve a set of operators and their interactions with ontological features of the lexical entries.
- To demonstrate how the distributive meaning of reduplication is based on the grammatical category and ontological properties (e.g., temporal vs spatial) of the reduplicated element and .
- To compare and contrast data from multiple Indian languages, spanning diverse language families and show that seemingly different semantic effects of reduplication across languages can be accounted for using the same operator-based framework.

By pursuing these goals, the study seeks to contribute not only to the semantic analysis of reduplication, but also to broader typological and areal discussions on how morphology interacts with meaning across languages.

2.2 Research Questions

The central questions guiding this study are as follows:

1. What are the basic semantic operations activated by reduplication in Indian languages?
2. How do these operations differ according to the grammatical category (e.g., noun, verb, adjective) of the reduplicated item?
3. Can a model for a unified explanation be proposed and if yes, can the model account for the full range of semantic interpretations associated with reduplication across different languages and categories?
4. What exactly is shared by languages in India as a Linguistic Area for this phenomenon?

These questions are designed to evaluate both the empirical coverage and the theoretical adequacy of the model.

2.3 Research Methodology

The empirical foundation of this study is built on originally elicited data from native speakers of several Indian languages, representing the Indo-Aryan, Dravidian, Munda, and Tibeto-Burman families. The goal of the data collection process was to systematically test how reduplication behaves across different syntactic categories—namely nominals, verbs, modifiers, and adpositions—and to determine which semantic effects (distribution, iteration, intensification) are licensed in each case.

Data was elicited through direct interaction with native speakers, using a semi-structured elicitation protocol designed to control for both form and context. The elicitation process involved the following steps:

1. Translation Tasks: Speakers were presented with sentences in English (or Hindi in some cases) and asked to translate them into their native language. These source sentences included contexts designed to test distributive, iterative, and emphatic meanings. For example, speakers were asked to translate sentences like "Every child went to their own house" or "He walked and walked until he got tired."
2. Contextual Prompting: In addition to translations, speakers were given specific contexts or situational descriptions and asked to form sentences that would naturally fit those contexts in their language. This allowed the observation of how reduplication emerges (or fails to emerge) spontaneously in discourse-relevant settings.
3. Category-wise Testing: These tasks were systematically repeated across four grammatical categories: nominals, verbs, modifiers, and adpositions. For each category, it was noted whether reduplication naturally occurred or whether the language lacked reduplicative constructions in that category. For example, if a language did not allow reduplication of adpositions, this was explicitly recorded.

After eliciting the sentences, a second round of inquiry was conducted to determine the interpretative range of the reduplicated forms. Speakers were provided with a set of possible meanings or interpretations commonly associated with reduplication (e.g., "Does this sentence mean the action happened again and again?", "Does this mean each person did it separately?", "Does this just mean 'very'?"). They were then asked whether each interpretation was available or acceptable for the sentence they had produced. This helped distinguish between what is grammatically acceptable and what is semantically accessible to native speakers, allowing for a more precise mapping of form to meaning.

The elicitation strategy thus combined structured translation with open-ended production and interpretation validation, ensuring that the data was both naturalistic and theoretically targeted. This approach also made it possible to track inter-speaker and inter-language variation, especially in cases where reduplication yielded ambiguous or non-compositional interpretations.

The resulting dataset forms the empirical basis for evaluating the R-D model developed in this study, allowing us to analyse which semantic operators are active or blocked in specific syntactic and ontological environments.

The theoretical framework presented in this study is based on the ontological features of the syntactic categories and how reduplication effects different categories. The analysis of the data collected is done to verify if the correlation between syntactic category and semantic effect is as predicted by the proposed theoretical framework.

3. Literature review

The following is the literature review of the papers that inspired this study.

In Abbi (1985) and Abbi (1990), Anvita Abbi shows how languages belonging to diverse language families in India, namely Indo-Aryan, Dravidian, Tibeto-Burman and Austroasiatic, show reduplicative phenomenon with shared interpretations. She claims that this similarity is due to areal influence in the South Asian Linguistic Area.

Balusu (2006) presents a detailed analysis of distributive reduplication in Telugu, arguing that reduplicated elements introduce event plurality. He utilises event semantics to explain the distributive meaning of Telugu numerals by positing a D-operator which has "spatial, temporal and participant key readings."

Expanding on this, Balusu and Jayaseelan (2013) compare Telugu data with Malayalam and Tamil, emphasizing that the phenomenon is pan-Dravidian but not uniform in form or function. They maintain that reduplicated quantifiers in these languages uniformly trigger a distributive interpretation, albeit with some morphosyntactic differences. One significant theoretical insight here is the claim that "reduplication serves as a strategy for quantificational strength in languages that otherwise lack overt distributive markers like 'each' or 'every.' ." This means that in languages which lack overt universal quantifiers, reduplication fills that role with the distributive meaning.

Together, these works draw a compelling picture: reduplication, far from being a mere stylistic or phonological quirk, is a rich grammatical strategy that various languages deploy to encode distributive semantics. Telugu and its Dravidian relatives offer particularly lucid cases, where reduplication interacts with syntactic scope, event structure, and quantificational force. These findings invite further inquiry into whether such constructions are universally available but parametrically varied in expression—or whether they reveal something deeper about how languages encode event individuation.

4. The r-d model: a compositional account of reduplication and meaning

4.1 Ontological Features of Syntactic Categories

The central innovation of the R-D model is the incorporation of ontological features that reflect the referential nature of lexical items. The interpretation of reduplication is

constrained by the ontological profile of the base, which can be captured via two binary features: [±time] and [±space]. These features are grounded in the referential domains that different syntactic categories typically invoke:

1. Verbs are expressions of eventualities, which refer to actions, events, states and, processes which unfold in time rather than being located in space. Therefore, they are marked as [+time, –space].
2. Prepositions encode spatial relationships without eventive reference, and are marked [–time, +space].
3. Nominals refer to entities that exist either in space or time or both space and time; thus, they are [+time, +space].
4. Modifiers such as adjectives, numerals and adverbs are non-referential in this ontological sense, lacking direct spatiotemporal anchoring, and are assigned [–time, –space].

These features underpin the hypothesis advanced in this study, which seeks to unify the effects of intensification, distribution, and iteration through an operator set (R-D) comprising a Reduplicator (R), an Iterator (I), and a Distributor (D). The operator set interacts with lexical items, and the interpretation is conditioned by their syntactic category and its interaction with the R-D set especially the D-operator.

4.2 The R-D Operator Set as an Areal Feature of India as a Linguistic Area

The main proposition of the study is that the R-D Operator set is a grouping of 3 distinct operators:

1. The Reduplicator (R) – The operator responsible for reduplication of the Phonological Form.
2. The Iterator (I) – The operator which iterates the instances of the lexical item.
3. The Distributor (D) – The operator which distributes the iterations of the lexical item through time or space depending on the ontological features of the lexical item.

The D-operator is the only one from the R-D set that is sensitive to the ontological features. It can distribute a lexical item only through the dimension of the feature(s) marked [+].

This means that the D-operator can distribute the instances of a lexical item L through time if and only if L has [+time] feature. This constraint allows derivation of the meanings based on syntactic categories as required. A sample example of how this works is depicted below:

Consider a word W with the features [-time, +space] and predicate function word(x). When R-D applies to this word, the following happens:

1. The R-operator reduplicates the word in its phonological form and gives the output W~W.
2. The I-operator iterates the instances of the predicate function of W and returns multiple predicate functions as a set {word(x), word(x), word(x) …}.

3. The D-operator then distributes the multiple instances generated by the I-operator through the spatial dimension because of the [+space] feature and returns a meaning of distribution through space. A meaning of distribution through time is blocked because W has [-time] feature.

The second main proposition of this study is that the R-D operator set is a shared feature in the Indian Linguistic Area. The operators themselves might be universal to all languages of the world. In India(and broadly South Asia), these operators are grouped together giving rise to the varied meanings of reduplication across syntactic categories which are uniform across languages in the region.

The model's application on real data taken from Indian languages will be explored in section 5.

5. Findings and analysis

Reduplication and its associated meanings of intensification, iteration and distribution are prevalent in the Indian subcontinent's various languages. However, there are nuances in the spread and depth of these features. The crosslinguistic data from different language families is analysed and these nuances are explored category-wise.

5.1 Nominals

Nominals are entities usually marked with [+time, +space] features. When R-D applies to them, the resultant meaning is distribution through the entities. In Balusu (2006), this reading is referred to as "the participant key reading." This study posits that the participant key reading is in fact the D-operator distributing through both time and space simultaneously. The entities within the universe of discourse would hence be the points of distribution as they are [+time, +space].

Common nouns, anaphors and pronouns are affected in the same way and give rise to distributive meanings.

Malayalam (Dravidian):
ellaːʊarum aʋaɾ-aʋaɾ-uɖe kaseːrajil iɾikʲkʲuka
everyone they~they-GEN chair-LOC sit
'Everyone sit in their own chairs'

Telugu (Dravidian):
saṉḍu-saṉḍu veṯakaɳḍi
street~street search
'Search every street'

Hindi (Indo-Aryan):
gʰəɾ- gʰəɾ kiː kəhaːnɪ hẽ
house-house of story is
'(It is) the story of every house'

Marathi (Indo-Aryan):
t̪umt͡ʃa:-t̪umt͡ʃa: kurt͡ʃivar basa:
your~your chair:LOC sit
'Sit in your own chairs'

Ho (Munda):
hora-hora te-lel hud͡ʒui-me
route~route PROG-see come-IMP
'Look at each and every route and come'

Bodo (Tibeto-Burman):
lama-lama tʰɯŋd͡ʒɯŋ tʰaŋ
road~road straight go
'Keep taking every straight road (do not turn)'

Ao (Tibeto-Burman):
pi:-pi: kidaŋi oaŋ
you~you house go
'Go to your own houses'

5.2 Verbs

Verbs have [+time, -space] features and are hence distributed through time by the D-operator. This results in the reduplicated verb perceived as being of an extended duration. An interesting point is that a reduplicated verb cannot be the main verb of the sentence, i.e. reduplication cannot happen to verbs in the matrix clause.

Indian languages differ in how reduplicated verbs are formed. The Indo-Aryan languages utilise *converb constructions* while the Dravidian languages utilise regular verb forms themselves. This leads to an interesting consequence: reduplicated verb forms in Dravidian languages can be used in their non-reduplicated versions while those in Indo-Aryan languages cannot be.

Malayalam (Dravidian):
n̪adan̪n̪ɯ-n̪adan̪n̪ɯ kʃi:ɳit͡ʃu
walk.PTCP~walk.PTCP became tired
'I got tired from walking.'

Marathi (Indo-Aryan):
lihun̪-lihun̪ ʈʰak le
write.PTCP~write.PTCP tired of
'(I) got tired of writing.'

Hindi (Indo-Aryan):
t͡ʃəlt̪e:-t͡ʃəlt̪e: mɛ:ne: ʊska: ha:t̪ pəkəɽa:
walk.PROG~walk.PROG I his hand held
'I held his hand while walking'

Ao (Tibeto-Burman):
ken aŋa-aŋa bilem-oko
song listen.PTCP~listen.PTCP got-tired
'(I) got tired of listening to the song'

Bodo (Tibeto-Burman):
aŋ gaɾi sɯlai-sɯlai fɯinai
I vehicle change~change to come
'I had to change vehicles many times to come'

Ho (Munda):
ɖuʔub-ɖuʔub-ʈe huʤuʔu-me
sit.PTCP~sit.PTCP come-IMP
'Sit and come' (implies sitting many times in different places)

Santali (Munda):
ɖuɽub-ɖuɽub-ʈiŋ laːŋga-enaiʔin
sit.PTCP~sit.PTCP bored-of
'Bored of sitting for a long time'

Telugu goes a little further and distinguishes the verbal reduplicated meanings based on the aspect of the verb. For progressive verbs, reduplication denotes only an extended duration of the verb, while for perfective verbs a second meaning of "multiple instances" is also available.

naɖiʧi-naɖiʧi alisipojaːnu
walk.PERF~walk.PERF got tired
(I) got tired of walking

naɖuʂʈu-naɖuʂʈu alisipojaːnu
walk.PROG~walk.PROG got tired
(I) got tired while walking

5.3 Adpositions:

In Indian languages, adpositions mostly follow the noun and are hence postpositions. They have the [-time, +space] features and are hence distributed through space by the D-operator.

Malayalam (Dravidian):
kuʈʈikaɭ aɖuʈʈu-aɖuʈʈu iɾuɳɳu
children near~near sat
'The children sat side by side'

Banjara (Indo-Aryan):
maːr paːt͡ʃ-paːt͡ʃ aːd͡ʒoː
me behind~behind come
'Come behind me' (keep following me)

Hindi (Indo-Aryan):
ʊske piːt͡ʃeː-piːt͡ʃeː jaːo
his behind~behind go
'Go behind him' (follow him)

Santali (Munda):
sur-sur-laŋ duɽub aː
near~near-both sit will
'Both will sit side-by-side'

Ho (Munda):
abu buru kuɽi-kuɽi-t̪e senuwa
we jungle side~side-from will go
'We will go through the side of the jungle' (to keep taking one particular side)

Based on the dataset and speaker sample available for this study, no instances of reduplication involving adpositions were attested in the Tibeto-Burman languages. One interesting finding was that some native speakers don't distribute the postposition and take an iterative meaning instead. In Telugu, for example, distributive reading of postpositions is absent for some native speakers. They instead perceive the reduplicated forms in an iterative way.

peŋkulanu pai-paina peʈʈu
Tiles up~up put
'Put the tiles on top of each other' (distributive reading)

peŋkulanu pai-paina peʈʈu
Tiles up~up put
'Cover (something else) with the tiles superficially' (pragmatic meaning derived from iterative reading)

5.4 Modifiers

Modifiers like adjectives, adverbs and numerals have [-time, -space] features. When R-D applies to them, the D-operator is unable to distribute the iterations generated by the I-operator. The study posits that this iteration without distribution is the reason for intensified meaning. The reason for this assumption of iteration is based on the meaning of the reduplicated form of the word for "again" in Indian languages.

Malayalam (Dravidian):
ɲa:n avano:ɖɯ ʋi:ɳɖum-ʋi:ɳɖum paraɲɲu
I him again~again told
'I told him again and again'

Marathi (Indo-Aryan):
mi puɳha:-puɳha: ʋit͡ʃa:ɾle
I again~again asked
'I asked again and again'
As can be seen in the above sentences, the intensification of "again" is iteration of the number of instances of "again". This can be extended to all other modifiers.

There is a difference in how adjectives and adverbs intensify as opposed to numerals. Numerals are quantifiable and hence intensification of numerals adds to the number of references.

Telugu (Dravidian):
pillalu mu:ɖu-mu:ɖu paɳɖlu konna:ru
kids three-three fruits bought
'The kids bought three fruits each'

Hindi (Indo-Aryan):
laɽkõ-ko pã:t͡ʃ-pã:t͡ʃ kɪʈa:bẽ mɪli:
boys five-five books got
'The boys got five books each'

For adjectives and adverbs, the intensification is qualitative and hence doesn't add to the number of references.

Malayalam (Dravidian):
ammu:mma paʈije:-paʈije: ɳaɖaɳɳu
grandmother slowly~slowly walked
'Grandmother walked very slowly'

Telugu (Dravidian):
ʋa:ɖu mella-mella-ga: ma:ʈla:ɖæ:ɖu
he slow~slow-ly speak:PAST:3P.SING.MASC
'He spoke very slowly'

Marathi (Indo-Aryan):
la:mb-la:mb rasʈa: a:he:
long~long way is
'It is a very long way'

Banjara (Indo-Aryan):
eɳa: moʈe:-moʈe: gʰar ʧa:
here big~big houses are
'There are very big houses here'

Ho (Munda):
suwe-suwe sene-me
slow~slow walk-IMP
'Walk very slowly'

Santali (Munda):
lage-lage taɽam-me
fast~fast walk-IMP
'Walk very fast'

Ladakhi (Tibeto-Burman):
kʰo gjokspa-gjokspa ʧenuk
he fast~fast go.PRES
'He is going very fast'

Bodo (Tibeto-Burman):
bejɯ kʰaileŋ-kʰaileŋ tʰabadɯŋg
He lean~lean walks
'He walks in a zig-zag way'

6. Conclusion

This study has introduced the R-D operator set positing a unified, operator-based account of how reduplication interacts with syntactic category and ontological features to yield distributive, iterative, and intensifying meanings in Indian languages. By positing three distinct operators—Reduplicator (R), Iterator (I), and Distributor (D)—and grounding their effects in binary time/space features of syntactic categories, the model successfully captures cross-linguistic patterns in Indo-Aryan, Dravidian, Munda, and Tibeto-Burman data. Empirical findings demonstrate that nouns systematically invoke spatial-temporal distribution, verbs yield extended or repeated events, adpositions distribute through space, and modifiers intensify via iteration without distribution. The R-D framework thus offers a compositional, typologically robust explanation for a wide array of reduplicative phenomena across South Asian languages.

The reliance of this study on elicited data calls for broader corpus-based validation, especially in under-documented Munda and Tibeto-Burman varieties. Future research could refine the categories (e.g. temporal entities like month), explore interactions with aspect and information-structure, and test the model's predictions in real-time processing and acquisition studies.

List of Abbreviations

R-D – Reduplicator-Distributor (Semantic Model)
R – Reduplicator (Operator for phonological reduplication)
I – Iterator (Operator for iterating instances)
D – Distributor (Operator for spatial/temporal distribution)
GEN – Genitive (case marker)
LOC – Locative (case marker)
IMP – Imperative (mood marker)
PTCP – Participle
PROG – Progressive aspect

L1 / L2 – First Language / Second Language
[+time] / [-time] – Temporal ontological feature (positive/negative)

NOM – Nominative (case marker)
1SG, 2SG, 3SG – First, second, third person singular
1PL, 2PL, 3PL – First, second, third person plural
AUX – Auxiliary verb
DSTR – Distributive
ADV – Adverb
V – Verb
NP – Noun Phrase
PP – Prepositional Phrase
SOV – Subject-Object-Verb (word order)

[+space] / [-space] – Spatial ontological feature (positive/negative)

References

Abbi, Anvita. 1990. Reduplication in Tibeto-Burman languages of South Asia. *Japanese Journal of Southeast Asian Studies*, 28(2), 171–181.

Abbi, Anvita. 1985. Reduplicative structures: A phenomenon of the South Asian linguistic area. *Oceanic Linguistics Special Publications*, 20, 159–171.

Balusu, Rahul. 2006. Distributive reduplication in Telugu. In *Proceedings of NELS 36*.

Balusu, Rahul and K. A. Jayaseelan. 2013. Distributive quantification by reduplication in Dravidian. In Lisa Matthewson (ed.), *Strategies of Quantification*, 60–86. Oxford: Oxford University Press.

वाक् मंथन
ISSN: 2426-2149

Phonological Processes in Bagri: A Descriptive Analysis of Reduplication

Madan Lal
&
Hari Madhab Ray
Jawaharlal Nehru University

Abstract

Bagri, a lesser-documented Indo-Aryan dialect spoken predominantly in parts of Rajasthan, Haryana, and Punjab, exhibits distinctive phonological processes contributing to its phonotactic structure and rhythm. This study investigates core phonological phenomena of Reduplication in Bagri. It explores how this process interacts to maintain phonological harmony, facilitate ease of articulation, and adhere to Bagri's phonotactic preferences. Using a descriptive linguistic approach of Auto segmental Phonology, this research captures the structural intricacies and underlying rules governing these processes in Bagri, thereby enhancing the typological understanding of Indo-Aryan languages.

1. Introduction

Bagri, a dialect belonging to the Western Rajasthani subgroup of Indo-Aryan languages, is spoken primarily in the northwestern regions of India, covering parts of Rajasthan, Haryana, and Punjab. Despite its wide geographic distribution and substantial number of speakers, Bagri remains underrepresented in mainstream linguistic research. Most existing studies on Indo-Aryan phonology tend to concentrate on Hindi, Urdu, Punjabi, and Bengali, thereby marginalizing dialects like Bagri, which exhibit unique structural features deserving of scholarly attention.

This article aims to describe the Reduplication that characterizes Bagri. By focusing on phenomena such as assimilation, vowel epenthesis, reduplication, and vowel shifting, this study seeks to provide a comprehensive account of Bagri phonology and highlight its divergence from and convergence with other Indo-Aryan languages.

2. Background of Research and Methodology

Bagri, a dialect belonging to the Western Rajasthani subgroup of Indo-Aryan languages, is spoken primarily in the northwestern regions of India, covering parts of Rajasthan, Haryana, and Punjab. Despite its wide geographic distribution and substantial number of speakers, Bagri remains underrepresented in mainstream linguistic research. Most existing studies on Indo-Aryan phonology tend to concentrate on Hindi, Urdu, Punjabi, and Bengali, thereby marginalizing dialects like Bagri, which exhibit unique structural features deserving of scholarly attention. This article aims to describe the reduplication that characterizes Bagri. By focusing on phenomena such as assimilation, vowel epenthesis, reduplication, and vowel shifting, this study seeks to provide a comprehensive account of Bagri phonology and highlight its divergence from and convergence with other Indo-Aryan languages.

This study adopts Goldsmith's auto segmental phonology (1976), which provides an analytical framework for understanding reduplication in Bagri as a multi-tiered phenomenon, rather than treating reduplication as a purely morphological operation. Auto segmental theory reveals how reduplication simultaneously manipulates independent phonological tiers, including segmental, tonal, and stress tiers. Goldsmith's auto segmental phonology posits that phonological representations consist of parallel, independently ordered tiers. These tiers are Skeletal Tier (CV-tier), segmental tiers, tonal tier, and stress or prosodic tier. These tiers represent templates of consonants (C) and vowels (V), as well as phonemes

and their features. Tone and stress are independent and associative, while metrical and prosodic structure are related. This analysis examines Bagri reduplication exclusively through the Goldsmith auto segmental lens, demonstrating how distinct reduplication types emerge from differential operations on the tier. To ensure comprehensive phonological coverage, the study uses selected lexical items from the standardized Swadesh list (core vocabulary resistant to borrowing). Moreover, there were terms unique to Bagri (e.g., agricultural tools, kinship terms, etc). The purpose behind using culture-specific terms is to capture the phonotactic patterns unique to Bagri, such as /tab≅r/ "child".

Data were collected from native Bagri speakers (N = 27) across four representative districts of Rajasthan (see Table). Participants were selected through stratified purposive sampling to ensure balanced demographic representation.

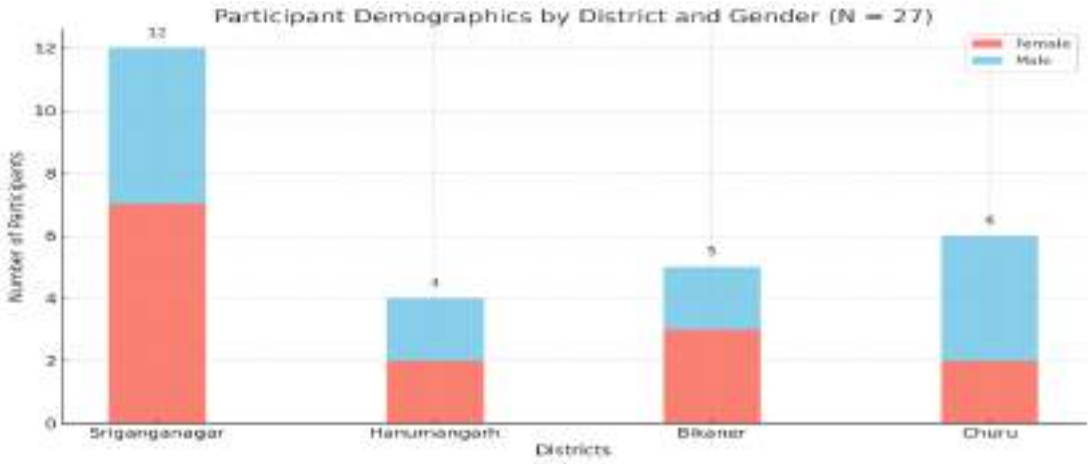


Figure 1: Demonstrating the data collected from various sources

3. Reduplication in Bagri

Reduplication in Bagri is often used to convey emphasis, plurality, or intensity, and it plays a significant morphological role in the dialect. Partial or complete repetition of the root word's syllables is applied based on semantic requirements. This is a common morphological phenomenon across many Indian languages, enriching Bagri's lexicon.

Base Form	Gloss	Reduplicated Form	Function
/na.no/	'small'	*na.no-na.no*	Emphasis (into small pieces)
/de. |o/	'give'	le. |o-de |o	Emphasis on giving

Table 1: Examples of Reduplication in Bagri

For instance, the reduplication of *"na.no" ("small") as "na.no-na.no"* intensifies the meaning, indicating a tiny size or quantity. This morphological strategy is critical in expressing varying degrees of meaning.

Complete reduplication

Bagri involves the full phonological copy of a root or word, typically used for semantic intensification or habitual aspect. From the auto-segmental perspective, this process represents a complete replication of the CV tier, segmental tier, tonal tier, and occasionally the prosodic tier, resulting in a structurally isomorphic second form.

Base Word	Reduplicated Form	Meaning	Segment change	Type
c&Hora	c&Hora-c&Hora	All boys	Same	Complete
mI□Ho	mI□Ho-mI□Ho	Kind of sweet	same	Complete
□ab↔r	□ab↔r-□ab↔r	All children	Same	Complete
admi	admi-admi	all men	same	Complete

kitab	Kitabã-kitabã	only books*	Same	Complete
c&Hori	c&Horijã-c&orrijã	all girls	Same	Complete
kar	karã-karã	all cars	Same	Complete

Table 2: Complete Reduplication (Total Reduplication)

In Feature Geometry, the second copy retains the complete [Root] structure, including [Laryngeal], [Supralaryngeal], [Place], and [Manner] nodes. There is no simplification or pruning of the tree. Instead, features like [±nasal], [±continuant], [±high], etc., are reinstantiated in the second item. This process illustrates a non-derivational duplication that is motivated semantically, rather than phonologically. Significantly, the prosodic structure may shift, often assigning secondary stress to the second form, revealing a suprasegmental effect even in identical segmental structures.

Base Form: c&Hora

Skeletal Tier: C V C V

| | | |

Segmental Tier: c&H o r a

Tonal Tier: (null for Bagri)

Stress Tier: σ (primary)

Reduplicated Form: c&Hora-c&Hora

Skeletal Tier: C V C V - C V C V

| | | |- | | | |

Segmental Tier: c&H o r a c&H o r a

Tonal Tier: (null)

Stress Tier: σ (primary) σ' (secondary)

Complete reduplication in Bagri (e.g., 'c&Hora-c&Hora' "all boys", 'mI□Ho-mI□Ho' "kind of sweet") represents the most transparent autosegmental mechanism: simultaneous copying of skeletal, segmental, tonal, and stress tiers.

The crucial observation is that the skeletal and segmental tiers are faithfully replicated, but the stress tier undergoes modification: the reduplicant receives secondary (or reduced) stress in many contexts. This demonstrates that Goldsmith's framework captures how identical segmental content can receive distinct prosodic prominence through independent tier manipulation. Autosegmentally, this reflects the principle of tier independence: the copying operation on the segmental tier does not automatically determine stress assignment on the stress tier. Instead, language-specific constraints govern stress distribution, allowing the reduplicant to be phonemically identical to the base yet prosodically distinct.

Echo reduplication (n+nonsense rhyming)

In autosegmental phonology, echo reduplication operates via floating consonantal nodes that override the initial segment on the second copy. While the CV-tier remains fixed (i.e., same number of syllables and syllable weight), only the onset features are detached and replaced.

Word	**Reduplicated form**	**Meaning**	**Segment Change**	**Type**
ro∝◊i	ro∝◊i-rY∝i	food etc.	o→Y	Echo
paNi	paNi-pYNi	liquids etc.	o→Y	Echo
gaba	gaba-gYba	clothes etc.	o→Y	Echo
pisa	pisa-pYsa	money etc.	o→Y	Echo
beli	beli-bYli	friends etc.	o→Y	Echo
kItab	kItab-kItYb	books etc.	o→Y	Echo
ga |a	ga |a-gY |a	songs etc.	o→Y	Echo

Table 3: Echo reduplication

Autosegmentally, this reflects the principle of tier independence: the copying operation on the segmental tier does not automatically determine stress assignment on the stress tier. Instead, language-specific constraints govern stress distribution, allowing the reduplicant to be phonemically identical to the base yet prosodically distinct.

Base: roti

Skeletal Tier: C V C V

| | | |

Segmental Tier: r o □ i

Reduplicated (Echo): rY∝i

Skeletal Tier: C V C V (IDENTICAL to base)

| | | |

Segmental Tier: r Y □ i (onset copied; vowel undergoes backening)

Echo reduplication (e.g., ro□i 'bread', kitab-kitYb 'books', pa |i-pY |i 'liquids') presents a more complex autosegmental operation. The skeletal tier remains identical, but segmental content undergoes substitution, particularly in the onset position. Critically, the CV-skeleton is entirely preserved, ensuring that the reduplicant maintains syllabic structure and weight. However, the segmental tier exhibits selective copying: the onset [r] is retained, but the vowel undergoes a feature change [o]→[Y] (backing and rounding). Echo reduplication demonstrates that the skeletal tier constrains phonological output independently of segmental content. The reduplicant's phonotactic well-formedness is determined by its skeletal structure, not its segmental realization. This reveals a fundamental principle of autosegmental theory: structure and content can be dissociated.

Numeral reduplication

Numeral reduplication (e.g., *ek-ek* 'one each', *do-do* 'two each') involves segmentally identical forms but with altered suprasegmental structure. Autosegmentally, this reflects the insertion of floating intonational boundary tones between the two reduplicants.

Word	Reduplicated form	Meaning	Segment Change	Type
ek	ek-ek	One each	Same	numeral
do	do-do	Two each	Same	Numeral
kitta	kitta-kitta	How much to each	same	Numeral

Table 4: Numeral Reduplication

Numeral reduplication (e.g., *ek-ek* 'one each', *do-do* 'two each') involves segmentally identical forms but with altered suprasegmental structure. Autosegmentally, this reflects the insertion of floating intonational boundary tones between the two reduplicants.

Base Numerals: ek-ek

Skeletal Tier: V C # V C

| | | | |

Segmental Tier: e k e k

Intonational Tier: L% (base) H% (boundary tone; phrase boundary)

According to Goldsmith's framework, the intonational tier encompasses floating tones that mark the boundaries of prosodic phrases. The insertion of a mid-high boundary tone (or pause) between the two identical numerals creates pragmatic distinctiveness without segmental change. This demonstrates that reduplication can be a fundamentally prosodic phenomenon, where the significant morphological operation occurs at the suprasegmental tier rather than at the segmental tier.

4. Conclusion

Goldsmith's auto segmental phonology elegantly captures Bagri's reduplication diversity through a unified principle: tier-selective copying. Complete, echo, intensification, and numeral reduplication each emerge as distinct manipulations of independent phonological tiers rather than as ad hoc morphological rules. The CV-skeleton, segmental tier, and suprasegmental tiers operate with their own associative logic, allowing the same root or word to be reduplicated in phonologically distinct yet functionally coherent ways.

Reduplication type	**Skeltel tier operation**	**Segmental tier operation**	**Suprasegmental tier operation**	**Tier feature**
Complete	Fully copied (CVCVlCVCV)	Fully copied	Distinct stress placement	Full tire replication
Echo	Copied with vowel change (o→Y)	Selective (onset)	Retained from base	Skeletal independence
numeral	Fully copied	Fully copied	Intonation boundary	Floating tone insertion

Table 5: Reduplication strategies in Bagri

This analysis demonstrates that Bagri reduplication is not morphological happenstance but rather a systematic phonological phenomenon governed by universal principles of multi-tiered organization. Goldsmith's auto segmental framework reveals that morphosemantic functions—distributivity, intensification, non-specificity, and pragmatic marking—are fundamentally encoded in phonological structure through tier-specific operations.

References:

Foley, W. A. (1997). Anthropological linguistics: An introduction. Blackwell.
Goldsmith, J. (1976). Auto-segmental phonology (Doctoral dissertation). MIT.
Gosain, L. (2002). Bagri bhasha ka vivaranatmak adhyayan. University of Rajasthan.
Grierson, G. A. (1916). Linguistic Survey of India, Vol. IX, Part I. Government of India.
Hyman, L. M. (2008). Universals in phonology. The Linguistic Review, 25(1-2), 83–137.
Jenkins, J. (2000). The phonology of English as an international language: New models, new norms, new goals. Oxford University Press.
Kachru, B. B. (1992). The Other Tongue: English Across Cultures (2nd ed.). University of Illinois Press.
Ladefoged, P., & Maddieson, I. (1996). The sounds of the world's languages. Blackwell.
Masica, C. P. (1991). The Indo-Aryan languages. Cambridge University Press.
Ohala, J. J. (1981). The listener is a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), Papers from the Parasession on language and behavior (pp. 178–203). Chicago Linguistic Society.
Shackle, C. (1976). The Siraiki language of central Pakistan: A reference grammar. School of Oriental and African Studies.
Shapiro, M. C. (2003). A Primer of Modern Standard Hindi. Motilal Banarsidass.
Singh, R. (1995). Reduplication in South Asian Languages. Linguistic Survey of India, Vol. VI.
Singh, R. K. (2001). Phonetics and phonology of the Bagri dialect. Jaipur University Press.

वाक् मंथन
ISSN: 2426-2149

Reduplication in Kokborok: Form-Function beyond Iconicity

Anirban Naha
Jawaharlal Nehru University

Abstract

Reduplication, one of the most productive morphological processes in Tibeto-Burman languages, provides crucial insights into the interface of phonology, morphology, and semantics. This paper investigates reduplication in Kokborok, the largest indigenous language of Tripura, by analyzing structural patterns and semantic functions across conversational, narrative, and elicited data. The analysis identifies multiple types of reduplication in Kokborok, including full, partial, echo, and expressive reduplication. These are linked to functions such as intensification, aspectual marking, distributive semantics, attenuation, and evaluative meaning. While iconicity often motivates reduplication cross-linguistically, Kokborok exhibits several instances of reduplication where Haiman's (1980) claim of more form is equal to more meaning is validated. The findings highlight the centrality of reduplication in Kokborok morphosyntax and contribute to broader typological discussions of morphological iconicity, and meaning extension.

Keywords: Reduplication, Kokborok, iconicity.

1. Introduction

Kokborok, a Tibeto-Burman language belonging to the Bodo-Garo group is spoken primarily in Unokoti district (earlier North Tripura) of Tripura, small scattered groups in Karimganj district of Assam in Northeast India and in Dhaka and Chittagong hill tracts of present day Bangladesh. Kokborok was declared as the official language of Tripura in 1979. It is the second official language of Tripura after Bengali. According to 2001 Census of India, the total population of Kokborok speakers in the State of Tripura was 7.62 lakhs. Despite being the mother tongue of nearly one million speakers (Benedict 1972; Bimal 2009), Kokborok remains under-documented compared to major South Asian languages. While early descriptive grammars (Darlong 2004; Debbarma 2017) briefly note the presence of reduplication, a systematic linguistic analysis is lacking.

Binoy Debbarma (2003) identifies 8 distinct dialects of Kokborok, namely Bru, Debbarma, Jamatia, Koloi, Murasing, Rupini, Tripura, and Uchoi. However, they have phonological variations among themselves but show mutual intelligibility. The closest linguistic relatives of Kokborok are Bodo (Boro), Mets, Dimasa Hodzai, Lalung (Tiwa), and Moran. Kokborok is an SOV language that exhibits rich morphological structures and typologically significant features, among which reduplication serves as a productive morphological and semantic process. Despite its importance, systematic studies of reduplication in Kokborok remain limited, especially those that show the mappings beyond iconicity.

Reduplication is a pervasive and multifunctional phenomenon in the world's languages (Moravcsik 1978; Rubino 2005; Inkelas 2014). It involves the repetition of a segment, syllable, or whole word to encode grammatical, semantic, or pragmatic information. Languages across diverse families, from Austronesian to Indo-Aryan, make extensive use of reduplication, often for iconic purposes such as intensification or plurality (Haiman 1980; Bybee et al. 1994). At the same time, reduplication can become routinized, losing its iconicity and functioning as a purely morphological marker (Inkelas and Zoll 2005). The study of reduplication therefore provides a window into the dynamics of form-meaning mapping, and typological diversity.

2. Literature Review

The interesting phenomenon of Reduplication across multiple language families in South Asia is explored in Anvita Abbi's ground-breaking research. A thorough examination of reduplication patterns, their structural forms, and the semantic nuances they communicate is given by *Abbi* in her 1992 book "*Reduplication in South Asian Languages: An Areal, Typological, and Historical Study*." The text discusses reduplication in a number of South Asian languages, including Bengali, Urdu, Hindi, and Tamil. It offers a thorough examination of the structures, purposes, and evolution of reduplication in various languages over time. Abbi explained and structured the text around three primary frameworks: Areal features, Typological perspective, and Historical study.

A landmark work in the field of linguistics, "Studies on Reduplication", edited by Bernhard Hurch, examines the phenomena of reduplication from a number of linguistic angles. The research of many academics is compiled in this volume to provide a thorough analysis of the structure, purpose, and typological variation of reduplication among languages. The main points of this volume are summarised in this review of the literature, which also places them in the larger context of reduplication studies. The first chapter by Bernhard Hurch gives a comprehensive introduction to the idea of reduplication and makes a distinction between full, partial, and echo reduplication. Hurch also addresses iconicity in reduplication, a process in which a segment's form frequently mirrors its pragmatic or semantic purpose.

The comparative study establishes that reduplication in South Asian languages shares similarities to other language families. Gil (2005), for example, talks about reduplication in Austronesian languages and observes patterns that are similar to those in South Asian languages. This comparison method aids in placing the book's results in a larger, worldwide context. The chapter by David Gil on reduplication in Austronesian languages makes a substantial addition to our understanding of reduplication typologically. Gil shows how reduplication is frequently used in Austronesian languages to denote intensity, plurality, or iterative features. His approach reveals the functional differences and similarities in the use of reduplication across languages.

'Echo Word Formation in Hindi and its Typological Implications' (1991), a seminal study by Anvita Abbi, focuses on a particular kind of reduplication common in Hindi and other South Asian languages. Echo-word formation is changing the first consonant or vowel of the underlying word to generate a semantically related or nonsensical reduplicative pair. Based on descriptive and comparative linguistic approaches, Abbi's work demonstrates different patterns of echo word generation in Hindi and identifies systematic phonological alterations. She supports the idea of South Asia as a linguistic region by highlighting the pragmatic uses of echo words, such as expressing non-specificity, generality, or a dismissive attitude. She also discusses the typological implications by drawing comparisons between similar processes in other South Asian languages. Anvita Abbi's key work "Echo-Word Formation in Hindi and its Typological Implications" offers thorough insights into a distinctive reduplicative occurrence in Hindi. Her comprehensive examination of the phonological, semantic, and pragmatic

features of echo words, along with a comparative typological viewpoint, makes this work a valuable asset for linguists with an interest in South Asian linguistics, reduplication, and language contact. Abbi has made a significant contribution to the discipline of typological linguistics with his work, which emphasises the value of analysing linguistic occurrences within their cultural and regional settings.

Subsequent research highlighted the role of reduplication in iconicity (Haiman 1980; Lakoff & Johnson 1980), while more recent studies emphasize its functional diversity, ranging from derivational to inflectional processes (Rubino 2005; Inkelas & Zoll 2005). The work '*Reduplication: Doubling in Morphology'* by Sharon Inkelas and Cheryl Zoll (2005) examines the theoretical foundations of reduplication in a variety of languages. Understanding the morphological process of reduplication which entails repeating a word whole or in part to communicate various meanings or grammatical functions requires familiarity with this book. Inkelas and Zoll tackle the topic theoretically, putting out a thorough framework to examine reduplication in the context of morphology and phonology. The Morphological Doubling Theory (MDT), which holds that reduplication entails the doubling of morphological structures rather than just phonological copying, is at the heart of their research. They contend that reduplication rather than just phonological redundancy should be interpreted as the duplication of morphological units, such roots, stems, or affixes, according to morphological criteria. The authors' research also incorporates phonological limitations, showing how these constraints combine with morphological rules to generate reduplication in three different forms: fixed-segment, partial, and entire. With a focus on Austronesian, Bantu, and Indo-European languages, the book offers a wealth of cross-linguistic data to demonstrate the various ways that reduplication manifests itself in different languages. This typological diversity bolsters the authors' case for a single theoretical framework by highlighting the adaptability and universality of reduplication. Reduplication has been better understood thanks to Inkelas and Zoll's work, which bridges the gap between morphology and phonology. The detailed examination of typological diversity and phonological limitations, in conjunction with the Morphological Doubling Theory, offers an extensive structure for comprehending this phenomenon in languages. As a linguist it has helped in understanding morphology, phonology, and the complex mechanisms that creates language structures.

Anvita Abbi's and other researchers' explorations of the phenomena of reduplication across different language families in South Asia offer important insights on the structural, semantic, and typological elements of this linguistic characteristic. Abbi's in-depth research on reduplication patterns in South Asian languages, which includes a thorough analysis of the development of echo-words, emphasises how common and important these patterns are throughout the region. Likewise, the theoretical works of Sharon Inkelas and Cheryl Zoll in 'Reduplication: Doubling in Morphology' and Bernhard Hurch's edited volume 'Studies on Reduplication' provide thorough frameworks for comprehending the morphological and phonological mechanisms underlying reduplication. Together, these studies highlight how crucial it is to examine reduplication in the context of culture and geography, offering a solid basis for linguistic studies. In Tibeto-Burman linguistics, reduplication has been described in languages such as Garo, Bodo, and Manipuri (Burling 2004; Sharma 1980; Joseph 2007). Kokborok, however, remains underrepresented. Early descriptions (Bhat 1969; Bhattacharya 1985) mention reduplication but provide limited data. Choudhury (2009) discusses Kokborok morphology broadly but does not analyse reduplication in detail.

Theoretically, reduplication has been framed in terms of **iconicity** (Haiman 1980; Lakoff and Johnson 1980). The notion is that formal repetition mirrors semantic iteration. However, more recent work highlights cases where reduplication has become semantically opaque or grammaticalised (Bybee et al. 1994; Stolz et al. 2011). This is crucial for Kokborok, where reduplication often encodes meanings that no longer bear iconic motivation.

3. Methodology and research questions

The central research questions are as follows:

1. What structural patterns of reduplication are attested in Kokborok?
2. What semantic and pragmatic functions do these reduplicative forms serve?
3. To what extent is reduplication in Kokborok iconic, and to what extent is it beyond iconicity?

Data for this study were collected through a combination of methods:

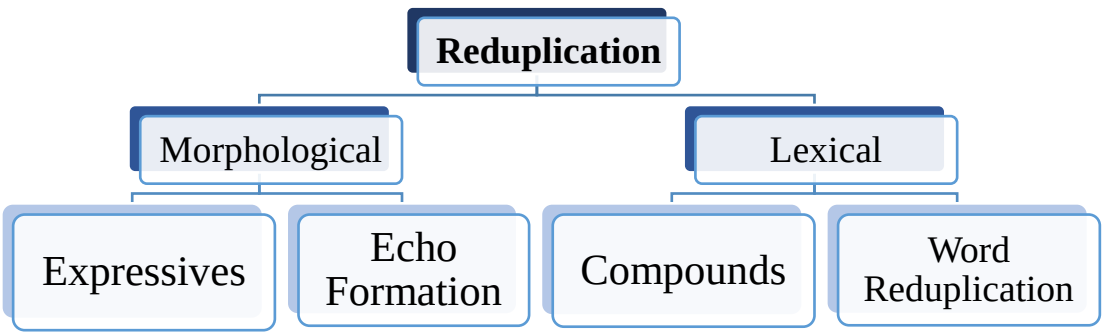
1. *Interview based questionnaire:* A short interview was held with native Kokborok speakers in Agartala, Tripura (W). They were asked to translate the sentences that were mostly in English and whenever there were constraints in graspinbg the reduplication, sentences were translated in either Bengali or Hindi for the ease of the participants. The same set of data was verified with 12 speakers to avoid any anomaly in the analysis.
2. *Analytical framework:* Data collected were transcribed using the Bloch and Trager method of transcription and glossed as per the Leipzig Glossing Rules. Semantic functions were categorized according to Rubino (2005) and Inkelas (2014).

The dataset for this paper are classified into full, partial, discontinuous, echo, and expressive reduplication.

4. Data Analysis

Reduplication is an extremely productive morphological process that enhances a language's lexical richness and expressivity by producing new words. Reduplication can convey plurality, intensity, or repetitive acts by repeating a word whole or in part, resulting in new forms with unique meanings. It enables the creation of idiomatic expressions, the conversion of nouns into adjectives or adverbs, and the expression of transitory states. Additionally, by improving phonological appeal and rhythm, this procedure increases the memorability and cultural significance of phrases. Reduplication broadens a language's semantic range overall, which adds to its dynamic and subtle quality.

Reduplicated structures are widely used, particularly in the field of modification, for a variety of semantic purposes. They work as effective language instruments to amplify words' descriptive qualities and communicate complex meanings. Reduplication modifies the original term to represent different degrees and states by indicating emphasis, plurality, frequency, and intensity.



4.1 Morphological Reduplication

Morphological Reduplication describes morphemes made up of repeated syllables that have limited significance and are segmentally indivisible. Therefore, the iterated portion and the base together make up a single morpheme, which is also a lexeme.

Expressives

Expressives are reduplicated and belong to a single lexical category. We do not think of expressives as 'acoustic symbols' exclusively because not all of them depict noises. Expressions are not "lexically discrete" in the same sense as other words and word classes. They don't look as iconic as they are supposed to.

i.	t̪ɯi	sɔu~sɔu-kʰe	kʰɯlai-t̪ɔŋgɔ
	water-3S-NOM	sound of water flowing-REDUP	be-PRES-IMPF
	The water is flowing down.		

ii.	bɔ	d̪ɔga	t̪ʰɔk~t̪ʰɔk	kʰai-t̪ɔŋgu
	he-3MS-NOM	door-3S-ACC	sound of knocking (repeatedly)-REDUP	be-PRES-IMPF
	He is knocking the door.			

In the sentences above, [sɔu~sɔu] is the sound of water flowing but [t̪ʰɔk~t̪ʰɔk] is not only the sound of knocking but the sound of repeated knocks. Hence, both qualifies for expressiveness due to its ideophonic characteristic. In the second sentence, the reduplication suggests that it is a resource for expressive intensification, extending beyond grammatical function into stylistic and ragmatic domains.

Again, due to the process of lexicalisation, the expressives become derivational modifier. Derivational modifiers enhance the expressive power of language by changing verbs to depict events or situations in a vivid and complex manner. But this is not overtly evident in Kokborok.

Echo Formation

A definition of an echo word is a form of the base word that is partially repeated, meaning that a different phoneme or syllable replaces either the initial phoneme (which might be a vowel or a consonant) or the base syllable. The replacer (phoneme/syllable) sound sequences are more or less fixed and rigid. (Abbi, 1992) While they might not always be distinct, the replacer sound sequences are unlikely to be numerous. Punjabi has [š-]. Hindi has [v-]. Bengali has [ṭ-] as their common replacer sound. Therefore, an echo construction is any construction in which an echo word comes after the basis word (or, in rare circumstances, before it). The language does not contain a single instance of the echo word or a distinct meaning for it. It only becomes a significant aspect once the "word" is attached to it.

i.	aŋ		kisa~misa	sɔm	naŋgu
	I-1S-NOM		a little bit-REDUP	salt-3PL-ACC	want-PRES
			I want a little bit of salt.		

[kisa] means 'a little bit' and in the above sentence has been reduplicated as [misa] that doesn't have a semanticity of its own but acts as an echo of the base form that together gives a synonymous semanticity that [t̪ʰora-vora] provides in Hindi.

ii.	aŋ	nɔk~hɔk-rɔgɔ	t̪ʰanai
	I-1S-NOM	house-REDUP-LOC	go-FUT-IMPF
	I'll go to house, etc.		

[nɔk] denotes 'house'. In [nɔk~hɔk], the phoneme /h/ carrying the echo where it itself has no meaning of its own. But the reduplication together means house, etcetera.

iii. nuŋ́ mai~ʈai kʰa-ɖe
you-2S food-REDUP eat-PRES-PERF-QM
Have you eaten food and all?

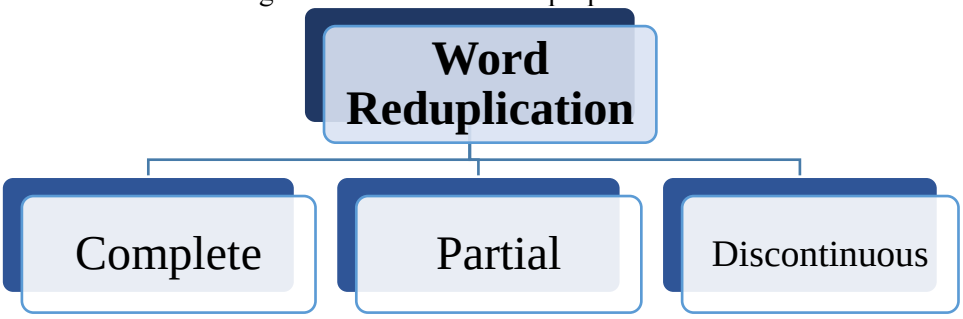
[mai~ʈai] together gives rise to a new semantic value. [mai] means meal or food where [ʈai] can mean drink or any complementary items taken with food, however, [ʈai] doesn't bear a meaning of its own.

4.2 Lexical Reduplication

Lexical reduplication occurs when a word or a part of word is repeated to produce a new lexical item with a particular meaning or purpose. Lexical reduplication has been examined by several linguists, who have classified it based on its forms, functions, and semantic characteristics. Reduplication can signify repeated activity, multiplicity, or increase meaning.

Edward Sapir, for example, highlighted the significance of reduplication in expressing complex meanings and pointed out that it may fulfil grammatical purposes including aspect, plurality, and tense.

Roman Jakobson studied phonological features, emphasising recurring patterns in languages. Charles F. Hockett examined the morphological procedure of reduplication in the creation of new words or forms, while Paul Kiparsky examined the generative phonology that underlies these forms. Frans Plank examined typological features in many languages, noting that reduplication fulfils distinct grammatical and lexical purposes.



Word Reduplication

Abbi defines word reduplication as the total or partial bimodal reduplication, meaning thereby repetition of the base of the word or of the stem. Reduplication can be of either a syllable or a larger constituent of a word or of the whole word. Whatever the unit of reduplication, the end result is a new word which has no parallel in its non-reduplicated counterpart.

Full/Complete Word Reduplication

Complete reduplication refers to those paired constructions when a single word or a clause is repeated in the same sentence without any phonological or morphological variation (Abbi, 1975).

Word reduplication can be either complete or partial. Complete reduplication involves structures where the entire word is repeated, resulting in two identical units. This means that any phonological sequence that forms a word can be duplicated, creating a structure where a word "x" becomes "xx". Because this form of reduplication involves the total repetition of the word, it has been referred to as complete word reduplication (CWR) in previous works by the author (Abbi, 1975, 1984).

i. bɔ kʰačikmaŋ~kʰačikmaŋ leŋ-ǰakʰa
he-3MS-NOM run-INF-REDUP tire-PST-PERF
He got tired of running and running.

ii. aŋ kok samaŋ~samaŋ ʈʰagɔi ʈʰaŋkʰa
I-1S-NOM word speak-REDUP stop be-PST-PERF
I stopped while speaking.

In the first sentence, [kʰačikmaŋ~kʰačikmaŋ], an INF meaning 'of running continuously' is an example of CWR. Post reduplication, it adds a continuity or progressiveness to the word 'run'. Similarly, the second sentence, the exact character is displayed by [samaŋ~samaŋ] (speaking continuously).

iii.	ʈuruk~ʈuruk-kʰe	himɖi	hayake	nɯŋ	kɯlainai
	slowly-REDUP	walk-PRES	else	you-2S-NOM	fall-FUT-IMPF
	Walk slowly, else you'll fall.				

The adverb of manner [ʈuruk] itself means 'slowly'. When it's reduplicated, it essentially decrease the pace which in turn means iconic increase in 'being slow' in walking and hence acts as a caution in the sentence above. Here, due to the adverb 'slowly' a negative intensification can be achieved. The semantic link is transparent. This reflects Haiman's (1980) claim that quantity of expression parallels quantity of meaning.

iv.	ɔ	kok	ǰɔʈɔ	nɔk~nɔk-ni	kɔʈʰɔma
	this	story-3S-NOM	all	house-REDUP-GEN	matter-3S-ACC
	This is a matter of every household.				

[nɔk] denotes house (singular). Upon reduplication, [nɔk-nɔk] denotes a plurality which is synonymous to [gʰar~gʰar] in Hindi.

v.	ɖɔ	kisa	kuʈuŋ~kuʈuŋ	mai	ʈubugaɖi
	please	some	hot-REDUP	rice	bring-PRES-IMP
	Please bring some piping hot rice.				

[kuʈuŋ] (hot) qualifies as a CWR as it adds on to the degree of hotness upon reduplication.

vi.	pini~pini	uaʈuɥi	wai	ʈɔŋgɔ	faʈar-ɔ
	slight-REDUP	rain	fall-IMPF	be-PRES-IMPF	outside-LOC
	It's drizzling outside.				

In the sentence above, [pini~pini], a reduplication for 'a little bit' or 'slight' expresses the quantity in totality.

Partial Reduplication

Structurally viewed, word reduplication can also be constituted either by (i) duplicating a part of the word (generally a syllable) or by (ii) disjoining the two iterated words by a syllable. The former may be termed as PARTIAL WORD REDUPLICATION (PWR). It should be noted that the syllable which is the constituent part of the word to be duplicated can either be a vowel (v) or a vowel consonant (CV), or consonant-vowel (CV). For instance, Santhali sen > seen or aium > aaium or ɖal > ɖa~ɖal. Partial Word Reduplication may be considered a subpart of partial lexical reduplication. (Abbi, 1975)

According to Abbi, Partial word reduplication may involve repetition of any of the constituents of a word. Root/stem reduplication without a reduplicating affix is an instance of partial word reduplication.

i.	ɔ	mui	ano	ʈʰɔkprai~prai	kʰai- ruɥkʰa
	that	curry	I-DAT	unsatiate-REDUP	do-PST-PERF
	The curry left me unsatiated.				

Partial reduplication is a rare phenomenon in Kokborok. But the sentence above has [ʈʰɔkprai~prai] as a PWR meaning 'the feeling of being unsatiated'. The syllable [prai] has been extracted from [ʈʰɔkprai] and reduplicated with the same to bring out the unsatiated-ness.

Discontinuous Reduplication

When the reduplicated structure is separated by an interference of a syllable that can be consonant or a vowel this phenomenon is called Discontinuous Reduplication. For example [lal~e~lal] (red all around), [ačč^{h}e~se~ačč^{h}e] (even the good ones), [kəm~se~kəm] (minimum or a little) in Hindi language.

i. kɯčak~bai~kɯčak simi nugu
red and red-REDUP only see-PRES-PERF
Only red and red is seen.

The example above mentions [kɯčak~bai~kɯčak] (red and red) as a discontinuous reduplication. It shows the high degree of the colour red that is seen all across. It unfolds an intensification of the visibility of the colour red.

ii. bɔ k^{h}a~bai~k^{h}a kɯrɯŋ-gəi samuŋ k^{h}ɯlay-ɔ
he-3MS-NOM heart-REDUP connect work do-PRES-HAB
He works with a heartfelt connection.

[k^{h}a~bai~k^{h}a] is synonymous to the Hindi discontinuous reduplication [d̪il~se~d̪il] (heartfelt).

Compound/ Semantic Reduplication

Compounds are said to be those combination of words in which the second word is not an exact repletion of the base word but shares somewhat similarities that can be either semantically or phonetically. Each word in Compound construction has its own meaning and they can stand alone in a sentence. However, when combined they create a new meaning, that too with a new reference. For example [ʊṭ^{h}na-bɛṭ^{h}na], in this construction [ʊṭ^{h}na] (to rise), [bɛṭ^{h}na] (to sit) the both the words has meaning on their own, but when used as compound [ʊṭ^{h}na-bɛṭ^{h}na] mean 'frequenting'.

Compounds are formed in two different ways-

1. When the words are semantically identical. For example, in a construction [d̪^{h}ən-d̪ɔlət̪] the first is from Hindi and the second word is from Urdu, they both mean the same thing.

2. When the semantically connected words are not phonologically duplicated yet they have a semantic link between them. For example- Hindi constructions like [ṭol-mol] (evaluation-bargain) and constructions like [šad̪i-viyah] where both the word means marriage.

In Kokborok, various forms of reduplication convey different semantic nuances, such as plurality, intensity, or specific attributes. Kokborok equally exhibits exocentric compound reduplication forms where the both the words don't carry a meaning of their own but together, they give rise to a new semanticity.

i. bɔ sal-hɔr samuɰŋ k^{h}wlay-o
he-3MS-NOM day-night-REDUP work do-PRES-HAB
He works day and night.

[sal] (day) and [hɔr] (night), individually denotes two extreme points of a 24 hour cycle. [sal-hɔr] meaning 'day and night' adds both the points of the day.

ii. ɔ sɯi kaham-kɯrəŋ t̪ɔŋgɔ
the dog-3MS-NOM well-built-REDUP be-PRES-HAB
The dog looks well-built.

[kaham-kɯrəŋ] suggesting that the dog is well-built is a semantic reduplication, synonymous with [mɔṭa-t̪aza] in Hindi. The individual words together posits a new semantic intensification.

iii. ayaŋ-ɔyaŋ ʈa ʈʰaŋɖi
here-there-REDUP NEG go-PRES-IMP
Don't go here and there.

Similarly, [ayaŋ-ɔyaŋ] (here and there) is synonymous for [iɖʰar-uɖʰar] in Hindi. In discourse, it indicates an aimless spatial entity.

iv. nɯŋ čamɯŋ~nɯŋmɯŋ- rɔk čakʰa-ɖa
you-2S-NOM food-REDUP-PLM eat-PRES-PERF-QM
Have you eaten food and all?

Here, [čamɯŋ] means 'food' and [nɯŋmɯŋ] means drink. Hence, [čamɯŋ~nɯŋmɯŋ] may seem to form Echo Formation due to its form. When considered the meaning of the parts, it's seen to constitute a copulative compound (having two heads).

In Echo Formation, the reduplicated part (here, [nɯŋmɯŋ]) generally doesn't contain meaning of its own but carries the echo of the base word that has a definite meaning. Generally, in Hindi when we say [kʰana-vana] ('food, etc' or 'food and all'), we end up intending 'drink' or 'water' through the echo [vana].

However, in this case, [nɯŋmɯŋ] has a meaning of its own (drink) and hence [čamɯŋ~nɯŋmɯŋ] together means 'food and drink' or a complete meal, in Hindi [kʰana-pina]. Hence, [čamɯŋ~nɯŋmɯŋ] is a semantic reduplication, judged beyond the traditional form.

v. bɔ hapuŋ-sapuŋ pɔriɔ
he-3MS-NOM day-night-REDUP study-PRES-HAB
He studies all day and night.

In the above sentence, [hapuŋ] solely means all day. [hapuŋ-sapuŋ] denotes all day and night, synchronically establishing an everyday phenomenon.

vi. ɔ lama asʈa ʈʰeʈrek~buʈrek
this road-3S-NOM entire coarse-REDUP
This entire road is coarse.

In the sentence above, [ʈʰeʈrek~buʈrek] together denotes coarseness (not smooth), and can be called a compound reduplication since individually none of the words exist or possess any semantic identity. Hence, despite the fact that [b] carries the echo of [ʈʰ], they don't qualify as Echo Reduplication.

Reduplication as iconic form-meaning mapping

Iconicity is one of the most frequently discussed motivations for reduplication cross-linguistically (Haiman 1980; Kouwenberg & LaCharité 2001). The principle of "more form, more meaning" is directly observable in Kokborok adverbial reduplication.

i. ɖaʈi~ɖaʈi
quickly-REDUP
'very quickly'

ii. baharɯi~baharɯi
jumping-REDUP
'jumping continuously'

The above examples index a process where the repeated linguistic form mirrors the semantic nuance of iteration and graduality in the action. [d̪at̪i] means quickly but when reduplicated, it increases the intensity, resulting in intensification. Similarly, [baharɯi] (jumping) shows a continuity or progression when reduplicated. Speakers often exploit this form when giving instructions, thereby leveraging the perceptible link between repetition in form and repetition in action.

iii. nɔk~nɔk
house-REDUP
'all houses'

The reduplicated nominal [nɔk] explicitly encodes plurality. Here, multiple tokens of the same word reflect multiple referents in the world. As Moravcsik (1978) and Inkelas & Zoll (2005) note, this is a cross-linguistically common pattern in reduplicative systems where duplication visually and auditorily corresponds to multiplicity.

iv. ĩ~ĩ
yes-REDUP
'yes yes/ of course'

In conversational exchanges, [ĩ~ĩ] does not iconically reflect doubling of agreement, but functions as a discourse particle signaling emphatic affirmation. In many cases, speakers perceive this as a fixed conventional form, not as a creative repetition of [ĩ]. The iconicity here is negligible; instead, the form serves a pragmatic, interactional function. This mirrors similar discourse reduplication phenomena in Japanese (hai-hai, Cook 1990) and Hindi [hã̄-hã̄], where repetition has become pragmatically entrenched.

5. Conclusion

The data show that Kokborok reduplication is multi-layered, ranging from iconic intensification to more conventionalized expressive uses. Full reduplication is closest to iconicity, where repetition signifies plurality or emphasis. Echo reduplication, while structurally similar to Indo-Aryan, appears more restricted in Kokborok, reflecting contact influence but with localized usage. These findings support theories that reduplication operates along a continuum from lexical to grammatical functions (Inkelas 2014; Singh 2005). The multi-functionality of Kokborok reduplication challenges a purely iconic account, showing that repetition has been conventionalized into distinct morphological strategies.

Moreover, the prevalence of reduplication in oral narratives highlights its role in discourse, where repetition reinforces rhythm, emphasis, and emotional engagement (Rubino 2005). This indicates that reduplication in Kokborok is not merely morphological but is integrated into communicative style and identity.

This study has examined reduplication in Kokborok with attention to its forms, functions, and meanings. Full, partial, and echo reduplication serve different morphological and semantic roles, from intensification to expressive uses. While some functions align with iconicity and others by discourse-driven conventions, suggesting that reduplication in Kokborok is both motivated by form–meaning mappings and shaped by conventionalized usage. The findings contribute to broader typological work on reduplication and highlight the need for further research on Kokborok, particularly corpus-based studies and sociolinguistic investigations of variation in reduplicative usage.

References

Abbi, A. (1992). Reduplication in South Asian Languages: An Areal, Typological and Historical Study. Allied Publishers.

Basu, D. (2002). A Grammar of Kokborok. Agartala: Tribal Research Institute.

Burling, R. (2004). The Language of the Modhupur Mandi. SIL International.

Darlong, H. (2017). Kokborok Grammar. Agartala: Directorate of Kokborok.

Dasgupta, P. (2003). *Bangla reduplication: Iconicity and grammaticalization*. Journal of South Asian Linguistics.

Debbarma, B. (2009). Kokborok Phonology and Morphology. Agartala: Tripura University Press.

Dwivedi, R. (1980). Kokborok: A Descriptive Grammar. Agartala: TRI.

Haiman, J. (1980). The iconicity of grammar: isomorphism and motivation. Language, 56(3), 515–540.

Heine, B., & Kuteva, T. (2007). *The Genesis of Grammar*. Oxford: Oxford University Press.

Inkelas, S. (2014). The Interplay of Morphology and Phonology. Oxford University Press.

Inkelas, S., & Zoll, C. (2005). Reduplication: Doubling in Morphology. Cambridge University Press.

Joseph, U., & Burling, R. (2006). Boro-Garo languages. In G. Cardona & D. Jain (Eds.), The Indo-Aryan Languages (pp. 630–653). Routledge.

Lakoff, G., & Johnson, M. (1980). Metaphors We Live By. University of Chicago Press.

Moravcsik, E. (1978). Reduplicative constructions. In J. Greenberg (Ed.), Universals of Human Language, Vol. 3 (pp. 297–334). Stanford University Press.

Radhamohan, D. (2002). Kokborok–English Dictionary. Agartala: TRI.

Rubino, C. (2005). Reduplication: Form, function and distribution. In B. Hurch (Ed.), Studies on Reduplication (pp. 11–29). Mouton de Gruyter.

Singh, R. (2005). The Structure of Reduplication: An Indian Perspective. Motilal Banarsidass.

Stolz, T., Kettler, S., Stroh, C., & Urdze, A. (2011). *Total Reduplication: The Areal Linguistics of a Morphological Pattern*. Berlin: Mouton de Gruyter.

वाक् मंथन
ISSN: 2426-2149

Property Concept (PC) expressions in Telangana Telugu in attributive modification position.

Nikhilesh Thirukovela
nikil.thirukovela@gmail.com

Abstract

Dixon (2004) classifies the world's languages into two types: **α**-type, which have adjectives representing all semantic classes, and **β**-type, which have a limited set of adjectives restricted to core semantic classes. Telangana Telugu (TT), a regional variety of Telugu, falls into the **β**-type category, where most adjectives are nominal in nature and are better analyzed as Property Concept (PC)-expressions rather than a distinct adjective class. This study investigates the morphological formation of adjectives in TT and their syntactic integration in attributive modification position. Based on fieldwork data, this paper identifies five distinct markers that occur with attributive PC-expressions in TT: *-ayina, -ti, -ni, -pu*, and a null marker. These markers exhibit specific distributional properties, determining how PC-expressions function in a sentence. The study further categorizes these markers into different subtypes:

1. null PCs: property concepts that appear without overt markers.
2. -ayina PCs with 2 subtypes:
    a. Obligatory -ayina PCs: Must always appear with the -ayina marker.
    b. Optional -ayina PCs: May occur with or without -ayina.
3. -ti/ni/pu PCs with 2 subtypes:
    a. PCs that occur with all three markers.
    b. PCs with ti/ni markers that are restricted in their co-occurrence with -pu.

The study demonstrates that TT, like other Dravidian languages, lacks a robust lexical adjective class, instead composing PC expressions derivationally from nouns and verbs via relativization. These findings support Dixon's classification of Dravidian as a (β)-type language, where most PCs are syntactically derived rather than lexical adjectives. This study reinforces the cross-linguistic evidence for **β**-type languages and their reliance on nominal strategies for adjectival expressions. It also provides new insights into the morphosyntactic diversity of Dravidian languages, particularly in how PC-expressions operate in attributive modification positions. The analysis underscores the role of copulas and relative clause morphology in PC formation, contributing to broader typological discussions on property-concept encoding.

Keywords: Property Concept (PC) - expressions, Adjectives, Telangana Telugu, Relative clauses, Copulas, Attributive modification, Dravidian syntax, Lexical categories.

1. Introduction: What is Property Concept (PC) expressions?

The Property Concept (PC)-expressions, by definition, are the semantic categories that appear as the syntactic category 'adjectives' in languages with lexical adjectives. For example, the lexical items such as: '*good', 'bad', 'smart',* and '*beautiful'* in English appear as lexical adjectives. This can be seen below:

The *good/bad/smart/beautiful* girl.

In languages with limited or no lexical adjectives, the majority of PC expressions belong to the word classes of nouns and verbs. For example, in Korean, PC expressions function as a subclass of verbs (Haspelmath, 2001). They necessitate the use of a relative suffix when modifying a noun, as demonstrated by the PC expression *noph* "high" in example (1). Similarly, in Hausa (Chadic), PC expressions appear as nouns, as shown by the PC expression *ƙarfī* "strength" in example (2)

1. **noph**-un san (Haspelmath, 2001)
   high-RELATIVE hill
   'a high hill'

2. munã dã **ƙarfī.** (Newman 2000:224)
   we.cont with strength
   'we are strong.'

Similarly, the PC expressions in the Dravidian language family also function as subclasses of nouns and verbs and use different strategies in order to appear in the attributive modification position. This can be seen below.

1.1. PC expressions in Dravidian language Family

In the Dravidian language family, Menon (2012) and Herur (2016) show that the adjectives are composed from other syntactic categories such as nouns and verbs in Malayalam and Kannada. For Telugu, Balusu (2015) shows that adjectives are composed from nouns. The examples are given below in (3-5),

3. koopa-da maatu-gaḷu (Kannada, Herur 2016: 39)
anger -GEN word –PL.NOM
'words of anger' (lit. anger's words)

4. santhosham ull -a kutti (Malayalam,Menon2014:ex14b)
happy COP-REL child
'happy child' (lit: child that is happy)

5. siita **erupu** (Telugu, Balusu 2015:2)
sita redness
'sita is red'

Amritavalli and Jayaseelan (2004) and Jayaseelan (2007) propose an incorporation theory for adjectives for Dravidian languages. They posit that all adjectives are formed by incorporating the nominals (N) into the dative case (K_{dat}), resulting in the base structure of,

DP (NOM) be AdjP e.g. I am happy/hungry.

However, in Dravidian languages, this process of incorporating adjectives is less frequent due to the stable case-system. As a result, the majority of the PC-expressions are expressed using dative suffixation or other denominal compositional structures. This results in the structure of,

DP-DAT be NP e.g. a. en-ikke santooSam uNDa (Jayaseelan 2007: 48)
(To me happiness is)

Here, the PC-nominals do not undergo incorporation into K_{dat} and remain as nominals. While this explains the frequent use of dative case in Dravidian for PC-expressions, this evidence crucially informs that PC-expressions in the Dravidian languages cannot be classified solely as adjectives. They are composed from other lexical categories such as nouns and verbs, indicating that they are compositional in nature. In terms of Dixon (2004), this makes Dravidian languages (***β***)-type languages, which contain only a small adjective class belonging to core semantic types, while majority of PCs are composed from other lexical classes within syntax. This raises the question of how PC-expressions are composed in Dravidian. In this context, this study looks at Telangana Telugu (TT), a, areal variant of Telugu, one of the four major Dravidian languages spoken in the areas of Telangana and Andhra Pradesh. This is discussed in the next section.

2. PC expressions in Telangana Telugu: An overview

For this study, a dataset of 57 PC expressions in TT is compiled and analyzed. Data is collected mainly from the four districts of northern Telangana: Karimnagar, Adilabad, Nizamabad, and Warangal. Participants completed picture-based tasks, i.e. describing objects with contrasting properties like *pedda* 'big' vs. *chinna* 'small' to elicit attributive PC constructions. To ensure reliability, the data was cross-checked with other native speakers of TT across the specified area.

2.1. Primary data and Findings:

Analyzing the primary data, it was found that in TT, PCs appear with at least 5 different morphological markers in attributive modification position and 3 different structures in predicative positions. For this paper, we focus only on PCs in attributive modification position.

Attributive modification: The noun modification structure where the PC appears within the same noun phrase as the modified noun and provides the modification semantics from within the NP is called attributive modification. An example of this can be seen in the prenominal modification in English such as

The *beautiful/angry/tall/smart* girl.

PC expressions can take 5 different markers in TT in attributive modification positions. They are divided as follows:

Out of 57 PCs, 24 PCs can appear in the attributive modification position without the need for any marker. For example, *pedda* 'big' and *chinna* 'small' in the examples (6-7)

6. pedda chettu
   big tree
   'big tree'

7. chinna chettu
   small tree
   'small tree'

Here, we can see that the PCs *pedda* 'big' and *chinna* 'small' do not need any overt markers and can appear directly juxtaposed to the modified noun in the attributive modification position. I call these 'null PCs' as they do not need any marker in order to appear in the attributive modification position. Out of 24 null PCs, 18 PCs are obligatory null PCs, i.e. they cannot take any other marker and will always appear without any marker in the attributive modification position. The above examples of *pedda* 'big' and *chinna* 'small' (6-7) belong to this category. The remaining 6 PCs are classified as optional null PCs as they can appear without any marker, but they can also optionally appear with other markers such as '*-ayina*'. These are discussed below.

Of 57 PCs, 26 PCs appear with *-ayina*' marker in the attributive modification position. Of these, 6 are optional i.e. they can take *-ayina* marker but they can also appear as null PCs (8) whereas 20 PCs obligatorily need *-ayina* marker in order to appear in the attributive modification position (9).

8. podug-(ayina) ammayi (optional)
   tall-(ayina) girl
   'tall girl' (lit.the girl who is tall)

9. teliv-*(ayina) ammayi (obligatory)
   intelligent-(ayina) girl
   'intelligent girl' (lit. the girl who is intelligent)

I call these PCs '*-ayina*' PCs as these PCs use an overt '*-ayina*' marker in order to appear in the attributive modification position.

Of 57 PCs, 13 PCs use either of *-ti, -ni or -pu* markers in order to appear in the attributive modification position. For the time being, I classify all of these as the same class as 5 of these 13 PCs can take either of *ti, ni*, or *pu* markers (10) and the remaining 8 PCs resist *-pu* markers but can appear with either of *ti* or *ni* markers (11).

10. sanna-(ti/ni/pu) teega (ti/ni/pu)
    slim-(ti/ni/pu) wire
    'slim wire'

11. tella-(ti/ni/*pu) angi (ti/ni/*pu)
white-(ti/ni) shirt
'white shirt'

I call these 'ti/ni/pu' PCs as these PCs appear with either a selection of or all of '*ti', '-ni*' and '*-pu*' markers on them in the attributive modification position. This is illustrated in the chart below:

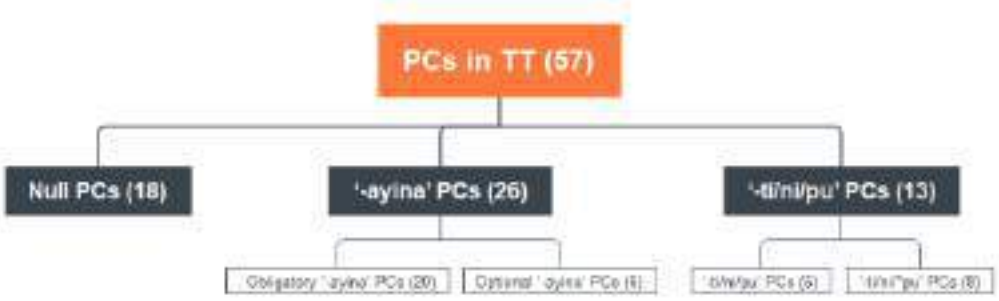


Fig 1: PC classes in TT (tentative)

This shows that out of 57 PCs in TT, there are 18 null PCs, 26 '*-ayina*' PCs and 13 *-ti/ni/pu* PCs. Of the 26 *'-ayina*' PCs, 20 are obligatory while 6 are optional. Of the 13 *-ti/ni/pu* PCs, 5 can take all *-ti/ni* and *-pu* markers while 8 can take only *-ti/ni* but not *-pu*. Each of these classes is discussed in further detail in the following section.

3. Analysis: PC classes in TT:

In this section, I provide a detailed discussion and analysis of all the 3 classes of PCs introduced in section 2.

3.1. Class 1: null PCs:

As seen in the previous section, these are PC-expressions that do not need any additional grammatical markers in order to appear in the attributive modification position. Some examples include the PCs *pedda* 'big', *chinna* 'small', *podugu* 'tall', and *vedalpu* 'wide'. When tested for syntactic categories, these PCs show adjectival properties. Therefore, they are considered as true adjectives in TT. For example,

12. pedda chettu
big tree
'big tree'

13. chinna chettu
small tree
'small tree'

Here, the PCs *pedda* 'big', *chinna* 'small', appear in the attributive modification position in their bare form without the need for any markers.

As shown by (Herur, 2014), a cross-linguistic parallel for these types of PCs can be found in another Dravidian language, Kannada. These are similar to the olleya-group of PC-adjectives found in Kannada.

14. olleya huduga (Kannada , Herur 2014:38)
    good boy
    'good boy'

These PCs can be classified as a small subset of non-deverbal/denominal adjectives present in all (***β***)-type languages. Dixon (2004) notes that all such PCs in (***β***)-type languages always belong to the four core semantic types: DIMENSION, AGE, VALUE and COLOR. This is also true for TT, as all 18 null PCs from TT belong to these 4 core semantic types. Further, null PCs in TT can be divided into two subtypes based on their optionality. These are,

3.1.1. Obligatory null PCs:

These are null PCs that must appear without any additional markers in the attributive modification position. These PCs show adjectival properties when tested for syntactic categories, indicating that they are, in fact, true adjectives. This is shown below,

15. pedda chettu
    big tree
    'big tree'

16. chinna chettu
    small tree
    'small tree'

3.1.2. Optional null PCs:

These are null PCs that can appear in their attributive position without any additional marker but can also appear with an optional *'-ayina'* marker. For example,

17. podug-(ayina) chettu
    tall-(ayina) tree
    'tall tree' (Lit: the tree that is tall)

18. vedalp-(ayina) chettu
    wide-(ayina) tree
    'wide tree'(Lit: the tree that is wide)

3.2. Class 2: '-ayina' PCs:

These are PCs that use the *-ayina*' marker in order to modify the subject noun in the attributive modification position. For example,

19. teliv-ayina ammayi
    intelligent-ayina girl
    'intelligent girl' (Lit: the girl who is intelligent)

This *-ayina*' marker is a morphological cluster made up of multiple markers. It is broken down and analyzed as follows.

3.2.1. Deconstruction of -ayina' marker:

The morphological cluster of *-ayina*' is composed of three distinct morphemes that have been agglutinated together. These include the Change of State (CoS) copula *avu* 'become', the past tense marker '*-in*', and the relative clause (RC) marker '*-a*'. The evidence for each of these markers is provided below.

3.2.2. The CoS copula -avu 'become':

The '-avu' morpheme in TT corresponds to the copula of the change of state type, which changes the state of the subject nominal within a temporal frame. For example, consider the sentences presented below:

20. ram manager
    ram manager-COP
    'ram is a manager'

21. ram manager ay-in-du
    ram manager COP-PST-3PSgM
    'ram became a manager'

In (20), the null equative copula in TT forms an individual-level predicate (ILP) that gives stative semantics. This sentence does not contain any temporality encoded in it and lacks any information regarding the timeframe of when Ram was a manager. It simply states that Ram is a manager. However, the CoS copula *-avu* 'become' in (21) forms a stage-level predicate (SLP) and gives Change of State semantics, indicating that the state of the subject noun 'Ram' has changed from not being a manager earlier to being a manager now. It encodes information that Ram 'became' a manager. This establishes that the morpheme '-avu' which appears as '-ay' is a CoS copula in TT.

3.2.3. The past tense marker [+PST] '-in'

The '-in' marker in TT corresponds to the past tense marker and is often affixed immediately to the copulas and verbs. This is shown in the examples below,

22. ram ninna vach-in-du
    ram yesterday come-PST-3PSgM
    'ram came yesterday'

23. sita market-ki poy-in-di
    sita market-DAT go-PST-3PSgF
    'sita went to the market'

24. ram doctor ay-in-du
    ram doctor COP-PST-3PSgM
    'ram became a doctor'

The above examples show that '*-in*' is a regular past tense marker in TT. Previously, the examples from the '*-ayina*' constructions in TT show that '*-in*' appears in a similar position, immediately affixed to the CoS copula '*-ay*' and prefixed to the agreement marker. Based on this evidence, it can be concluded that '*-in*' in *'ayina'* is a past tense marker.

3.2.4. The Relative clause marker '-a'

The '*-a*' marker in TT corresponds to the relative clause marker. Subbarao (2012) gives the following sentence (25) for Externally headed relative clauses (EHRCs) in Telugu.

25. nenu tin-(i)n-a aaku (Subbarao 2012:178)
    I eat-PST-adjr leaf
    'the leaf that I ate'

Here, we can see that the '*-a*' marker affixes immediately after the past tense marker '*-in*' in the morphological cluster of Verb-Tense-Relativeclausemarker (V-T-REL). This is similar to the morphological cluster seen in the '*ayina*' construction. This makes the '*-ayina*' the same V-T-REL cluster seen in the relative clause constructions in TT. Following this, I refer to the '*-ayina*' PCs as relative clause (RC) PCs. They are represented as follows,

26. teliv-ay-in-a ammayi
    intelligent-COP-PST-REL girl
    'intelligent girl' (Lit: the girl who is intelligent)

Here, we can see that the '*-ayina*' marker is deconstructed into a morphological cluster containing 3 morphemes: '*ay-in-a*', glossed as COP for the copula '*-ay*', PST for past tense marker '*-in*' and REL for relative clause marker '*-a*'. This gives the literal transcription of the PC modification structure as 'the girl that is intelligent'.

3.3. Class 3: ti/ni/pu PCs

The third and final class of PCs that appear in attributive modification context in TT is the ti/ni/pu class PCs. These PCs require the use of one of the three markers '*-ti', '-ni*' or '*-pu*' in order to appear in the attributive modification position. This class contains two types of PCs. Some PCs take any of the three markers (*-ti/-ni/-pu*) such as *sanna* 'slim' in (27) while others can only take either *-ti* or *-ni* but not *-pu*, such as *tella* 'white' in (28).

27. sanna-(ti/ni/pu) teega (ti/ni/pu)
    slim-(ti/ni/pu) wire
    'slim wire'

28. tella-(ti/ni/*pu) angi (ti/ni/*pu)
    white-(ti/ni) shirt
    'white shirt

Two key observations can be made from the data of the PCs from this class. Firstly, all 3 markers *ti/ni/pu* are optional for all PCs except for three PCs (*nunna* 'smooth', *thiyya* 'sweet' and *sakka* 'nice') in this class. Except for these 3 PCs, all other PCs can also appear without any markers. Secondly, all PCs from this class end in '*-a*', as seen in examples such as *sanna* 'slim', *tella* 'white', *mettha* 'soft', *salla* 'cold', *nunna* 'smooth', *sakka* 'nice'.

This suggests the possibility that these structures are also relative clauses, similar to the RCPCs seen previously. The only difference is that in '*-ay-in-a*' class RC PCs, there is a CoS copula '*avu*' become which accompanies the relative clause marker '*-a*,' while in this class, an equative copula, which is null in TT, is accompanied by the relative clause marker '*-a*' in the same 'V-T-REL' morphological cluster. Additionally, these relative clauses can take *-ti/-ni/-pu* markers optionally. This is represented as follows,

29. sann-a-(ti/ni/pu) teega (ti/ni/pu)
    slim-COP-REL-(ti/ni/pu) wire
    'slim wire'

In (29), the '*-a*' marker found at the end of all *ti/ni/pu* PCs is considered as a relative clause marker which is preceded by the null equative copula. The example (20), repeated in (30) below shows that the equative copula in TT is null.

30. ram manager
    ram manager-COP
    'ram is a manager'

As a result, I classify '*-ti/ni/pu*' PCs as Equative RCPCs. This categorization divides the PC classes in TT into two types: null PCs and RC PCs. Additionally, RC PCs are further divided into two subcategories based on the copula used within the RC. CoS RCPCs use a CoS copula, while equative RCPCs employ an equative copula. This chart is shown below.

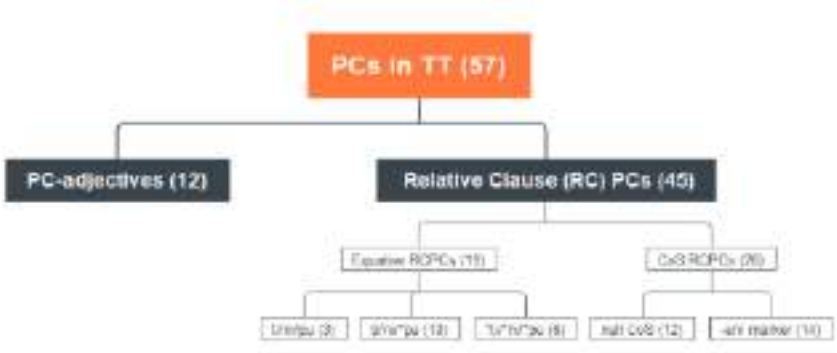


Fig 2: PC classes in TT (updated)

4. Conclusion:

This study explains the compositional nature of PC expressions in Telangana Telugu, revealing a systematic division between null PCs (true adjectives) and relative clause-based PCs (derived via copular and tense morphology). Some of the key findings in this paper include:

Class 1: Null PCs (e.g., *pedda* 'big') are restricted to core semantic types and exhibit adjectival properties, aligning with cross-linguistic patterns in (β)-type languages.

Class 2: '*-ayina*' PCs decompose into a CoS copula '***ay-***', the past tense marker '***-in***', and the relative clause marker '***-a'***, forming stage-level predicate that gives Change of State semantics. (e.g., teliv-ayina 'intelligent').

Class 3: '*-ti/-ni/-pu*' PCs (e.g., *sanna* 'slim') are analyzed as equative relative clauses with a null copula, expanding the relativization strategy to stative predicates.

The data support Amritavalli & Jayaseelan (2003) incorporation theory for Dravidian, where PC expressions are syntactically derived via nominal/verbal roots and case-marking rather than a dedicated adjective class. TT's reliance on relativization for PC formation mirrors patterns in Kannada and Malayalam, reinforcing the family's typological consistency.

Abbreviations: PC: Property Concept; TT: Telangana Telugu; CoS: Change of State; RC: Relative Clause; ILP: Individual-Level Predicate; SLP: Stage-Level Predicate

References:

Amritavalli, R, 2014, Separating tense and finiteness: anchoring in Dravidian. Natural Language & Linguistic Theory, 32 (1), 283–306

Amritavalli, R., & Jayaseelan, K. A, 2003, The genesis of syntactic categories and parametric variation. 4 , 19–41.

Baker, M. C, 2003, Lexical categories: Verbs, nouns and adjectives (Vol. 102). Cambridge University Press.

Balusu, R, 2015, Comparison, predication and lexical semantics of pc nouns in Telugu. Indian Linguistics, 76 (1-2), 1–18.

Dixon, R. M., & Aikhenvald, A. Y, 2004, Adjective classes: A cross-linguistic typology. OUP Oxford.

Haspelmath, M., et al, 2001, non-canonical marking of core arguments in European languages. Typological studies in language, 46, 53–84.

Herur, S, 2016, Expression of adjectival meaning in Kannada. Formal Approaches to South Asian Languages.

Jayaseelan, K, 2007, The argument structure of the dative construction. Argument structure, 37–48.

Krishnamurti, B., & Gwynn, S. S. B., JPL, 1987:285, A grammar of modern Telugu. Journal of the American Oriental Society, 107 (3).

Menon, M, 2012, The apparent lack of adjectival category in Malayalam and other related languages. Proceedings of GLOW-in-Asia IX, 157–171.

Menon, M., & Pancheva, R, 2014, The grammatical life of property concept roots in Malayalam. In: Proceedings of Sinn und Bedeutung (Vol. 18, pp. 289–302).

वाक् मंथन
ISSN: 2426-2149

Khasi Phonotactics & the Limitation of Classical Sonority Scale through the perspective of Beats and Binding Theory (B&B)

Ronald Kharsohtun
NEILAC
ronaldkhsohtun@gmail.com

Abstract

Sonority constraints may be examined through two complementary lenses: articulatory (production-based) and perceptual (listener-based) perspectives. From the production standpoint, sonority sequencing aids in facilitating smoother articulation by aligning jaw movement with varying sonority levels. When this alignment is disrupted, it can complicate the articulation process, making speech production more challenging. On the perceptual side, sonority modulation enhances auditory salience, which improves the recognition of speech sounds and helps in distinguishing different speech segments more easily. Khasi consonant clusters display complex voicing of 'deliberate dissimilation' (Henderson 1990). In Khasi, certain consonant clusters deviate from typical sonority sequencing principles. For instance, clusters such as *'kp', 'bt', 'bt^h-'*, and *'dk-'* are permissible in initial positions, which contrasts with the Sonority Distance Principle. According to the OSDP, obstruents like stops and fricatives generally belong to low sonority classes, while resonant consonants and vowels are classified into high sonority classes. This principle usually dictates the sequencing of consonants and vowels within syllables. The Beats and Binding (B&B) theory proposed by Katarzyna Dziubalska-Kołaczyk (2002) provides an alternative framework. The theory proposes that in initial consonant clusters, non-nuclear segments form bindings with the nuclear element, with phonotactic organisation driven primarily by perceptual contrast rather than adherence to the Sonority Sequencing Principle. Additionally, Baroni's classification system (2012) utilises Net Auditory Distance (NAD) to identify well-formed consonant clusters in Khasi. This model evaluates clusters based on manner of articulation, place of articulation, and voicing, preferring configurations that differ maximally in these features and later using salience to rank them accordingly. Acoustic cues are then employed to identify the sesquisyllables present in Khasi and true clusters, providing a comprehensive understanding of their phonological structure.

Keywords: sonority constraints, clusters, NAD, salience, sesquisyllables.

1. Introduction

Khasi belongs to a Mon-Khmer branch, an Austroasiatic (AA) language, the only one in the northeastern part of India. Other Austroasiatic languages spoken within the Indian subcontinent include Mundari and Nicobarese. Khasi, by its name, is both a tribe and a language. Despite being in close contact with several Tibeto-Burman and Indo-Aryan languages, the Khasi language has preserved its fundamental structural characteristics. Sidwell notes the lack of scholarly consensus concerning the internal classification of Austroasiatic (AA) branches, as well as debates surrounding the antiquity and internal diversity of the family (Sidwell 2010: 117). Sidwell and Blench propose that the proto-Austroasiatic migration occurred approximately 3800 years ago from the Mekong River basin—currently a region

central to major AA languages (Sidwell and Blench 2011: 338–339). Over time, speakers of AA languages became linguistically isolated due to the expansion of neighboring language groups.

The term 'Khasian' designates a subgroup within the Mon-Khmer branch of Austroasiatic languages, with Khasi as the dominant and standardised variety. The Khasi language did not possess a native orthographic tradition; instead, the Bengali script was utilised for written representation until the Roman script was formally adopted in 1841 through the efforts of the Welsh Calvinistic Mission. Post-independence, Khasi has undergone significant functional expansion, encompassing education, judiciary, media, and other formal domains. According to the 2011 Indian census, the speaker population is estimated at approximately 1.2 million, while Ethnologue (2012) records 1,431,344 speakers. UNESCO no longer classifies Khasi and Jaintia as endangered languages since 2012.

2.1 Aims and Objectives

This study aims to investigate the phonotactic structure of the Khasi language, focusing specifically on consonant clusters that defy the classical Sonority Sequencing Principle (SSP). Khasi's rich inventory of onset clusters, including configurations like /*kp*/, /*b̪t̪*/, and /*dk*/, challenges the predictions made by traditional sonority-based models. The central objective is to evaluate whether these anomalous clusters can be more effectively explained using the Beats and Binding (B&B) theory and Baroni's Net Auditory Distance (NAD) model. This research seeks to explore the perceptual basis of phonotactic legality in Khasi and to determine whether auditory salience and phonetic contrast offer a more accurate account of consonant cluster patterns than sonority hierarchies alone. This study seeks to investigate the structure and phonological function of sesquisyllables in Khasi, a feature commonly attested in Austroasiatic phonological systems.

2.2 Research Questions

Some of the research questions that will guide us during this study are;

1. To what extent do Khasi onset consonant clusters violate the classical Sonority Sequencing Principle?
2. Does the research further explore whether the Beats-and-Binding (B&B) framework and Baroni's Net Auditory Distance (NAD) model offer more explanatory adequacy in analysing Khasi consonant clusters?
3. Do perceptual salience, voicing, manner, and place of articulation are assessed in terms of their contribution to the well-formedness and acceptability of onset clusters in Khasi?
4. What role do sesquisyllables and minor syllables play in the overall phonological structure of Khasi, and how are they acoustically realized?

2.3 Research Methodology

This study adopts a descriptive-analytical approach, combining both qualitative and quantitative methods. Primary data will be drawn from phonetic transcriptions of spoken Khasi, supplemented by existing descriptions and previous research (e.g., Henderson, Rabel, Nagaraja). A corpus of onset clusters will be compiled and categorized according to their conformity or deviation from the SSP. Each cluster will be analyzed using Baroni's NAD model, assessing Manner of Articulation, Place of Articulation, and Voicing parameters. Additionally, acoustic analysis tools (e.g., PRAAT) will be employed to examine phonetic cues that indicate minor syllables and sesquisyllables in natural speech. These methods will help determine whether NAD and B&B provide a viable alternative to sonority-

based constraints in explaining Khasi phonotactics. The methodology also includes cross-linguistic comparisons with other Austroasiatic languages to contextualize Khasi's phonological patterns within a broader typological framework.

3. Existing Literature Review

The phonotactic structure of Khasi has intrigued linguists due to its unusually rich inventory of consonant clusters, many of which challenge established phonological principles like the Sonority Sequencing Principle (SSP). Traditional models struggle to account for the wide variety and irregular patterns observed in Khasi, prompting the development of alternative theoretical frameworks. Rabel (1961) was among the first to document Khasi phonology extensively, identifying over 120 initial consonant clusters, many of which defy typical cross-linguistic constraints. He described these clusters as "absolutely fantastic," especially when compared with neighboring Sino-Tibetan languages, which largely avoid such complexity. His work laid the foundation for future studies into the distinctive phonotactic behavior of Khasi. Building upon Rabel's foundational work, Henderson (1976, 1990) identified consonant clusters in Khasi that deviate from the Sonority Sequencing Principle (SSP), including sequences such as /*bt̪*/, /*dk*/, and /*pd*/, where sonority does not increase from left to right.He introduced the idea of "deliberate dissimilation," where voicing contrasts enhance segmental distinction within a cluster, potentially improving perceptual salience. These observations suggested that Khasi may prioritize perceptual cues over strict sonority constraints. Wright (2004) further supported this perceptual approach, arguing that languages may permit certain "marked" clusters if they offer sufficient internal cues for recognition. In response to such anomalies, alternative theoretical models such as the B&B framework (Dziubalska-Kołaczyk, 2002) have been developed. These models de-emphasize the syllable as the primary phonological unit and instead focus on perceptual contrast among segments. This approach aligns closely with Baroni's (2012) concept of Net Auditory Distance (NAD), which quantifies the perceptual difference between adjacent segments based on manner, place, and voicing. Clusters with high NAD scores are considered more perceptually distinct and, thus, more acceptable. Applying this model to Khasi, clusters like */kp/, /ks/,* and */bt̪/*—though problematic under SSP—are well-formed due to their strong perceptual contrast. This provides a robust explanation for Khasi's tolerance of complex and seemingly irregular clusters. We will go into details below.

4.1 Consonant clusters

Khasi onset clusters present a compelling area of inquiry, as they frequently violate the SSP. According to Wright (2004), consonants that are not flanked by vowels, liquids, or glides referred to as "stranded consonants" tend to be dispreferred in phonotactic structure unless they possess strong internal acoustic cues capable of sustaining perceptual salience in the absence of formant transitions. Consonant types that are more likely to be perceptually robust in such contexts include sibilant fricatives, other fricatives to a lesser extent, and nasals. Some of the consonant clusters found in Khasi in the onset position are:

/p + t̪,d,n, ɟ,l,r,s/ ,

/pʰ + n,ŋ,l,r /,

/ b + t̪, t̪ʰ,n,s,l,r,ʃ/ ,

/ t̪+ b,d,m,n,ŋ,l,r,w/,

/ t̪ʰ + m,ŋ,n,l,r,w/ ,

/d +p,k,kʰ,ŋ/ ,

/k + p,b, t̪, t̪^{h},d ɟ,s,ʃ,m,n,l,r,w,j/ ,

/ k^{h} + m,n,ɲ,l,r,w/ ,

/ ɟ^{h} + r,l,n/ ,

/s + p,b, t̪,d,k,k^{h},m,n,ŋ,l,w,ʔ/,

/ ʃ + t̪,d,k,ʔ,n,ɲ,r,ŋ,l/ ,

/ m + t̪,l,r/ ,

/l + p,b, t̪,d,k,k^{h},ʔ,h,m,ŋ,w/ ,

/r + b, t̪,k^{h},m,n,ŋ,h,w/

We find that most consonant clusters happen at initial positions and in the medial positions. The final positions are relatively fewer. However, there are no consonant clusters with the sounds / n,h,w/ as in /nh, hn, hw, wh, nw, wn/.

- Only /*p^{h}, t̪^{h}, k^{h}*/ are seen occurring at the initial position when combined with liquids and nasals.
- Aspirated consonants are relatively rare in the second position of consonant clusters, indicating a marked distributional constraint within Khasi phonotactics.
- Nasals are rarely found as the initial consonant in clusters, with the exception of /m/, which commonly precedes liquids.
- Some consonant clusters at the onset go against the sonority hierarchy /*pt̪-, bt̪-, t̪b-*/

Prior research by Henderson (1976a), Nagaraja (1985, 1990), Khriem (2013), and others has substantially contributed to our understanding of Khasi cluster patterns. Jenny, Weber, and Weymuth (2014) observe: “Most syllables are closed, and there are many combinations of onset consonants... Some of them go against the sonority hierarchy... Dental+alveolar clusters such as /*t̪n-, t̪r-, t̪l-, t̪^{h}n-, t̪^{h}r-, t̪^{h}l*/ do occur, however” (Jenny, Mathias & Sidwell, Paul 2014).

Although Khasi displays a remarkable array of consonant clusters, this phenomenon is not exclusive to the language. Similar patterns are observed across various languages, as Henderson (1976: 523) notes. One of Khasi's most typologically salient features is the extensive range of word-initial clusters. This pattern aligns more closely with other Austroasiatic languages such as Khmer and Old Mon than with neighbouring Sino-Tibetan languages, which typically restrict initial clusters to plosive–liquid or plosive–semivowel combinations. Rabel (1961: 21–29) documents 127 distinct two-consonant onset clusters and refers to them as “absolutely phantastic” (Rabel 1973: 15).

Khasi displays the possible onset clusters:

(1) the possible combination of clusters in Khasi:

(a) /p/ + Coronal (voiced/voiceless)

(b) /p^{h}/+Coronal [+ Nasals & above]

(c) /b/ + coronal

(d) / t̪/ + Labial, Coronal(voice/voiceless)

(e) / t̪ʰ/ + Nasals & above

(f) /d/ + Labial(voiceless), coronal

(g) /k/ + Labial, Coronal

(h) /ɟ/ + Nasals & above

(i) /s,l,r/ + Labials (voiced/voiceless), Coronals, dorsals, nasals.

(2) the following combination of clusters are not possible in Khasi.

(a) *Labial + glottal

(b) *Labial + Labial

(c) *Labial + Dorsal

The data strongly suggest that Khasi onset clusters do not uniformly conform to the Sonority Sequencing Principle.

4.2 Sonority sequencing principle

Sonority-based constraints, as developed by Frisch (2004), nonetheless remain instrumental in evaluating the phonotactic structure of consonant clusters. These constraints suggest that certain restrictions on the manner of articulation exist within consonant clusters, with traditional categorizations placing obstruents (stops and fricatives) in low sonority classes, while resonant consonants and vowels fall into high sonority classesThe classification and number of sonority sub-classes differ across scholarly frameworks, reflecting varying theoretical approaches to sonority hierarchies. The preference for consonant clusters that maximize sonority differences extends to vowels, and this sonority classification can potentially elucidate common patterns in syllable structure.

Two overarching perspectives exist in understanding the functionality of sonority constraints: the production side and the perception side. From a production standpoint, sonority sequencing is seen as a result of the ease of articulation when organizing speech into units. This involves a cyclical jaw movement, transitioning from low sonority sounds (jaw closing or raised for obstruents) to high sonority sounds (jaw opening or lowered for vowels). Violations of this sequencing involve challenging and precise jaw movements or can lead to a reorganization of suprasegmental sound structure.

On the perception side, sonority modulation is posited to enhance the perception of speech sounds and phonological structure. Alterations in amplitude and dominant frequency spectra contribute to the easier identification of speech segments and syllabic structure. Research indicates that vowel and consonant information can be recovered from different portions of the speech signal, making speech more robust in noisy environments.

Building upon these perspectives, specific quantitative hypotheses can be formulated. For sonority sequencing, consonant cluster types with larger steps in sonority difference toward the peak of the syllable are expected to be preferred. In the context of sonority modulation, consonant clusters characterized by greater sonority differentials are generally predicted to be more acceptable, regardless of their linear order. As Frisch argues, clusters exhibiting larger sonority steps tend to be phonotactically preferred. From a functionalist perspective, this view posits that the phonotactic well-formedness of a given structure correlates with its frequency of occurrence in a language. However, if frequency does not systematically align with the degree of sonority difference within a language, the functional quantitative model's explanatory power is called into question (Frisch 2015: 27(1), 9-27).

4.3 Sonority of consonant clusters

Khasi allows certain consonant clusters in the initial positions which are generally not seen in many languages and which violates the sonority principles mentioned below,

plosives < fricatives < nasals< laterals < flaps < high vowels < mid vowels < low vowels

Although clusters of voiced + voiceless stops are found to be common where the voiced consonant clusters are closer to the nucleus/vowel according to Greenberg's generalisation, but Khasi displays initial clusters like /*bt-, bth-, bs-, bsh-* [bʃ], *dk, dkh-, dp-/* as well.

Henderson (1989–90: 62) observed a systematic voicing dissimilation within stop–stop clusters in Khasi. For instance, alongside sequences such as /bt-/, /dk-/, and /dp-/, we also find their voiced counterparts /tb-/, /kd-/, and /pd-/. Clusters like /pt-/ and /tk-/ are largely confined to onomatopoeic or expressive lexemes.

Table 1: Voicing dissimilation among Khasi clusters

bt̪- bt̪ɛpˀ	'to cover with the earth'
sb - sbɔʔ	'fertile'
bs- bsa	'feed'
dk - dkɔʔ	'cripple'
pd - pdɔt̪ˀ	'Adam's apple'
kd - kdɛu	'to point out with a finger'
dp - dpɛi	'ash'
t̪b – t̪ba	'to search; feel'

Greenberg generalizations no. 21 highlights that the only permissible sequence of voiced+voiceless consonant is to have voiced nasal+ homorganic unvoiced obstruent. However, Henderson once again showed examples of Khasi clusters like *mt- mt̪ain* 'to drift' and *mt̪h- mt̪hin* 'strong and stout', *mt̪-mt̪ung* 'in a large heap', though these words might be considered 'expressives.'

4.4 Distribution of velar stops

Khasi does not have a voiced velar */g/* and therefore, it is replaced with */k/*. As a result we see clusters as; *k+ p,b,t,t^h,ɟ,s, ʃ*. Thus, there are sequences of *k*+voiceless stop as in *kp-kpa* 'father', and *k*+voiced stop as in *kb-kba* 'wheat'.

Table 2: Clusters of velar stops

kp - kpa	'father'
kb - kba	'paddy'
kt̪- kt̪i	'hand'
kt̪^{h} - kt̪^{h}a	'chew'

kɟ - kɟaʈ̪	'leg'
ks - ksuiʈ̪	'pus'

Though the clusters provided above violate the sonority sequencing principle as noted, there are yet permissible combinations in Khasi. To account for these clusters, we use Baroni's classification of possible clusters in terms of perceptual salience.

4.5 Baroni's classification

Baroni (2012) offers an alternative phonological model—the Beats and Binding (B&B) theory—originally introduced by Dziubalska-Kołaczyk (2002, 2009). This model dispenses with the syllable as a central unit in favour of a perceptually grounded phonotactic framework. Within the B&B framework, phonotactic preferences are governed by principles of perceptual salience and contrast, operationalised through the Net Auditory Distance (NAD) metric, which integrates MoA, PoA, and Voicing. High NAD values, corresponding to greater perceptual distinction between adjacent segments, are associated with greater phonotactic acceptability (Dziubalska-Kołaczyk 2009: 55; Baroni (2012, p.47). Within the Beats and Binding (B&B) framework, the "beat" corresponds to the syllabic nucleus, while "non-beat" elements refer to non-nuclear segments. The interaction between these elements is termed a *binding*. Phonotactic constraints in this model are governed by the concept of Net Auditory Distance (NAD), which evaluates contrast across three phonetic dimensions: Manner of Articulation (MoA), Place of Articulation (PoA), and Voicing. A core principle of the theory asserts that segmental sequences are phonotactically more well-formed when adjacent consonants exhibit maximal disparity along these dimensions, thereby yielding a higher NAD score. For instance, sequences such as /brV/ or /grV/ are considered more optimal than /drV/ due to greater place of articulation contrast.

Any sequence C1C2V is well-formed iff NAD C1C2 ≥ NAD C2V

Baroni assigns the following values for the different consonants as per their POA and MOA (Baroni, 2012, p.49).

Table 3: Values for MoA and PoA in NAD (Baroni, A. 2012: 49)

4			3	2		1	0	**MOA**		
Obstruents			**Sonorant**							
Stop	Affricate	Fricative	Nasal	Liquid		Glide	Vowels		**POA**	
				Lat	Rh					
p b	pɸ bβ	ɸ β	m			ɥ	y	Bilabial	**Labial**	3
	pf bv	f v	ɱ			ʊ		Labio-dental		
t d	ts dz tθ dð	s z θ ð	n	l	r		a	Dental Alveolar	**Coronal**	0
c ɟ	cç ɟʝ	ç ʝ	ɲ	ʎ		j	i	Post-alveolar/ Palatal		1
k g	kx gɣ	x ɣ	ŋ			ɯɰ	ɯ	Velar	**Dorsal**	2
						w	u	Labio-velar		2.5

As per this table if we were to examine the onset clusters in Khasi, these are the generalisations that we can make: Assuming the vowel to be uniformly /i/.

1. Aspirated stops */p^h, b^h and j^h/* (labial+palatal) occur only with coronals, nasals and liquids. This combination of C1C2 would have the values C1 = (4,3,0) and C2 = (3,0,1); for C2V, C2 (3,0,1) and V (0, 1, 1)

So, C1C2 (4-3, 3-0,0-1)= 1+3+1=5

C2V= (3-0, 0-1, 1-1)=3+1+0=4

So, NAD C1C2 is ≥ NAD C2V and hence, a possible combination word initially.

The other possible clusters are:

2. Voiceless or voiced labial stops */p,b/* occur only with voiceless /voiced coronal (stops, nasals and liquids).

3. Voiceless dorsal stop */k/* occurs only with coronal or labial stops.

4. Voiceless coronal stops */t̪,d/* occur with labial stops, coronal stops, nasals, and liquids

Consonant Cluster Types	**C1**	**C2**	**NAD C1C2**	**C2**	**V**	**NAD FOR C2V**	**NAD C1C2 And NAD C2V**	**Well formed- Yes/No**
Type 1 /p^h, b^h and j^h/+ coronal obstruents, nasals and liquids	(4,3, 0)	(3,0, 1)	(4-3, 3-0, 0-1) =5	(3,0,1)	(0,1, 1)	(3-0, 0-1, 1-1)=4	NAD C1C2> NAD C2V	Yes
Type 2 /p,b/+ vl or vd coronals (stops, nasals and liquids)	(4,3, 0/1)	(2-4, 0, 0/1)	(4-2 to 4, 0 to 1-0, 1-0 to 1)= 0 to 2+0 to 1, 0 to 1= 2-4	(2-4, 0, 0/1)	(0,1, 1)	2 to 4-0, 0-1, 0 to 1-1= 2-4	NAD C1C2= NAD C2V	Yes
Type 3 /k/+coronal or labial stops	(4,2, 0)	(4,0/ 3, 0)	(4-4, 2-0/3, 0-0)= 0+2 to 5+0= 2 to 5	(4,0/3, 0)	(0,1, 1)	(4-0, 0 to 3-1, 0-1)= 4+1 to 2+0 to 1= 5 to 7	NAD C1C2 < NAD C2V	No

Type 4 / t̪,d/+ labial stops, coronal stops, nasals, liquids	(4,0, 0 or 1)	(4 to 2, 0, 0 or 1)	(4-2 to 4, 0 to 1-0, 0 to 1-1) = 2 to 4+0 to 1+0 to 1)= 2 to 6	(4 to 2, 0, 0 or 1)	(0,1, 1)	(2 to 4-0, 0-0 to 1, 0 to 1-1) =2 to 4+0 to 1+0 to 1= 2 to 8	NAD C1C2 < NAD C2V	No
Type 5 /s,l,r/+vl and vd. labial stops, coronal stops, dorsal stops, nasals, liquids and glides	(2 to 4, 0, 0 to 1)	(1 to 4, 0 to 3, 0 to 1)	(2 to 4- 1 to 4, 0-0 to 3, 0 to 1 -0 to 1)=	(1 to 4, 0 to 3, 0 to 1)	(0,1, 1)	(1 to 4 -0, 0 to 3 -1, 0 to 1 -1) = 3 to 8	NAD C1C2 < NAD C2V	No

5. The three coronal sounds */s,l,r/* occur with voiceless and voiced labial stops, coronal stops, dorsal stops, nasals, liquids and glides. */s/* C1 = (4,0,0),C2= (4,0,0), */l,r/* C2= (2,0,1) and also followed by (4/3,0-4,0/1)

Table 4 (above): Illustration of NAD for each cluster.

As per the values given above for Types 1-2, C1C2 is well formed as the MoA, PoA, Lx values of C1 is greater than the MoA, PoA, Lx values of C2.

However, in the case of Types 3-5, the values of MoA, PoA, Lx of C1 is lesser than the MoA, PoA, Lx values of C2.

4.6 Plateau

Apart from the notion of NAD, Baroni uses the notion of salience to resolve issues of a plateau. According to Baroni, an optimal consonant sequence is typically characterized by a rise in sonority, defined as the degree of vocal tract openness, preceding the beat, followed by a decline thereafter. When adjacent consonants fail to contrast in sonority, this results in what is termed a sonority or manner plateau. Similarly, if no distinction exists between adjacent segments in terms of place of articulation (PoA) or voicing—assuming their manner is identical—these are referred to as PoA and voicing plateaus, respectively (Baroni 2012: 53). More broadly, the notion of a *plateau* signifies a lack of phonological contrast at a particular level. When there are plateaus, there are some plateaus that are better than others.

For instance, for Baroni, "a sequence of two obstruents is a plateau" (Baroni,2012:53). According to SH (Sonority Hierarchy) and B&B (Beats and Binding) FS (Fricative+ Stop) is not any better than SF (Stop+ Fricative). This refers to sequences like */sp/, /ps/* and */tp/* are possible but we know that */sp/* has more chances to occur in many languages than */ps/*. This can be accounted for in terms of **salience.**

Salience relies on acoustic factors. Salience is acoustically defined as the extent to which a non-beat segment's manner and place of articulation can be independently perceived in relation to its distance from the beat (Baroni 2012: 55).

4.7 Salience scales

Salience hierarchies relevant to obstruents include:

- Fricatives > Stops
- Within Stops: /k/ > /p/ > /t/
- Within Fricatives: /s/ > /f/ > /θ/
- Nasals: /m/ > /n/ > /ŋ/
- Liquids: /r/ > /l/

(cf. Ladefoged & Maddieson 1986; Hume et al. 1999; Wright 2004; Jun 2005; Greenlee & Ohala 1980; van der Hulst 2004)

The notion of salience accounts for Types 3 and 5 in Khasi, namely */k/*+ a labial or coronal stop (as */k/* is the most salient of all stops), */s/* followed by stops, nasals, liquids as */s/* as a strident is more sonorous than other obstruents. The only cases that are difficult to account for are Type 4 (i.e., */t̪,d/* followed by a labial or coronal stop and */l,r/* followed by voiceless or voiced stops, and nasals.

5.1 Minor syllables

Minor syllables are common among the AA languages. They are a reduced form of a vowel. The occurrence of minor syllables in Khasi is quite frequent. They are usually combined with major syllables. The first syllable is usually shorter and less intense in a disyllabic word compared to the second syllable and the vowel on the second syllable tends to be heavier or longer. We find a combination of one minor syllable with one major syllable or sometimes two minor syllables with one major syllable. Rabel (1961) offers a series of exemplifications that substantiate his theoretical perspective.

Table 5: List of minor syllabic words in Khasi

[hⁱn-nin]	minor+major	'INDEF.yesterday'
[mⁱn-hⁱn-nin]	minor+minor+major	'DEF.yesterday'
[kⁱnmaw]	minor+major	'remember'
[pⁱnkⁱnmaw]	minor+minor+major	'to remind'

Anderson discusses the sesquisyllabic structure characteristic of many Austroasiatic languages, including Khasi, in which prosodic words consist of minor-major syllable sequences (Anderson 2014: 378–379).

5.2 Sesquisyllable

The concept of sesquisyllable derives from the Latin prefix *sesqui-* meaning "one and a half" and denotes a phonological structure characterized by the reduction of a minor syllable—often realized as /Cɨ/ or /CɨN/—preceding a stressed tonic syllable. In the Khasi language, this sesquisyllabic structure can be observed in examples such as *kyndon* /kn̩dɔːn/ meaning 'rule', *syrwet* /sr̩wɛ̄t/ meaning 'sign', *kylla* /kl̩la/ meaning 'transform', *symboh* /sm̩bɔːʔ/ meaning 'seed', and *tyngkai* /t̪ŋ̩kaːi/ meaning 'conserve.'

Since the length of a syllable can vary depending on the language, the speed of speech, and individual pronunciation, there's no fixed duration for a sesquisyllable in seconds. However, in a general context, if we consider the average duration of a syllable in normal speech (which can range from 0.1 to 0.25 seconds), a sesquisyllable could roughly range from 0.15 to 0.375 seconds. This is an approximation and would depend on the specific speech context. American linguist James Matisoff (1973) termed the iambic pattern that Khasi follows is sometimes called sesquisyllabic. Thomas attempted at typologizing sesquisyllabic languages in 1992. The */ɨ/* in Khasi replaces the centralised vowel */ə/* (Jenny, Mathias & Sidwell,P 2014: 1148).

Khasi lacks the central vowel */ə/* and instead features */ɨ/* as its sole mid-high central vowel. Notably, all vowels in Khasi, with the exception of */ɨ/,* can be both long and short (Jenny, Weber & Weymuth 2014: 31). The short vowel */ɨ/,* represented orthographically as '*y*', serves as a junctural element within initial clusters whose second consonant is neither a glottal stop nor a frictionless continuant (The Handbook of Austroasiatic Languages: 1148). A PRAAT analysis (Boersma & Weenink, 2013) was conducted to classify them as sesquisyllables. As observed, in the word *kyndon* 'steps', the duration of */ɨ/* is notably shorter in comparison to */ɔ:/.* This is also because Khasi has an iambic stress pattern. Two other examples of the words *pynkynmaw* ' to cause someone to remember' and *symboh* 'seed' are given below.

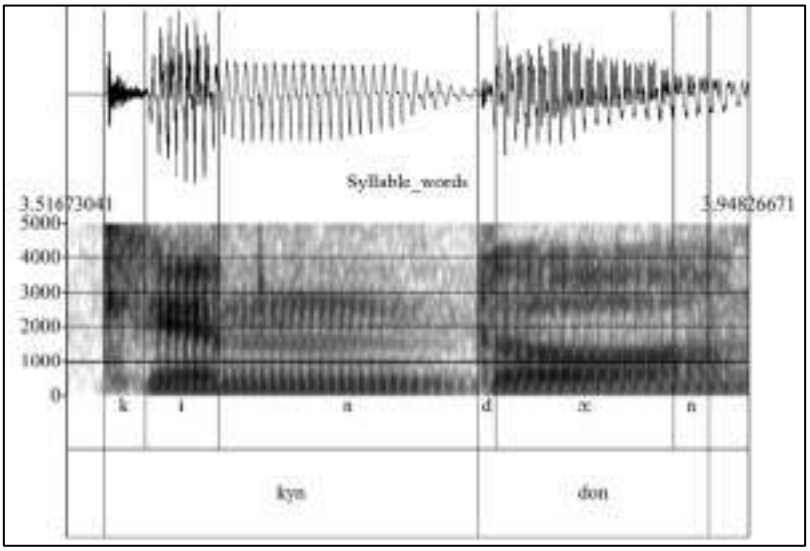


Figure 1: Illustration of 'kyndon' to show a sesquisyllable /ɨ/ using PRAAT software.

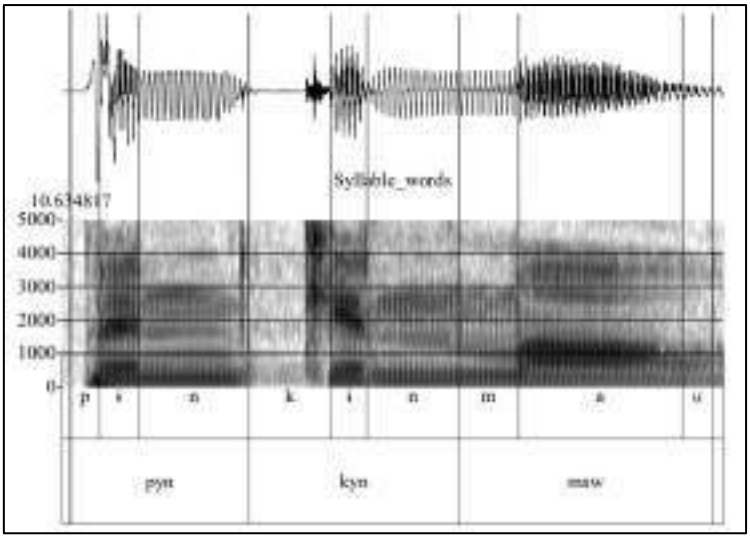


Figure 2: Illustration of 'pynkynmaw' to show a sesquisyllable /ɨ/ using PRAAT software.

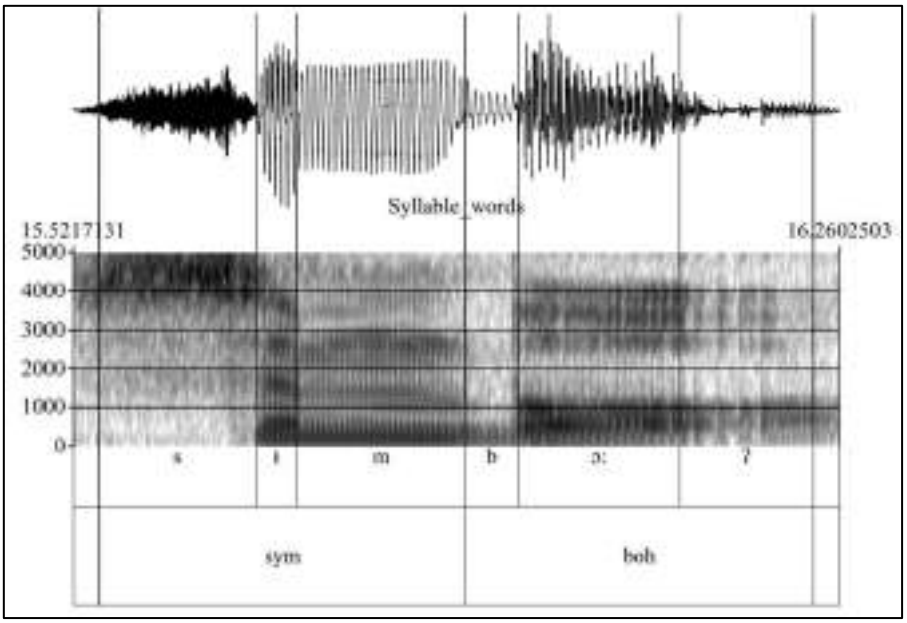


Figure 3: Illustration of 'symboh' to show a sesquisyllable /ɨ/ using PRAAT software.

The traditional classification of clusters by Henderson using Frøkjær-Jensen combined oscilloscope and mingograph, with attachments, had identified clusters belonging to those categories that are true clusters and those with sesquisyllables (See Henderson 1976).

In this study, we will utilise PRAAT analysis to identify the true clusters in Khasi.

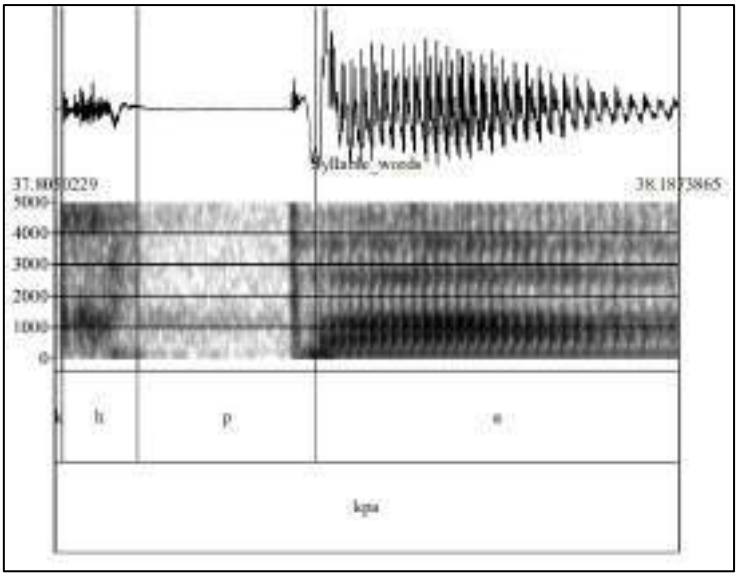


Figure 4: Illustration of true clusters 'kpa' using PRAAT software.

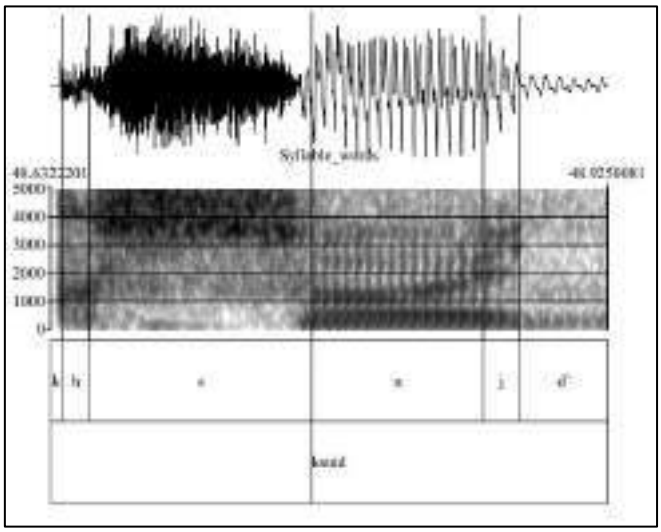


Figure 5: Illustration of true clusters 'ksuid' using PRAAT software.

Here are the results after having observed through the words using PRAAT software.

Acoustic analysis using PRAAT of items such as kpa 'father' and ksuid 'devil' reveals no intervening vowel between the initial consonants. However, in clusters problematic under both SSP and B&B theories (Type 4), a short vowel appears in the first syllable, thereby classifying them as sesquisyllabic rather than true clusters.

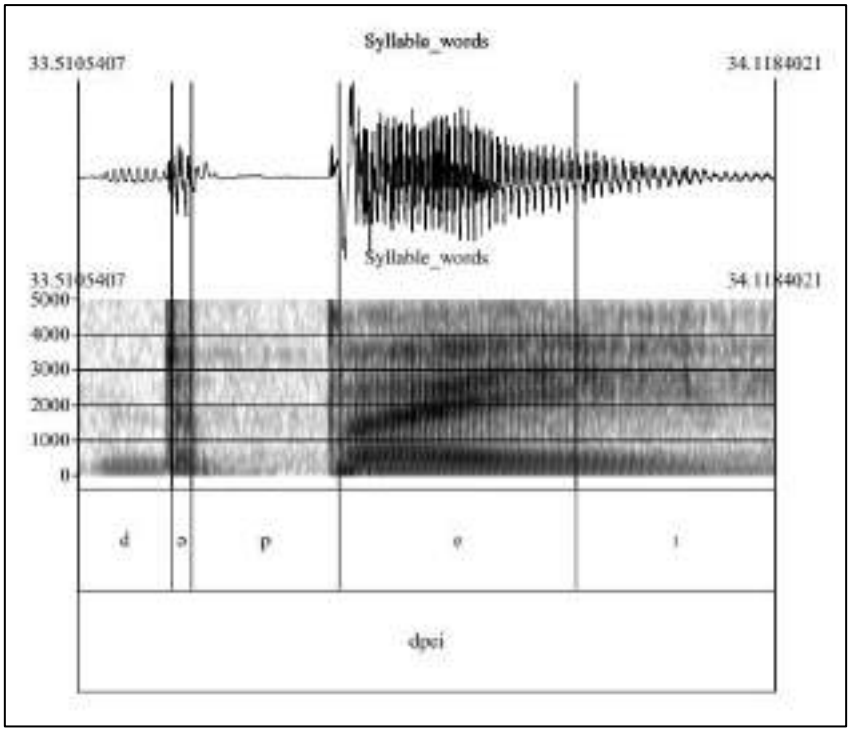


Figure 6: Type – 4 syllable 'dpei'

Thus, we have seen that the occurrence of onset clusters in Khasi cannot be accounted for using the Sonority sequencing principle. However, Baroni's classification of plausible clusters in terms of NAD and salience is better equipped to account for the possibilities in Khasi. Apart from the consonant clusters discussed in this section, which are traditionally considered as 'true' clusters in Khasi grammar, Khasi also has a number of sesquisyllables, which are a combination of a minor + major syllable.

6. Findings and Conclusion

This study has examined the phonotactic structure of Khasi through the lens of Beats and Binding (B&B) Theory, challenging the explanatory adequacy of the Classical Sonority Scale (CSS). The Khasi language, with its complex onset clusters and allowance for sonority-violating sequences, presents a particularly compelling case against the universality of CSS-based accounts. By demonstrating that CSS fails to consistently predict permissible onset structures in Khasi, the analysis highlights the limitations of a strictly scalar approach to sonority and syllable well-formedness.

Central to this inquiry was the observation that in Khasi, consonant clusters such as /*pn*/, /*kn*/, and /*pl*/—which should be marked or ill-formed under CSS—occur naturally and frequently, especially in word-initial positions. This challenges the CSS premise that syllable onsets must rise in sonority, exposing the inadequacy of this scale to accommodate language-specific phonotactic preferences and morphophonemic constraints. In contrast, the B&B Theory, which foregrounds rhythmic and prosodic binding relations over hierarchical sonority sequencing, offers a more flexible and explanatory framework. By viewing permissible clusters through the lens of binding domains and rhythmic licensing, B&B accommodates sequences that are anomalous under CSS but phonetically and prosodically plausible within Khasi. Further, this analysis reaffirms that phonotactic well-formedness cannot be entirely captured by abstract scalar hierarchies. Instead, it must consider language-specific morphophonological rules, rhythmic timing, and auditory-perceptual salience. The Khasi data suggests that linguistic systems may encode preference patterns that diverge significantly from cross-linguistic norms yet remain internally coherent. This calls into question the presumed universality of CSS and

aligns with a growing body of research arguing for more empirically grounded, typologically sensitive models of syllable structure. Moreover, the application of B&B Theory to Khasi phonotactics opens a promising pathway for revisiting phonological structures in other understudied or typologically atypical languages. This study suggests that languages with apparent violations of CSS may be more systematically understood through an approach that prioritizes prosodic timing and perceptual coherence rather than strict sonority sequencing.

While this study provides significant insights, several avenues for further investigation remain. First, a more extensive phonetic analysis of Khasi onset clusters could provide deeper insight into how rhythmic timing and auditory salience shape phonotactic preferences. Acoustic studies focusing on timing, duration, and intensity of consonant sequences in natural speech would enhance our understanding of the perceptual cues that govern binding relations.

The applicability of B&B Theory across different dialects of Khasi, or within other Austroasiatic languages, merits exploration. Comparative studies could help determine whether the patterns identified here are unique to Khasi or reflect broader regional or genetic phonotactic tendencies. This would contribute to developing a more robust typology of onset licensing patterns that better accounts for variation outside Indo-European or well-documented phonological systems. A computational implementation of B&B Theory, perhaps within an Optimality-Theoretic or constraint-based framework, could formalize the binding relations proposed and test their predictive power against corpus data. This would aid in determining how binding constraints interact with other phonological processes, such as stress assignment, vowel reduction, or morphophonemic alternation. Lastly, the role of language acquisition and processing in the internalization of non-sonority-based phonotactics in Khasi remains an open question. Psycholinguistic studies with native speakers—particularly children acquiring Khasi—could reveal whether prosodic binding is cognitively more salient than sonority scaling during early language development.

In short, the Khasi language not only challenges the sufficiency of classical phonological models like CSS but also demonstrates the value of alternative frameworks like B&B that prioritize rhythmic and perceptual structures. Future research should continue to expand these models across diverse linguistic data to better understand the full range of human phonological systems.

List of abbreviations

NAD : Net Auditory Distance
CSS : Classical Sonority Scale
B&B : Beats and Binding
AA: Austroasiatic
OSDP: Optimal Sonority Distance Principle
PoA: place of articulation
MoA : Manner of articulation
SH: Sonority Hierarchy
SF : Stop + Fricative
C1C2: Consonant 1, Consonant 2
SSP: Sonority Sequencing Principle

References

Dziubalska-Kołaczyk, K. 2002. *Beats-and-Binding Phonology*. Germany: Peter Lang, Ch. 4, 50-80.

Jenny, M. & Sidwell, P. (Ed (s).) 2014. *The Handbook of Austroasiatic Languages*, Vol. 1. Leiden: Brill.13–220, 1145–1185.

Nagaraja, K. S. 1985. *Khasi: A Descriptive Analysis*. Pune: Deccan College Research Institute.

Nagaraja, K. S. 1990. *Khasi Phonetic Reader*. Phonetic Reader Series-29. Mysore: CIIL.

Rabel, L. 1961. *Khasi, A Language of Assam*. Baton Rouge: Louisiana State University Press.

Henderson, E. J. A. (1976). *Khasi Initial Clusters*. In P. Jenner, L. Thompson, & S. Starosta (Eds.), *Austroasiatic Studies*: Part I Honolulu, HI: University of Hawai'i Press, 523–538.

Morelli, F. 1999. *The Phonotactics and Phonology of Obstruent Clusters in Optimality Theory*. Unpublished doctoral dissertation, University of Maryland at College Park.

Bareh, C. 2013. *Languages of North East India: A Comparative and Contrastive Analysis on Some of the Phonological Features*. *International Journal of Linguistics*, 5, ISSN 1948-5425.

Frisch, S. A. 2015. *A Preliminary Investigation of Quantitative Patterns in Sonority Sequencing*. *Italian Journal of Linguistics*, 27/1, 9–27.

Henderson, E. J. A. 1989–1990. *Khasi Clusters and Greenberg's Universals*. *The Mon-Khmer Studies Journal*. 18–19, 61–66.

Nagaraja, K. S. 1990. *Sounds of Khasi Dialects*. *Bulletin of the Deccan College Research Institute*. Vol. 50, 291–296.

Rabel, L. H. 1976. *Analysis of Loanwords in Khasi. Oceanic Linguistics Special Publications*, 13, Honolulu, HI: University of Hawai'i Press, 971–1034.

Baroni, A. 2012. *A Beats-and-Bindings Account of Italian Phonotactics*. *Padua Working Papers in Linguistics*, 5, Padova: Università di Padova. ISSN 1827-0522, 45–72.

Baroni, A. 2013. *The Phonotactics of Plateau Clusters: Typological and Experimental Evidence. International Workshop on Phonotactics and Phonotactic Modelling*, Scuola Normale Superiore, Pisa, November 21–22.

Misra, M., Lyngdoh, S. & Khyriem, B. 2022. *Morphological Analysis (affixes) of Aesop's Fables*. *Quest Journals Journal of Research in Humanities and Social Science,* Volume 10 ~ Issue 3 ISSN(Online): 2321-9467, 27-41.

वाक् मंथन
ISSN: 2426-2149

Negative Verbal Clitics and Negative Copula in Kumaoni

Moomal Majee
JNU, Delhi
moomal02majee@gmail.com

Abstract

This paper aims to examine how negation operates in the Kumaoni language, and how it differs from other Indo-Aryan languages. Negation can be defined as a basic operation in languages in order to reverse the truth value of a proposition. Each language has its own specific ways of negating the truth of an utterance. Kumaoni performs sentential negation with the negative particle */nə/, /ni/* or */nɛ/*, which can be semantically translated into the English equivalent 'not'. However, verbs in Kumaoni also display a special negative verbal clitic in certain conditions. The copula is used in Indo-Aryan languages for denoting existential and possessive constructions. These constitute a separate category of negation in Kumaoni because there is a special suppletive negative copular verb which shows agreement in terms of number and honorificity. This study attempts to identify the exact cases where these elements appear, with illustrative examples. It also attempts to make a comparison with some neighbouring and related languages like Garhwali and Nepali to gain a better understanding of the phenomena. The paper thus looks at some interesting features of negation in Kumaoni which distinguish it from its sister languages and indicate at a rich linguistic inventory.

Keywords: negative verbal clitic, negative copula, negation, Kumaoni, comparison

1 Introduction

Negation in languages has been a rather fertile area of research. Languages across the world perform negation with various devices, be it lexical, phonological, morphological or syntactic. The focus of this paper is how Kumaoni treats this phenomenon of negation somewhat differently from its related languages. Belonging to the Indo-Aryan family, it is a non-Scheduled Central Pahari language spoken in the Kumaon region in the state of Uttarakhand, and in parts of Nepal. It has been attested in Sir George Grierson's *Linguistic Survey of India, IX(IV),* in the volume on *Indo-Aryan Family: Central Group: Specimens of the Pahari Languages and Gujuri* (1916). It is presently dominated by the overarching influence of Hindi.

2.1 Aims and Objectives

The paper aims to examine the uniqueness of the operations of negation in the Kumaoni language. Kumaoni shows evidence of having a negative verbal clitic and a negative copula for constructions in the present tense. The study attempts to examine the conditions of occurrence for this clitic and copula with appropriate examples, and compare it with neighbouring and related languages too.

2.2 Research Questions

The issues under investigation are the negative verbal clitic and the negative copula in Kumaoni. The study tries to answer questions like exactly where and in which contexts they occur, and also explores their respective agreements patterns. Meanwhile, similar constructions in neighbouring languages are kept in mind for purposes of comparison and gaining a better understanding of potential contact phenomena with respect to these features.

2.3 Research Methodology

The data for this paper has been acquired from a field visit to Almora in Uttarakhand during last December, as part of the coursework for my ongoing degree. A detailed questionnaire had been prepared to test out various patterns of negation and the occurrence of negative polarity items in the Kumaoni language, and data was collected on that basis from 17 willing native speakers in Almora and some adjacent semi-urban and rural areas. The sentences were spoken out in Hindi or English, whichever was deemed comfortable, and the speakers were asked to translate them into Kumaoni. Owing to the constraints of time and resources, the convenience sampling method was used. However, this included participants from different socio-economic backgrounds, ages (18-70) and geographic mobility. This data was further cross-checked multiple times. Some data on Garhwali and Nepali have also been collected from one native speaker each, for the purpose of comparison. Then the data was analysed in close detail. It is to be noted that all the transcriptions have been done using the modified IPA version by Bloch and Trager (1942).

3 Literature Review

Negation can be defined as a basic operation in languages in order to reverse the truth value of a proposition. Each language has its own specific ways of negating the truth of an utterance, be it a lexical, morphological, phonological, syntactic or pragmatic device. Symmetric structures are those which show no structural differences between the affirmative and the negative constructions other than the addition of the negative marker(s) in negation. Due to the presence of a negative verbal clitic, the negative marking in Kumaoni has an asymmetrical structure, according to Miestamo's framework (2005). This can be compared with a neighbouring language like Nepali to explore the extent of this asymmetry, which Sharma calls a 'morphological peculiarity' (1987:120) of Kumaoni. He

elaborates that all the three tenses have a full-fledged negative conjugation corresponding to the affirmative conjugation, e.g. in the present tense, the negative conjugation is the use of the durative participle /-n-/ and all flectional markers are affixed to this base. Moreover, he finds less free variation in the negative conjugation, as opposed to the affirmative, except for the negative copula.

The copula is a verb that connects the subject with a predicate nominal or adjective. Many languages of the world have a different negative copula that is used only in negative copular, existential, and/ or possessive constructions. For example, Tagalog and most Austronesian languages distinguish between plain negatives and negatives of existence (Payne 1997). Iraqi Arabic employs different particles in verbal and verbless predicates (similar to copular constructions). Negative copula can also be found in Mongolian, Ladakhi, Duan Nu, Malayalam, Bangla etc.

The review of existing literature shows that there has been work in negation in many languages which display varying patterns of negative markers, auxiliaries and particles. However, Kumaoni as such has not been worked on so much, hence, the use of the negative verbal clitic as well as the negative copula appear to be fascinating traits of the language, as opposed to its neighbouring languages.

4 Discussion and Data Analysis

4.1 Negative Verbal Clitic

Sentential negation in Indo-Aryan languages is usually done with the help of a negative particle. This holds true for Kumaoni as well, which uses the negative particle */ni/, /nə/* and */nɛ/*, usually before the verb. However, the verbs in Kumaoni have a somewhat different form in a negative sentence, i.e. a special negative marker is added to the verb. For reasons that will be explained in the further sections, this marker is called a negative verbal clitic. Examples of this can be seen from the following pairs of affirmative-negative sentences, to demonstrate the change in the verb form.

1	ɾam-kɛ̃	yə	šaɡ	bʰəl	lɑg-ɔ
	Ram-3MS-Dat	this-Adj	vegetable-3S-Nom	good-Adj	seem-3S-Pres-Indf
	Ram likes this vegetable.				

2	rɑm-kɛ̃	yə	šɑg	bʰəl	nə	lɑg-ən
	Ram-3MS-Dat	this-Adj	vegetable-3S-Nom	good-Adj	not-Neg	seem-Neg-3S-Pres-Indf
	Ram does not like this vegetable.					

3	həri	məstə	kɑ̄m-ø	kər-ũ
	Hari-3MS-Nom	much-Adj	work-3S-Acc	do-3MS-Pres-Indf
	Hari works a lot.			

4	həri	bɪlkʊl	kɑ̄m-ø	ni	kər-ən
	Hari-3MS-Nom	absolutely-Adv	work-3S-Acc	not-Neg	do-Neg-3MS-Pres-Indf
	Hari does not work at all.				

In the present indefinite affirmative sentence (1), the verb */lag-ɔ/* ends at */-ɔ/,* and in (2), when the exact same sentence is negated with */nə/*, this verb ending changes into */-ən/*. Similarly, the */-ũ/* of (3) changes into */-ən/* in the negative version (4). Therefore, in the present indefinite, the verb contains the morpheme */-ən/* to become semantically negative, hence it is a clitic. Further evidence of this clitic morpheme follows, especially with varying the number and gender of the grammatical subject of the sentences, since it has been observed that Kumaoni verb agreement depends on person, number, and in some aspects, on gender as well.

5	sitɑ	məstə	kɑ̄m-ø	kər-ɛ̃
	Sita-3FS-Nom	much-Adj	work-3S-Acc	do-3FS-Pres-Indf
	Sita works a lot.			

6	sitɑ	bɪlkʊl	kām-ø	ni	kər-ən-i
	Sita-3FS-Nom	absolutely-Adv	work-3S-Acc	not-Neg	do-Neg-3FS-Pres-Indf
	Sita does not work at all.				

So, while in sentences (2) and (4), the negative clitic was */-ən/*, in sentence (6), the clitic seems to be */-ən-i/*. The */-i/* at the end of the verb form marks feminine gender here. In the affirmative sentence, the feminine marker in the verb form is not */-i/* but */-ɛ̃/*, so, it cannot be said that the negative clitic stays the same, and just gets attached within the verb form itself. This also supports the argument that it is not an inflectional affix, but a clitic. The feminine marker is also changed, or at least, there is a vowel change taking place – the word-final vowel is raised in the negative form of the verb. Therefore, for now, it can be said that the negative verbal clitic for masculine grammatical subjects is */-ən/*, and for feminine grammatical subjects, it is */-ən-i/*. This means that the clitic is inflected for gender. Now, it also needs to be seen if the clitic gets inflected for number as well.

7	rɑm-kɛ̃	šɑg	bʰəl	lɑg-i
	Ram-3MS-Dat	vegetables-3Pl-Nom	good-Adj	seem-3Pl-Pres-Indf
	Ram likes vegetables.			

8	rɑm-kɛ̃	šɑg	bʰəl	ni	lɑg-eɳ⁻ˡ
	Ram-3MS-Dat	vegetables-3Pl-Nom	good-Adj	not-Neg	seem-Neg-3Pl-Pres-Indf
	Ram does not like vegetables.				

In sentences (7) and (8), the grammatical subject is the plural nominal */šag/*. Though there is no morphologically realised plural marker on the noun, the agreement pattern of the verb helps identify the plurality. It can be seen that this clitic does get inflected for number, and that the feminine and plural ending for negative sentences in the present indefinite are identical, though there may be some dialectal and idiolectal phonetic differences like variations in stress on the last syllable, leading to the differing vowels */ə/* and */e/*.

4.2 Other Tenses and Moods

9	rəmeš-əl	kʰɑɲ-ø	bəɲ-ɑ
	Ramesh-3MS-Erg	food-3S-Abs	make-3S-Pres-Perf
	Ramesh has made food.		

10	rəmeš-əl	kʰɑɲ-ø	ni	bəɲ-ɛ
	Ramesh-3MS-Erg	food-3S-Abs	not-Neg	make-Neg-3S-Pres-Perf
	Ramesh has not made food.			

11	mə-kɛ̃	bɔhtɛ	bʰɔl	lɑg-nɑ
	I-1S-Dat	very-Adv	good-Adj	seem-3S-Pres-Prog
	I am feeling very well.			

12	mə-kɛ̃	bɪlkʊl	bʰɔl	ni	lɑg-nɛ
	I-1S-Dat	absolutely-Adv	good-Adj	not-Neg	seem-Neg-3S-Pres-Prog
	I am not feeling well at all.				

Sentences (9) and (10) both have a transitive verb and are in the perfect aspect. Since Kumaoni too is a split-ergative language like Hindi and many other North Indian languages, these sentences mark the subject with an overt ergative case */-əl/*. The agreement of the verb therefore shifts to the direct object in these sentences, i.e. */kʰaɲ/*, which is overtly unmarked. The object is singular, so the agreement patterns should be similar in both sentences. But the verb form differs in (9) and (10), ending at */-a/* and */-ɛ/* respectively. The negative form of the verb is different in the present progressive aspect too. As can be seen, the ending of the verb in (11) is */-na/,* while in the negative sentence (12), it is */-nɛ/*. The difference is only in the vowel. Therefore, the negative verbal clitic is seen to be the same for the perfect and progressive aspects in the present tense.

This specification of the tense till now was important because testing similar pairs of sentences in past and future tenses did not reveal any existence of the negative verbal clitic in the negative sentences. This is different from the neighbouring language Nepali. As shown in these following sentences in the future tense, there is no visible negative verbal clitic in (13), and the verb forms in (13) and (14) are identical.

13	ləkšmi	bʰōl	ɪskul-ø	jɑ-l-i
	Lakshmi-3FS-Nom	tomorrow-Adv	school-3S-Loc	go-3FS-Fut-Indf
	Lakshmi will go to school tomorrow.			

14	ləkšmi	bʰōl	ɪskul-ø	ni	jɑ-l-i
	Lakshmi-3FS-Nom	tomorrow-Adv	school-3S-Loc	not-Neg	go-3FS-Fut-Indf
	Lakshmi will not go to school tomorrow.				

Another thing observed from the data is that this negative verbal clitic is seen only in in the indicative mood, and not in imperative or interrogative constructions, as illustrated in the following sentences.

15	wɑ̃	bəsɑ
	there-Adv	sit-2S-Hon-Imp
	Sit there.	

16	wɑ̃	jʰən	bəsɑ
	there-Adv	Neg-Imp	sit-2S-Hon-Imp
	Do not sit there.		

17	tʊm	ɑj	šɪkɑr-ø	ni	bəɳɑ-lɑ
	you-2S-Hon-Nom	today-Adv	chicken-3S-Acc	not-Neg	make-2S-Hon-Fut-Indf
	Won't you cook chicken today?				

4.3 Why a Clitic?

Shopen defines clitics as 'words that occur in a fixed position in relation to some other sentence element' (1985:53). Clitics may occupy a fixed position in relation to a verb, and are usually positionally defined. It may be postulated, given the phonological similarity between the */-ən/* morpheme and the usual negative marker */ni/* or */nə/*, the negative clitic may have developed from the negative particle itself by means of metathesis, because it clearly adds some semantics to the verb form. Though it cannot be said to be a contracted form for sure, it may have developed historically from the negative particle.

A special clitic, unlike a simple clitic, can appear in very restricted positions in a clause. They are special because their syntax differs from the simple clitics. Their syntactic position and syntactic meaning are different from the free word from which they are derived, in this case, presumably, the negative particle */ni/*. This negative verbal clitic may be therefore called a special clitic. As a verbal clitic, it may share some features with inflectional affixes, but they are not the same.

In the present indicative, the presence of the negative particle */ni/* in the sentence, by signaling a critical semantic distinction in terms of negation, has a significant impact on the verb form. It shows a bipartite negation, where the negative verbal clitic and the sentential negative particle are local, i.e. belong to the same clause. This particle within the verb is definitely a clitic because it is easier to posit a process of morphological derivation rather than having two different sets of conjugations for all Kumaoni verbs in the mental lexicon, according to the principle of economy.

If one is to compare this with Bangla or Nepali, the negative particle and the verb do not get fused together in Kumaoni but stay separate, with an additional marker within the verb itself. In Bangla, there is an asymmetry in the simple and the perfect aspects in negation. In simple aspect, the negative particle is separate: */na/*, but in the perfect aspect, it becomes fused with the verb, and also becomes */ni/*. On the other hand, in Nepali, the negative morpheme */-ɪnə/* or */-enə/* gets fused with the verb form, at least in the simple present and perfect aspects. Moreover, similar to Kumaoni, Nepali also changes the verb conjugation a little in the negative constructions, e.g. */həri besi kam gərčʰa/* (Hari works a lot), when negated, becomes */həri ekdəm-əyi kam gərd-enə/* (Hari does not work at all). So, the negative

conjugation in Nepali too is different, but it seems to be more of an inflectional affix than a negative clitic. Further research may shed some more light on this.

It can be concluded that the negative construction is marked in discourse by these alterations. As Payne (1997) observes, negative constructions may involve multiple operators because a negative assertion is communicatively very distinct from the affirmative correspondent, so languages always tend to develop strong and easily perceivable devices to express this difference.

4.4 Negative Copula

Copular constructions are those where the subject is connected to its predicative adjective or noun with the 'be' verb or the copula which forges a relationship between the nominal(s) and the property described. Existential constructions talk about the existence or presence of something, e.g. 'There is…'. Possessive constructions show possession, e.g. 'She has…', and these are both formed in most Indo-Aryan languages by using the 'be' verb, e.g. */hona/* in Hindi. However, in Kumaoni, such constructions constitute a separate category of negation, which is discussed below.

The 'be' verb in Kumaoni is */čʰʊ/* for singular subjects and */čʰən/* for plural and honorific subjects. However, there is a special form of the 'be' verb in Kumaoni which is used only in negative copular, existential and possessive constructions – this form is */nʰa/* or */nʰɛ/*. In other words, it is in complementary distribution with the copula */čʰʊ/* or */čʰən/*, so it can be considered as suppletive negative copular verb. This is similar to a negative copula in another Pahari language called Pahari-Pothwari – Hussain (2024). In such sentences, there is no need for the syntactic negative element */ni/* because the verb itself semantically negates the utterance. The reason why this morpheme has been identified as a verb and not as a special kind of negative particle is because it shows agreement in terms of number and honorificity, as will be evident from the following pairs of sentences.

18	my-ər	čʸəl	yɑ̃	čʰʊ
	I-1S-Nom	son-3MS-Nom	here-Adv	be-3S-Pres-Indf
	My son is here.			

19	my-ər	čʸəl	yɑ̃	nʰɑ
	I-1S-Nom	son-3MS-Nom	here-Adv	be-Neg-S-Pres-Indf
	My son is not here.			

In the sentences above, the subject is singular and non-honorific. So, for singular and/ or non-honorific subjects, */nʰa/* can be used. Whether the subject is male or female makes no difference.

4.5 Agreement with Number and Honorificity

Verb agreement in Kumaoni in general treats a plural nominal and an honorific entity as identical. In such cases, it *becomes /nʰa-tɪ/, /nʰa-tʰɪ/, /nʰa-tən/, /nʰɛ-tɪ/, /nʰɛ-tʰɪ/* or */nʰɛ-tən/*, based on dialectal variation or speaker choice. However, it was observed that not all speakers make this distinction.

20	my-ər	čʸəl-čʸeli	yə̃	čʰən
	I-1S-Nom	son-3MS-daughter-3FS-Nom	here-Adv	be-3Pl-Pres-Indf
	My son and daughter are here.			

21	my-ər	čʸəl-čʸeli	yə̃	nʰɑ-tən
	I-1S-Nom	son-3MS-daughter-3FS-Nom	here-Adv	be-Neg-Pl-Pres-Indf
	My son and daughter are not here.			

22	my-ər	bə̄jʊ	yə̃	nʰɑ-tən
	I-1S-Nom	father-3MS-Hon-Nom	here-Adv	be-Neg-3S-Hon-Pres-Indf
	My father is not here.			

Possessive constructions also show the negative copula when negated. That can be seen in sentences (23) and (24). Distinctions in honorificity have not been shown here, but they can be assumed to follow the same pattern, based on their similarity in the other kinds of constructions containing the copula.

23	my-ər	pɑs	ek	bəkər	čʰʊ
	I-1S-Gen	near	one	goat-3S-Nom	be-S-Pres-Indf
	I do not have any goat.				

24	my-ər	pɑs	koye	bəkər	nʰɑ
	I-1S-Gen	near	any-Adj	goat-3S-Nom	be-Neg-S-Pres-Indf
	I do not have any goat.				

25	həm-ər	pɑs	bɪlkʊl	lɛ	ḍəbəl	nʰɑ-tɪ
	we-1Pl-Gen	near	absolutely-Adv	Emph	money-3Pl-Nom	be-Neg-Pl-Pres-Indf
	We do not have any money.					

The negative copula can only be seen in the present tense, and also, mostly in statements. For forming questions with the copula, speakers vary between using the negative copula */nʰa/* and the (affirmative) copula with the usual negative particle */ni/*. Copular constructions in the past tense too treat negation like any other verb, i.e. with */ni/,* as can be seen from the sentences (26) and (27). So, any tense other than the present do not show the occurrence of the negative copula.

26	rɑm	dʊkʰi	čʰɪ
	Ram-3MS-Nom	sad-Adj	be-3S-Pst-Indf
	Ram was sad.		

27	rɑm	dʊkʰi	ni	čʰɪ
	Ram-3MS-Nom	sad-Adj	not-Neg	be-3S-Pst-Indf
	Ram was not sad.			

A similar phenomenon can be observed in the neighbouring language Garhwali as well. In some rural dialects of Garhwali, existential constructions have no copula in the affirmative, but in the negative constructions, */na/* is the counterpart of */nʰa/* in Kumaoni, and it is only phonetically similar to the negative particle */nə/*. Something similar happens in Bangla too – existential constructions in the present are zero-copula in the affirmative, but when negated, the negative copula */nei/* is used, which is different from the usual negative particle

/na/. The negative marker concerns the illocutionary aspect of the linguistic act itself. It is of pragmatic significance, and may be marked. It makes sense therefore to ask why only the present tense shows this special use of the negative copula. That may be because the present is of the most importance in an active discourse, and hence shows such features of prominence.

5 Limitations and future directions

This paper is constrained by time and resources, so it was not possible to look at how negation functions in other complex aspects in the past and future tenses. More extensive data is required for closer investigation in this regard. Some comparisons have been done between Kumaoni and Garhwali, Nepali, Hindi, and Bangla. However, investigation in finer detail is required to identify any interesting contact phenomena with regard to diffusion of features. Such endeavours could point the way for future directions of research.

6 Conclusion

In conclusion, after having analysed instances of sentential negation in Kumaoni, the conditions of use of the negative verbal clitic and the negative copula have been identified. To summarise very briefly, the negative verbal clitic */-ən/* is attached to the verb form in the present indicative, and it is inflected for both number and gender. The negative copula */nʰa/* or */nʰɛ/* is used in negative copular, existential and possessive constructions. This also agrees with number or honorificity, though this distinction exhibits idiolectal variation. These features prove the pragmatic markedness of the operation of negation. The data is Kumaoni has also been compared with neighbouring languages like Garhwali, Nepali, and Hindi, and more detailled exploration could potentially reveal some interesting contact phenomena in relation to negative clitics and copula.

List of Abbreviations

1 1st person	*Dat* dative
2 2nd person	*Emph* emphatic
3 3rd person	*Erg* ergative
Abs absolutive	*F* feminine
Acc accusative	*Fam* familiar
Adj adjective	*Fut* future tense
Adv adverb	*Gen* genitive

Hon honorific

Imp imperative mood

Indf indefinite aspect

Loc locative

M masculine

Neg negative

Nom nominative

Perf perfect aspect

Pl plural

Pres present tense

Prog progressive aspect

Pst past tense

S singular

References

Bhatia, T. K. 1995. *Negation in South Asian Languages*. Patiala: Indian Institute of Language Studies.

Grierson, G. A. 1916. *Indo-Aryan Family: Central Group: Specimens of the Pahari Languages and Gujuri, Linguistic Survey of India*, IX(IV). Kolkata: Office of the Superintendent of Government Printing.

Hussain, A. 2024. *Clausal negation in Pahari-Pothwari.* MA by research thesis, University of York. Date of access: 17 February 2025. Available online: <oai:etheses.whiteroseac.uk:34946>.

Kumaon Mandal Vikas Nigam. 2025. *Language*. Date of access: 13 March 2025. Available online: <https://www.kmvn.in/language#:~:text=These%20dialects%20are%20most%20closely,Rajasthani%2C%20Kashmiri%2C%20and%20Gujarati>.

Miestamo, M. 2005. Standard Negation. *The Negation of Declarative Verbal Main Clauses in a Typological Perspective*. Berlin; New York: Mouton de Gruyter.

Peterson, J., & Chevallier, L. 2022. Towards a Typology of Negation in South Asian Languages. *Bhasha*, 1(1), 17-62. Date of access: 9 February 2025. Available online: <10.30687/bhasha/8409-3769/2022/01/005>.

Sharma, D. D. 1987. *The Formation of Kumauni Language: Part II (Morphology & Syntax)*. New Delhi: Bahri Publications.

Shopen, T. 1985. *Language Typology and Syntactic Description: Volume 3*. Cambridge University Press.

Subbārāo K.V. 2012. Case and agreement. In: *South Asian Languages: A Syntactic Typology*. Cambridge University Press.

Tatarovska, O. 2023. The Pragmatic Aspects of Linguistic Negation. *Research Trends in Modern Linguistics and Literature*, 6. 65-75. Date of access: 15 February 2025. Available online: <https://doi.org/10.29038/2617-6696.2023.6.65.75>.

वाक् मंथन
ISSN: 2426-2149

Neologisms across North East Indian Languages

Prof. Bidisha Som & Co-author: (1) Mirdom Teronpi
Co-author: (2) Tsukhumla Yimchunger
IIT Guwahati

Abstract

Neologism refers to the words or lexical units newly coined to express innovative things and concepts in society. Novel terms are produced or invented daily to accommodate modern ideas in a speech community. The paper explores neologisms in the context of North East Indian languages, focusing on the process of coinage and integration into the standard vocabulary. It also investigates metaphorical extension to neologism. The method of research is predominantly qualitative through structured interviews, participant observation, textual and media analysis, and surveys in languages such as, Yimkhiung, Bodo, Mizo, Lotha, Karbi, Ao, Simtee and Khasi. The emergence of technologies necessitates communities to borrow or invent recent terms and fit them into the grammatical structure and lexicon of language. In Lotha, 'mobile phone' is termed *ɬantʰapʰen*, where *ɬantʰa* means 'to talk', and *pʰen* 'tool/device' which translates to 'a tool or device to talk with'. This is an example of semantic innovation. In Yimkhiung, 'car' is coined *loluŋpe* where 'lo' means 'soil' and *luŋpe* means 'to roll' which translates as 'something to roll on the soil'. This phenomenon gains importance in the northeastern Indian context, as neologism helps expand and promote many languages, which are otherwise dwindling in number and usage.

Keywords: neologism, northeast Indian languages, borrowings, coined terms.

1. Introduction

Neologism is a term derived from the Greek words *neo* meaning 'new' and *logos* meaning 'word'. According to Merriam Webster's definition, neologism is "a new word, usage, or expression" (Čilić & Plauc, 2021:117). First used in 1772, neologism as a concept, *neologie* or *neologisme* (in French) came about in 1758 in France and its study in linguistics commenced in the twentieth century (Kakimova, et al. n.d.: 1). The old meaning "practice of innovation in language, the use of new words or old words in new senses" was modified to "new word or expression" in 1803 in English (Januševa & Jurukovska, 2015:101). With the development and advancement in technology, people have adapted to new and improved ways of interaction. The use of the internet to communicate has further led to creation of new words and new

terms. Neologisms arise because of "the development of science and technology, changes in everyday and social life", according to Arnold I.V., as a result of "the emergence of new concepts, phenomena and qualities", to N.S. Valgina and D.E. Rosenthal, and "in response to a particular need", to P. Newmark (Ibid, n.d.: 2). Language is evolving continually; as older words gradually turn archaic and lose their significance, the requirement to describe new concepts increase. Every year, each language is known to have acquired at least 3000 new words. However, quantifying it is problematic since such created words face the challenge of being accepted in the lexicon and many tend to disappear after a short period (Newmark, 1988:140).

Neologism is a phenomenon in linguistics that is evident in the languages of North East region of India. The region is culturally and linguistically rich and diverse; it is home to three major language families: Tibeto-Burman, Indo-Aryan, and Austro-Asiatic. Most of the languages belong to the Tibeto-Burman language family, with 20 independent subgroups and 100 to 300 individual languages. The North East Indian languages account for 75% of the total languages spoken in the country. The languages under this study, namely Mizo, Lotha, Karbi, Yimkhiung, Ao, Bodo, Simtee and Khasi are spoken in this region. According to the Census of India 2011, the population of Mizo is 8,30,846, Lotha 1,79,467, Karbi 5,28,503, Yimkhiung 83,259, Ao 2,60,008, Bodo 14,82,929, Simte 6728 (not included in the census), and Khasi 14, 31, 344.

This is a preliminary study on neologisms across the North East Indian languages and further research can explore the evolution of these neologisms, their broader impact on society, and the attitudes of speakers towards them. The study also investigates how metaphors are extended in conceptualisation of neologisms.

2. Research objectives and methodology

2.1. Objectives of the study

The primary objective is to investigate neologisms in the lexical systems of North East Indian languages and to examine their structures, formation processes, and semantic functions, and compare the significance of neologisms across these languages. It also aims to examine the use of metaphors in conceptualizing modern concepts from the perspective of cognitive linguistics.

2.2. Research Questions

(1) What are the common types of neologisms across the North East Indian languages?

(2) What are the morphological strategies used in integrating neologisms in North East Indian languages?

(3) How are metaphors used in the creation of neologisms across these languages?

2.3. Method of research

The method of research is predominantly qualitative through structured interviews, participant observation, textual and media analysis, and surveys. The data are collected from Karbi, Yimkhiung, Bodo, Lotha, Khasi, Simtee, Ao, and Mizo, through sources like media (TV, News programs) and social networking sites (Facebook, X (Twitter), Instagram and Youtube).

3. Literature Review

3.1 Defining neologisms

The Cambridge Dictionary defines neologism as 'a new word or expression, or a new meaning for an existing word'. In other words, neologism refers to the words or lexical units newly coined to express new things and concepts in society. Newmark defines it as, "newly coined lexical units or existing lexical units that acquire a new sense" (Occulares & Consad, 2014:1). Neologisms fill the gaps when a language lacks appropriate terms for certain ideas, thoughts, emotions, or social phenomena. Newmark reiterates that neologisms are "new words naming newly invented or imported objects or processes, or new expressions that suddenly fill one of the innumerable gaps in a language's resources for handling human thought and feeling at some level of formality" (Newmark, 1988:122).

Cabré (2015), forty years later, added to this definition, which can be translated as, "as objects of knowledge, neologisms are relative units that can only be identified when placed in a specific time period, discursive context and enunciative perspective" (Guerra, 2016:529). Cabré established four parameters: the date of appearance in a lexicon, exclusion from dictionaries, formal or semantic instability, and the perception speakers have of an item's novelty (Ibid, 2016:530).

3.2. Types of neologisms

L. Gilbert classified the structure of neologism into three: phonological neologisms; morphological neologisms; and borrowing (Kakimova, et al. n.d.: 3). According to Newmark (Newmark, 1988:141-149), old words with new senses, new coinages, derived words, abbreviations, collocations, eponyms, phrasal words, transferred words, acronyms, and pseudo-neologisms, are types of neologisms. Depending on the language, the formation of neologisms can be identified. For instance, English neologisms are created through blending, shortening or clipping, borrowing, back-formation, semantic drift, compounding, and affixation. The types of neologisms are discussed in the following:

a) Old words with new senses: They are words that already existed in the language but have new meanings. Some such words in English are 'gay', 'scene', etc.

b) New coinages: Words that have originated from linguistic innovation. Examples are brand names such as 'Bisto', 'Bacardi', 'Revlon'.

c) Derived words: Words that have been derived from ancient Greek and Latin morphemes with suffixes *-ismo, -ismus, -ija*.

d) Abbreviations: They are a common type of pseudo-neologism more common in French.

e) Collocations: Noun compounds or adjectives with nouns. Some examples are 'cold-calling', 'wishful thinking', 'lead time', 'domino effect'.

f) Eponyms: They are terms derived from names of people or objects. E.g. Parkinson's Law.

g) Phrasal words: These are restricted to English. E.g. 'work-out', 'trade-off', 'check-out'.

h) Transferred words: The words which have been imported but have less dependence on their contexts. E.g. 'Adidas', 'Gallini', 'Levi'.

i) Acronyms: Examples are UNESCO, UNICEF, ASEAN.

j) Pseudo-neologisms: Words that appear as newly coined but are not actually truly new.

There are different morphological processes involved in the creation of neologisms: derivation, back-formation, compounding, conversion, borrowing, abbreviations, clippings, acronyms, and onomatopoeia.

4. Main findings and Discussion

The neologisms across these languages fall under the types of neologisms put forward by Levchenko (2010:122). According to Levchenko, they are of three types: an entirely new word, a wholly new meaning from an old word, and a new seme from an older word (Akunna, 2012: 26). However, the selected languages consist of neologisms that are completely new terms and new terms that have added new meanings but have not lost the significance of the old ones. As stated under Newmark's classification, some of the data fall in the 'new coinages' category: words that have originated from linguistic innovation.

4.1. Neologisms across the Northeast Indian languages are created through various morphological processes. These various processes are classified into the following:

4.1.1. **Compounding**

Compounding is one of the major morphological operations common in many languages through which words and word forms are generated. It is a combining process, "a joining of two separate words to produce a single form" (Yule, 2016:54). A compound word consists of "at least two bases that are both words, or

at any rate, root morphemes" (Lieber, 2016:55). In this study, compounding is a recurring word-formation process in almost all of the selected languages.

The following examples illustrate this process:

1. *lam-ri-hiŋ* [Karbi]
   speech-rope-hand
   'Mobile phone'

2. *biak-hlatna* [Mizo]
   to talk- distance
   'Mobile phone'

3. *tʰil-tʰawnna* [Mizo]
   object-to transfer
   'Bluetooth'

4. *ɬantʰapʰen* [Lotha]
   to talk-device
   'Mobile phone'

5. *poto-jak-chamkʰikʰi* [Yimkhiung]
   general-rule-separate
   'Social distancing'

6. *tʰijushak-han-liu* [Yimkhiung]
   news-carry-rope
   'Mobile phone'

7. *pʰü-müküp* [Yimkhiung]
   health-strong
   'Immunity'

8. *kʰɔwl-huah-nei* [Simte]
   machine-brain-possess
   'Computer'

In examples (1) from Karbi and (6) from Yimkhiung, the terms for 'mobile phone' are constructed by integrating the function of the object, i.e. it is a voice and hand-held device or tool that carries word or news through a rope or a line. Examples (2) from Mizo and (4) from Lotha for the same object 'mobile phone' denote a device for speaking or talking from a distance. Other examples also indicate linguistically creative and innovative forms to represent neologisms.

4.1.2. Hybridization

Hybridization is another morphological process present in some of the selected languages. It refers to "the mixing of words and phrases of two different codes or languages" (Khan, 2013:822). Hybridization is a common process of constructing neologisms in the sample languages where elements have been mixed from one language to another. In the following, a few examples have been presented:

9. *hisapna kʰɔwl* [Mizo]
   hisap (Hindi) 'calculation' + *kʰɔwl* (Mizo) 'machine'
   'Computer'

10. *ka kali* [Khasi]
    FEM (Khasi) + *kali* (Hindi 'gəri')
    'Car'

11. *ka trok* [Khasi]
    FEM (Khasi) + *trok* (English 'truck')
    'Truck'

12. *motɔr gaɾi* [Bodo]
    məʊtə(ɹ) (English) + *gəri* (Hindi)
    'Car'

4.1.3. Blending/Clipping

Blending is a word-formation operation that can be defined as "the combination of two separate forms to produce a single new term" (Yule, 2016:55). Clipping is a type of blending that involves reducing the elements of words when "a word of more than one syllable is reduced to a shorter form" (Ibid, 2016:55). In English, the concept of clipping is confined to the beginning of the syllable; however, in the selected languages, clipping can be seen even at the final syllable.

13. *bai-anɔŋ → bainɔŋ* [Karbi]
    work-reward= 'Salary'

14. *(o)ɬu-ran-pʰen → juranpʰen* [Lotha]
    water-to scoop out-device or tool = 'Mug'

These examples (13) and (14) show how a few syllables have been reduced or removed to construct shorter forms of the complete words.

4.1.4. Loan words

A loan word refers to "a word taken from another language, adapted to the sound system, spelling system or morphological system" (Sijens & Van de Velde,

2020:8). Some words are usually borrowed from English and modified according to the sound and spelling systems of the target languages.

15. *biteri* [Khasi]
'battery'

16. *battiɾ* [Ao]
'battery'

17. *waipʰai* [Meitei]
'wifi'

18. *kaɾen* [Bodo]
'current/electricity'

19. *senitaizaɾ* [Simte]
'sanitiser'

In these examples, (15) from Khasi, (16) from Ao, (17) from Meitei, (18) from Bodo, and (19) from Simte, the words have been incorporated from English and some sounds have been altered to the sound systems and phonotactics of the target languages.

4.1.5. Loan Translation or Calque

As defined by Crystal (Khan, 2013:823), calque is "a type of borrowing where morphemic constituents of the borrowed word are translated item by item into equivalent morphemes in the new language". It is "a word taken from one language and translated in a literal or word for word way to be used in another", according to the Cambridge Dictionary (Cambridge Dictionary).

20. *set-sluŋ* [Khasi]
lock-completely close
'Lockdown'

21. *tʰah-hriak* [Mizo]
to kill-oil
'Sanitiser'

In these two examples, (20) from Khasi and (21) from Mizo, the main functions of the word meanings have been literally translated in the target languages. 'Lockdown' and '*set slung*' serve the same purpose of being confined or locked up to a particular place; similarly, 'sanitiser' and '*tʰah hriak'* have the same function of a liquid to kill germs.

4.1.6. Coinage

Coinage refers to "invention of new terms" (Yule, 2016:53). These totally new terms have no technical origins formed from completely non-existing words or

terms in the language. These terms have been invented to stand for concepts or objects that were previously non-existent. Some of the examples are given below:

22. *hmaibu* [Mizo]
'Facebook'

23. *zotoro* [Lotha]
'car'

24. *ɬambuŋ* [Bodo]
'Mobile phone'

25. *andw* [Bodo]
'Isolation'

4.2. Metaphorical Extension to Neologisms

Metaphor in Cognitive Linguistics is largely associated not with rhetoric or a figure of speech but with understanding how it demonstrates the way people think. Lakoff and Johnson in *Metaphors We Live By* (1980) makes it clear that metaphor is merely a characteristic of language but a matter of thought and action (Lakoff & Johnson, 1980:4). Metaphors give us a glimpse into people and cultures' view of the world and their surrounding environment. This perspective is termed as "Conceptual Metaphor"; it views metaphor as "a mental mapping from source domain to target domain based on the similarity between the two things" (Zheng, 2015:1381).

Languages tend to use old meanings and extend them to new and modern concepts. Lakoff's understanding of metaphor seems to accommodate this process. Metaphor is defined as, "understanding and experiencing one kind of thing in terms of another" (Lakoff & Johnson, 1980:5). The function of metaphor in this sense is relevant in that languages usually employ old words and designate them to new objects. Therefore, old concepts are used to understand and experience new concepts. In the following examples, the similar shape and structure/function from old word meanings are extended to modern objects such as 'car' and 'computer'.

26. *ochi-suŋen* [Ao]
truth-box
'Computer'

In example (26), *suŋen* meaning 'box' is extended to the modern concept of computer. This metaphor is based on the mapping of 'DEVICES ARE CONTAINERS' where devices are considered as containers pertaining to the truths that they can find from the box.

27. *hɛm-kɔŋlɔŋ* [Karbi]
house-to roll
'Car'

In this example from Karbi, the metaphor is based on 'VEHICLES ARE HOUSES' where the function of a house *hem*, which is an enclosure or protective space with roof, doors and windows, has a visual representation of a 'car'.

28. *me-ri-hiŋ* [Karbi]
fire-rope-bind
'Electricity'

29. *bor-diŋ* [Khasi]
power-fire
'electricity'

In these examples, the mapping of source domain 'fire' to target domain 'electricity' can be seen in the metaphor 'ENERGY IS FIRE'. Fire and electricity serve a similar purpose of delivering or generating light; therefore, the names are transferred to conceptualise the new object in these languages.

The words under compounding, loan translation or calque, and blending, are also metaphors such as - in Karbi, *lam-ri-hiŋ* (speech+rope+bind), mobile phone is conceptualized as a tool that binds speech to the hand, reflecting both its function and the embodied interaction with the device. Similarly, mobile phone is Mizo is *biak hlatna*, a combination of *biak* 'to talk' and *hlatna* meaning 'distance'. It means to talk from a distance. Additionally, *tʰil thɔnna* means 'Bluetooth' a combination of *tʰil* 'object' and *thɔnna* 'to export' which means to export/transfer an object. Here the abstract things are viewed as an entity and, it means even pictures, songs, or videos can be 'transferred' without direct manual intervention but through aerial medium (from device to device). Interestingly, in Lotha *jantʰapʰen* is made up of *jantʰa* 'to talk' and *pʰen* 'device' which translates to 'a talking device'. A device used for speaking and a device that speaks. An entity that is personalized like any human who can communicate. Many northeast Indian languages comprise of rich metaphors which has a scope for future studies.

5. Conclusion

In conclusion, neologism as a phenomenon in the languages of North East India is predominant and it illustrates the effect of technological advancement and development, which thereby leads to the dynamic evolution and expansion of language. This study demonstrates that neologisms or the adoptions of new terms in languages like Karbi, Yimkhiung, Bodo, Lotha, Simte, Ao, and Mizo are pivotal in facilitating effectiveness in communication and reflect contemporary concepts in society where globalization continues to spread. Compounding, Coinage, Hybridization, Clipping and Blending are some of the morphological processes that are used to construct new terms; these constructions enrich the lexicons of these languages and help promote them. These formation processes highlight the creativity and adaptability of the speakers in integrating new concepts while retaining their integrity.

An analysis of metaphor extension to neologisms found that metaphors allow languages to extend old meanings to new contexts, as illustrated in examples from Ao and Karbi. The transition and evolution of languages can be seen through the adoption of modern words by coining or through calque. The fluidity of new words blending into the languages through creative innovations can be seen in the examples. These extensions reveal linguistic creativity on the part of language speakers through neologism. The study observe that some communities have a patriotic and strong attitude towards their language in terms of adding new lexicons while some are resistant to the adaptations. Metaphorical extension studies to neologism are novel ways of looking into language adaptation through cognitive lens. Neologism, therefore, can be considered as an essential tool to help expand and preserve minority languages.

References:

Calque. Cambridge Dictionary, Cambridge University Press, https://dictionary.cambridge.org/dictionary/english/calque. Accessed 7 May 2025.

Cilic, I. S., & Plauc, J. I. 2021. Today's Usage of Neologisms In Social Media Communication. *DrusTvene I HumanisticKe Studije*, 1(14), 115-140. https://doi.org/10.51558/2490-3647.2021.6.1.115

Crystal, D. 1997. A Dictionary of Linguistics and Phonetics. Blackwell Publishing.

Fang, Y. 2021. A Glimpse at the Study of English Neologism. *Business Prospects*, 2(2), 78-84. https://doi.org/10.52288/bp.27089851.2021.12.13

Guerra, A. R. G. 2016. Dictionaries of Neologisms: A Review and Proposals For Its Improvement. *Open Linguistics*, *2*, 528–556. https://www.degruyterbrill.com/document/doi/10.1515/opli-2016-0028/pdf

Haokip, Pouthang. 2011. Socio-Linguistics Situation in North-East India. Concept Publishing Company, New Delhi- 110059.

Janssen, M. 2005. Orthographic Neologisms Selection Criteria and Semi-Automatic Detection. In *Instituto De Linguística Teórica E Computacional* [Journal-article].

Janusheva, V., & Jurukovska, J. 2015. Conceptual Formulation of Neologisms in Various Dictionaries and Primary School Macedonian Language Course Textbooks. *Journal of Language and Linguistic Studies*, *11*(2), 99–116. http://files.eric.ed.gov/fulltext/EJ1105185.pdf

Kakimova, M. E., et al. n.d. The New Content of The Concept Of "Neologism" At Present Stage in the Development of Linguistic Science. https://doi.org/10.48371/PHILS.2024.72.1.010

Khan, Mohsin. 2013. Neologism In Urdu: A Linguistic Investigation of Urdu Media. *Language in India*, 13 (6). http://www.languageinindia.com/june2013/mohsinneologism.pdf

Lieber, Rochelle. 2016. Introducing Morphology. Cambridge University Press

Maria, C. 2024, January 4. Neologisms: How New Words Shape Our Society. *TPR teaching*. https://www.tprteaching.com/neologisms/

Medvid, O., et al. 2022. Ways Of Generating Neologisms in Modern English. *Filologicni Traktati*, 14 (2), 73-84. https://www.doi.org/10.21272/Ftrk.2022.14(2)-8

Neologism. In *Merriam-Webster.com Dictionary*. Retrieved May 6, 2025 from https://www.merriam-webster.com/dictionary/neologism

Miri, Mrinal. 2002. Linguistic Situation in North-East India. Concept Publishing Company, New Delhi- 110059

Newmark, P. 1988. A Textbook of Translation. http://ilts.ir/Content/ilts.ir/Page/142/ContentImage/A%20Textbook%20of%20Translation%20by%20Peter%20Newmark%20(1).pdf

Oculares, Genesis & Ramos Consad, Rizza. 2014. Lexical Blends in Pinoytechblog. 10.13140/RG.2.2.31881.11364.

Ortiz, R. 2021. English Neologisms in Modern Times. Unpublished thesis. https://addi.ehu.es/bitstream/handle/10810/55740/TFG_Ortiz.pdf?sequence=1

Other Word Formation Processes. n.d. Morphology. https://semanticsmorphology.weebly.com/other-word-formation-processes.html#:~:text=Though%20not%20a%20very%20common%20word%20formation%20process%2C,words%20to%20denote%20previously%20nonexistent%20objects%20or%20concepts.

Sijens, H., & Van De Velde, H. 2020. The Formation of Neologisms in a Lesser-Used Language: The Case of Frisian. *Dictionaries, Journal of the Dictionary Society of North America*, *41*(1), 45–67. https://doi.org/10.1353/dic.2020.0004

Yule, G. 2001. The Study of Language. https://doi.org/10.1017/cbo9780511757754

Zheng, Y. 2015. A Metaphorical Study On Chinese Neologisms. ISSN 1798-4769. *Journal of Language Teaching and Research*, Vol.6, No. 6, 1379-1383. DOI: http://dx.doi.org/10.17507/jltr.0606.28

वाक् मंथन
ISSN: 2426-2149

Word Formation Processes in Zyphe[1]

Parag Kumar Deka
Co-Author: Dr. Arup Kumar Nath
Tezpur University
pkdeka123@gmail.com
arupjnu@gmail.com

Abstract

This study explores the word formation processes in Zyphe or Zophei, a Kuki-Chin language spoken primarily in Chin Province Myanmar, and Northeast India. Zyphe, like many other languages of the Kuki-Chin group, exhibits rich morphological structures that contribute to its unique word formation mechanisms. The research identifies and examines key processes, including affixation, derivation, compounding, reduplication and onomatopoia, conversion and borrowing.

Keywords: Word formation, prefix, suffix, derivation, reduplication, onomatopoeia.

1.0 Introduction:

The Zyphe or Zophei [2]is a Chin ethnic minority group inhabiting in Mizoram, India, and southern Thantlang township in western Chin state, Myanmar. The Zyphe language belongs to the Maraic branch of the Kuki-Chin group of the Tibeto-Burman Language Family (Vanbik,2009). There are three major dialects of the Zyphe language, viz. Upper Zyphe, Lower Zyphe A and Lower Zyphe B.The Zopheis identify ethnically as part of the Laimi tribe, a linguistically diverse group which contains speakers of many Chin languages including the Maraic languages Mara, Senthang, Lutuv (or Lautu), and Zotung (Lotven,2021). According to Ethnologue, there were approximately 17,000 Zophei speakers living in Burma/Myanmar and India with 20,000 speakers worldwide. In Mizoram, India, the Zyphe population is estimated to be about 4000.

---

[1] This paper is part of the NEC-funded project *Documentation and Revitalization of Nthyni, Khongdei, and Zyphe: Three Endangered Languages of North-East India*. The authors gratefully acknowledge the financial support provided by the North Eastern Council, Shillong, for facilitating this research.

[2] Zyphe or Zophei are two names for the same language spoken by a Chin ethnic minority community living in Chin State, Myanmar and Mizoram.

1.1 Genetic Classification

The first comprehensive subgrouping of the Kuki-Chin languages was undertaken by a Catholic priest, Fr. André Bareigts, who identified three main subgroups: Southern, Central, and Northern. Bareigts' classification was not based on linguistic criteria. Within the Central Chin languages, he further divided them into Central Northern, Central Chin and Southern Central Chin. The Southern Central Chin subgroup constitutes a separate branch called Maraic, a term coined by James Matisoff.

In Bareigts' classification of the Kuki-Chin family, Zyphe, or Zophei, is categorized under the South-Central group within the Maraic branch. (VanBik, 2009)

Kenneth VanBik, in his thesis on Proto-Kuki-Chin, also identifies the Maraic branch as a distinct subgroup within the Kuki-Chin family. According to VanBik(VanBik, 2009), the Maraic group is subdivided into three smaller groups: Mara, Zotung, and Sengthang. Zyphe or Zophei is included in the Mara group.

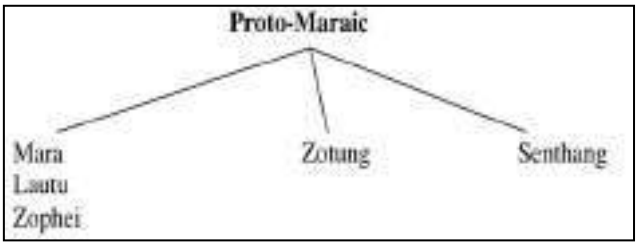


Figure 1**: Kenneth VanBik's Soubgrouping Schema for the Maraic group.**

2.1 Aims and Objectives:

The primary objective of this paper is to study and explain the internal structure of the complex words which are constructed through different morphological processes in Zyphe. It tries to describe the grammatical and lexical expansion of the Zyphe language. It also explores how many possible morphological processes are employed in the language to form new lexemes of different grammatical categories.

2.2 Methodology

The method employed in this study is descriptive in nature. The data described and analyzed here is the outcome of a fieldwork that was undertaken to Siata and Iana, two remote villages of the Saiha district, Mizoram.

Map 1: Siata and Iana (Ainak) village,

The primary data collected for the study through a questionnaire (pertaining to this research) and the interview method.

3.0 Literature Review:

Zyphe, also known as Zophei, is a lesser-studied language within the Maraic branch of the Kuki-Chin group. Samson Lotven's doctoral dissertation titled "The Sound System in Zophei Dialects and other Maraic Languages" gives an in-depth analysis on the phonetic and phonological account of the dialects of Zophei or Zyphe. Moreover, apart from his phonetic and phonological account, Lotven has also examined the agreement system on one of the Zophei dialects titled "Lawngtlang Zophei Verbal Complex"

Gabrial Gilbert (2024) has also explored the applicative construction of Zyphe titled "Applicative Morphosyntax of Zophei," where he discusses on the general functionality and related morphosyntax of applicatives found in Zophei.

4.1 Typological Features

i) The Zyphe language follows SOV word order.

ii) It is an agglutinating language

iii) It is a future vs. non-future language.

iv) It has three tones- high, middle and low

v) The language has a complex agreement system.

vi) It is a highly pro-drop language.

4.1.1 Word order:

The basic word order of Zyphe language is SOV.

1)	*raŋlai-tah*	*tso*	*a-ruh*
	Ranglai-ERG	cow	3SG-steal
	Ranglai stole the cow.		

In Zyphe, affixes are added linearly to form complex words, each representing a single grammatical category while retaining individual identity and ensuring morphological transparency, which makes it a classic agglutinating language.

2) *ka-e-tʰa*
1SG-EAT-PFV
I have eaten.

3) *a-vɔŋ-beh*
3SG-come-NEG
He didn‘t come

The language is highly pro-dropped. The subject and object pronouns are often dropped, and agreement markers are concatenated with the verb.

4) *raju-tah a-pa-dai*
raju-ERG 3SG-1SG-beat
Raju beat me.

4.2 Derivation

Derivation is a kind of word formation process in which a new word is created by adding affixes to a base or root word, often changing its meaning or grammatical category.

Zyphe demonstrates several types of derivation, including processes where verbs are nominalized, and adjectives or nouns are transformed into verbs through the addition of suffixes.

The following examples illustrate how nouns are transformed into verbs by adding the prefix *pa*, and verbs are nominalized in Zyphe through the addition of the suffixes *tu* and *na*. The suffix *tu* marks the agent or the doer of the action, while *na* serves to nominalize the verb, transforming it into a noun.

4.2.1 Derivation by prefixation

In Zyphe, nouns are transitioned to verbs by using the prefix *pa* marker.

5)	a)	*vɔŋ*	black	pa-vɔŋ=blacken
	b)	*sai*	red	pa-sai= redden
	c)	*raŋ*	white	pa-raŋ=redden
	d)	*kʰɔŋ*	partition	pa-kʰɔŋ= block or defend
	e)	*tsy*	stable	pa-tsy=sterngthen
	f)	*saŋ*	high	pa-saŋ=highten
	g)	*ny*	soft	pa-ny=soften
	h)	*tsaŋ*	hard	pa-tsaŋ=harden

Example:

6)	a)	*kʰua-so*	*hlaitah*	*mediŋ-tah*	*avai*
		rain-fall (falling rain)	before	cloud-ERG	sky
		a-pa-vɔŋ			
		3SG- blaken			
		Clouds blacken the sky before it rains.			
	b)	*rɔŋ-tah*	*tlaipia*	*a-pa-rɔŋ-a*	
		colour-ERG	wall	3SG-whiten-FUT.	
		The paint will whiten the wall.			

In sentences 6a and 6b, the prefix *pa* is concatenated with the color nouns *vɔŋ* and *rɔŋ*, transforming them into verbs such as *blacken* and *whiten*. If we consider example 4, it is evident that the prefix *pa* can also serve other functions as well. In this case, *pa* indicates first-person object agreement when attached to the verb *dai* (beat). In sentence 6b, the morpheme *a* also performs different functions depending on its position. When used as a prefix before a verb, it marks third person singular in the positive paradigm of the agreement system of Zyphe. The suffix *a* on the other hand, serves as a future tense marker.

4.2.2 Derivation by Suffixation:

4.2.1 Derivation of noun from verb by suffixing *tu*

The suffix *tu* is an agentive marker which is concatenated with the action verb which indicates the doer of a particular action.

7)	a)	*patsu*	teach	*patsu-tu* teacher	
	b)	*laŋ*	dance	*laŋ-tu*	dancer
	c)	*tʰɔ*	kill	*tʰɔ-tu*	killer

Example:

8)	a)	*patsu-tu*	*a-tsang*	
		teach-NMLZ		3SG-be
		He is a teacher.		
	b)	*laŋ-tu*	*a-tsaŋ*	
		dance-NMLZ	3SG-be	
		She is a dancer.		

Sentence 8 a) & 8b) show that the suffix *tu* as an agentive marker, deriving the verb to teach and to dance respectively into nouns or the doers of the partiular action.

4.2.2 Derivation by suffix *na.*

In Zyphe, nouns are derived from verbs by adding *na*. The suffix *na* primarily functions as a nominalizer, converting verbs into nouns. It serves to denote the name of an action or an entity associated with the verb.

9)	a)	*atsu*	learn	*atsu-na*	learning
	b)	*rei*	read	*rei-na*	reading
	c)	*tsia*	write	*tsia-na*	writing

Example:

10)	*tsa*	*tsia-na*	*le-tah*	*a-tɔh*	*ŋetei*
	book	write-NMLZ	LOC-FOC	3SG-good	very

She is good at writing.

Again, the nominalizer *na* derives a noun from an adjective.

11)	a)	*ziŋ*	dark	*ziŋ-na*	darkness
	b)	*niŋdo*	kind	*niŋdo-na*	kindness
	c)	*alɔŋ*	happy	*alɔŋ-na*	happiness
	d)	*ngetsia*	sad	*ngetsia-na*	sadness
	e)	*tʰɔdu*	weak	*tʰɔdi-nah*	weakness

Example:

12) *aluŋ-na-tsɔ* *a-vatsu-ka* *a-kʰuŋ*
happy-NMLZ-DEF 3SG-arrive-DECL 3SG-vanish
kɔwtei
repeatedly
Happiness comes and goes'.

In the sentence 12, the NMLZ marker *na* is attached to the adjective *aluŋ,* followed by a definite marker *tsɔ* which denotes happiness.

5.0 Compounding

5.1 Endocentric Compound:

An endocentric compound is a type of compound word where one part, known as the head, determines the primary meaning and also defines the grammatical category of the entire compound.

5.1.1 Left-Headed endocentric compound:

Left-headed endocentric compounds are observed, where the first element (head) determines the compound's overall meaning and grammatical category, while the second element (modifier) provides additional specification.

13)	a)	*tui+kʰɔŋ*	(water+rat hole)	= tuikʰɔŋ (water body)
	b)	*rua + kuŋ*	(big bamboo + tree)	= ruakuŋ (all kind of bamboo)
	c)	*lu+kʰu*	(head+cover)	= hat.
	d)	*thui+lukʰu*	(fruit+hat or helmet)	= A fruit that has a cap-like covering on one part.
	e)	*lu+raŋ*	(head+white)	= dandruff

Example-

14) a) *lukhu a-ŋepaw a-kʰu-tʰa*

hat 3SG-beautiful 3SG-wear-PFV

she has worn a beautiful hat.

Examples 13a, 13b, and 13c are all left-headed compounds. In each case, the second element modifies the first element, which is the head. The head noun *tui, lu, tʰui*, and *rua* are dependent on *kʰɔŋ, kuŋ, kʰu*, and *lukʰu* respectively, which alter the meaning of the compound word in each instance.

5.1.2 Right-headed endocentric compound

Right-headed endocentric compounds are also prevalent in this language. In these compounds, the second element (head) determines the compound's overall meaning and grammatical category, while the first element (modifier) provides additional specification.

15)	a)	*tʰiŋ+kuŋ*	(wood+stem)	= tʰingkuŋ (tree)
	b)	*me+tʰiŋ*	(fire+wood)	= metʰiŋ (firewood)
	c)	*tlaŋ+rai*	(mountain+work)	= tlaŋrai (social work)

Example:

16) a) *raŋlai-tah* *raaŋpaʂile-tah* *methiŋ*
ranglai-ERG jungle LOC-FOC firewood
a-hlaŋ-leimei
3SG-cut-PROG
Ranglai is cutting firewood in the jungle.

In sentence 16a, the compound word *methiŋ* is formed with two nouns, *me* and *thiŋ*. While both have distinct individual meanings, ***thiŋ*** act as the head and determines the overall meaning of the compound. The first element, ***me***, simply modifies the head, resulting in, ***methiŋ***, means firewood.

5.2 Exocentric compound

An exocentric compound is a headless compound where neither element functions as the head, and its overall meaning is not directly derived from its individual components.

17) a) *tlai+pia* (stamp+going out) = Tlaipia (wall)
b) *nu+vɔ* (woman+stream) = nuvɔ (couple)
c) *thiŋ+baihlɔ* (tree+banana) = thiŋbaihlɔ(Papaya)
d) *tsaŋ+tha* (finish+new) = tsaŋtha (youth)
e) *lai+sɔ* (big+meat) = *laisɔ* (spinster)

Example:

18) a) *khɔtei* *le-tah* *tsaŋtha* *tsai* *a-m̨aŋ*
village LOC-FOC youth time 3SG-spend
He spent his youth in the village.
b) *laisɔ* *uthui-tsɔ* *ama* *teitah* *iŋ teitei*
spinster old-DEF 3SG 3GS alone house small
le-tah *a-uŋ*
LOC-FOC 3SG-EXST
The old spinster lived alone in the small house.

In sentences 18a, and 18b, the compound words are headless. Neither element is functioning as the head in both examples. Each part contains its own meaning if we break down the elements as *tsaŋ* (finish) and *tha* (new) in 18a and *lai* (big) and *sɔ* (meat).

5.3 Conjunctive Compounds

In this type of compound, the nominal elements are joined using the conjunctive particle *le*.

19) a) *tsiŋ le zai* (day and night) = everytime

b) *nupui le pase*(wife and husband) = wife and husband

c) *bu le sɔ*(food and meat) = curry

d) *meh le ŋ̍ɔ* **(**fire and ear) = facial expression

Examples:

20) a) *kɔŋtsuana* *le-tah* *bu le sɔ* *hlupui* *a-uŋ*

feast LOC-FOC food and meat much 3SG-EXST

There are lots of curries in the feast.

b) *a-meh* *le ŋ̍ɔ* *a-siŋ* *ŋetei*

3SG -face and ear 3SG -clever INTENS

His gesture (facial expression) shows that he is very clever.

In sentences 20a and 20b, conjunctive compounds are formed by conjoining two nouns with the conjunctive particle ***le.*** In 20a, ***bu*** and ***sɔ*** have separate meanings, but when conjoined with ***le,*** they form a new word meaning "curry." Similarly, in 20b, ***meh*** and ***ŋ̍ɔ*** are conjoined with ***le*** to create a new meaning - facial expression.

5.4 Noun+noun

21) a) *luimai* (tax) *lui+mai* (soil+rate or price)

b) *luŋtʰiŋ* (mind) *luŋ+tʰiŋ* (heart+liver)

Example:

22) *kuŋtiŋtah* *ka-luimai* *ka-peh*

every year 1SG-tax 1SG-give

I pay my tax every year.

In sentence 22, the two independent nouns ***lui*** and ***mai*** have been combined, which means tax.

5.5 Noun+Verb

It is a kind of compound where noun and verb are joined together to form a different compound word.

23)

a)	*meh + bai*	(eye+to clip)	spectacles
b)	*pʰe + pakɔŋ*	(feet+ to wear)	shoes
c)	*lu+kʰu*	(head+ to cover)	hat
d)	*tui+so*	(water+ to rain)	waterfall
e)	*tʰiŋ+e*	(wood+ to eat)	log of firewood
f)	*tʰɔ+tʰi*	(strength+ to die)	lazy

Example:

24)	*zeletah-mɔ*	*ka phepakɔŋ*	*a-uŋ*	
	where-Q 1SG shoes		3SG-EXIST	
	Where is my shoes?			

In Zyphe, sentence 24 demonstrates the involvement of two different grammatical categories — a noun and a verb, which together form a new word with a different grammatical category. For instance, *pʰe* means 'feet' (a noun), and *pakɔŋ* means 'to wear' (a verb). When combined, they create a new noun meaning 'spectacles.'

5.6 Verb+Noun

In this compound, verbs are combined with a noun to derive a different word.

25)	*ve+hlo*	(visit+wage)	labour

26)	*vehlo-tsɔ*	*buitsuŋ*	*le-tah*	*rai*	*a-ʂe-leimei*
	labour-DEF	garden	LOC-FOC	work	3SG -work-PROG
	The labour is working in the garden.				

Zyphe also employs verb+noun construction together to form a new lexeme. Here in sentence 26, *vehlo* means labour. It is combined with two words verb *ve* means to visit and noun *hlo* means wage or salary. Together it denotes labour.

5.7 Noun+Adjective

27)	a)	*luŋ+lai* (heart+big)	stubborn
	b)	*rei+tsɔ* (enemy+good)	brave
	c)	*rei+tsia* (enemy+bad)	timid
	d)	*kaŋ+tsia* (side of body+bad)	feel shy
	e)	*vai+sang* (heaven+high)	sorrow/grief

In Zyphe, words can be formed by combining a noun and an adjective. The examples of 27 demonstrate the word formation through noun+adjective construction.

5.8 Adjective+Noun

28) *lɔŋ+mai* (happy+rate) Prize

Adjective + Noun constructions are relatively less productive as compared to other compounds in Zyphe.

6.0 Reduplication:

Reduplication is a morphological device that involves the repetition of a word or part of a word to express grammatical or semantic distinctions such as plurality, intensity, distribution, or repetition of action (Abbi, 1992). The following section will discuss the reduplicated structure of the Zyphe language.

6.1 Reduplicative Adverb

29) a) *karu~ru* = associated with dirt, which intensifies it.

b) *kalu~lu* = indicates black and dark.

c) *kala~la* = indicates very small.

d) *balu~lu* = indicates very dirty water.

e) *batle~tle*= indicates level of bitterness.

f) *bahla~hla*= indicates sourness

kalu~lu = Very black

30)	a)	*ma*	*stɔŋtsɔ*	*kʰy*	*a-vɔŋ*	*kalu~lu*
		DEM	person	DIST.	3SG-black	REDUP.

That person looks very black.

kaliŋ~ liŋ = Blackish

	b)	*ma*	*tsɔŋstɔ*	*kʰy*	*a-vɔŋ*	*kaliŋ~liŋ*
		DEM	person	DIST.	3SG-black	REDUP.

That person looks blackish.

kare~re is used with the word thin to intensify its meaning, like very thin.

c)	*ka*	*satei-tsɔ*	*a-zo*	*kare~re*
	1SG	sister-DEF.	3SG-thin	REDUP.

My sister is very thin.

Examples 30a, 30b, and 30c show how intensity or reduction of a thing or concept is expressed. These reduplicative adverbs do not contain any individual meaning but act to modify the intensity or degree of a particular thing or concept. In example 30a, kalu~lu appears before *vɔŋ* (black), indicating a strong or very black color. In example 30b, *kaliŋ~liŋ* changes the intensity of black to mean less black than *kalu~lu*. In example 30c**,** *kare~re* is used with any thin object or human being.

6.2 Reduplicative Verb:

In this reduplication process, the root of the verbs is repeatedly used, which indicate the continuity and iterativity of a particular action.

Example:

31) a) *e-e* : eat

ram-tah batui *e-e-ka* *e-tʰlu pai*

Ram-ERG rice eat-eat-DECL eat-finish-happen

Ram kept eating the food and finished it.

b) *y-y*: call

ka-no *y-y-ka* *ka-rei* *pai*

1SG-mother call-call-DECL 1 SG-tired happen

My mother got tired of calling me repeatedly.

c) *rei-rei* : read

tsa *rei-rei-ka* *a-ʂuh* *pai*

book read-read-DECL 3 SG-mad happen.

He became mad after reading a lot.

Example 31a shows the continuity of the action, while examples 31b and 31c indicate the repetition of an action for a long time.

6.3 Reduplicative Nominal

It is a common phenomenon in some Kuki-Chin languages. In reduplicative nominals, the same nouns or pronouns are repeatedly used.

Example:

32) a) *meru* *tle-pɔ-tah* *police-tah* *kʰua~kʰua-tah* *a: hui*

thief catch-NF-FOC police-ERG village village-FOC 3PL- search.

The police search village to village to catch the thief.

b)	*rai*	*a-chai~ chai-tah*	*a-ʂe*
	work	3SG-step step-FOC	3SG-work
	He did the work step by step.		

In Zyphe, reduplicative nominal constructions are formed by the repetition of two nouns followed by the focus marker *tah.* Example 32a) shows that the action takes place in every village. Example 32b) indicates a sequential process, meaning “one by one" or actions occurring in a step-by-step manner.

7.0 Onomatopoeia:

Onomatopoeia is a linguistic process where words are formed by imitating natural sounds, often found in expressive and ideophonic constructions in South Asian languages (Abbi, 1992). In Zyphe, the following onomatopoeic sounds indicate the plurality.

Example-

33)	a)	*miau miau*	-	sounds of cat.
	b)	*bu bu*	-	sounds of cow
	c)	*tada tada*	-	sounds of hen
	4)	*ŋhɔr ŋhɔr*	-	sounds of tiger
	5)	*ŋhɔ ŋhɔ*	-	sounds of snoring
	6)	*pʰai pʰai*	-	sounds of walking with heavy steps

Example-

34)	a)	*iŋ-kʰaŋ*	*tai-tah*	*pʰai-pʰai-tah*
		house-room	ABL-FOC	sound of walking-ADV
		a-pia		
		3SG-go out		
		He walked out of the room with heavy steps.		

Examples of 33 show the sound imitations used in the Zyphe language to represent specific actions or the sounds of animals and birds. In sentence 34a, the onomatopoeic word *phai-phai* is followed by the adjectival marker *tah*, which indicates the manner of walking of an individual.

7.0 Borrowing:

Borrowing is the process of adopting words from other languages due to linguistic contact, often for reasons of prestige or necessity.

7.1 Borrowing from English

English Word	Nativized in Zyphe
school	*sikul*
class	*kalas*
hall	*hɔl*
motor	*mɔtɔr*

Table:1.0

In this process, the words from English are directly borrowed with some phonological modification or adapted into their phonological pattern.

7.2 Borrowing from Hindi

Hindi Word	Nativized in Zyphe	Gloss
tsutti	*tsuti*	Holiday
istri	*istari*	Iron
tsabi tala	*sabi tala*	key and lock

Table: 2.0

As with borrowings from English, the lexical items borrowed from Hindi into Zyphe often undergo some phonological changes.

7.3 Borrowing from Mara

Mara Word	Nativized in Zyphe
daiti	time
su	Place
hlo	salary or wage

Table: 3.0

While borrowing from Mara, words are directly borrowed without any phonological changes.[3]

[3] Mara is a Kuki-Chin language and also a dominant language of Siaha district, Mizoram.

8.0 Conclusion:

This research primarily examines three-word formation processes in Zyphe: affixation, compounding, and reduplication. Zyphe employs a derivational prefix *pa* that converts nouns into verbs. The language also uses two derivational suffixes, *tu* and *na***,** which change the grammatical category when attached to verbs. Zyphe features various types of compounds, including endocentric, exocentric, and conjunctive compounds. Additional compound patterns include noun+noun, noun+adjective, verb+noun, and noun+verb constructions. Reduplication occurs in nouns, verbs, and adverbs, with adverbial reduplication being particularly notable. The agreement system in Zyphe is not only well-structured but also remarkably complex.

Abbreviation:

DECL	declarative
DEF	definite
DEM	demonstrative
DIST	distal
ERG	ergative
EXST	existential
FOC	focus
FUT	future
LOC	locative
NEG	negative
NMLZ	nominalizer/nominalization
PFV	perfective
PL	plural
PROX	proximal
SG	singular

Figure: 1.0 Genetic classification of the Maraic branch of the Tibeto-Burman language family.

Map: 1.0 Siata and Iana village of Siaha District, Mizoram.

Table 1.0 English words borrowed into Zyphe.

Table 2.0 Hindi words borrowed into Zyphe.

Table 3.0 Mara words borrowed into Zyp

वाक् मंथन
ISSN: 2426-2149

Challenges in Translating and Developing Teaching-Learning Materials for Tiwa and Rabha in Assam's MTB-MLE Programs[1]

Harshita Talukdar (Corresponding Author)
Arup Kumar Nath (Co-author)
Tezpur University
harshitatalukdar18@gmail.com
arupjnu@gmail.com

Abstract

This paper examines the primary challenges encountered in translating and developing effective school materials for the Rabha and Tiwa languages within Assam's Mother Tongue-Based Multilingual Education (MTB-MLE) program. This research is based on fieldwork and makes use of a qualitative approach to gather primary data from two major regions: Goalpara district (for Rabha) and Morigaon district (for Tiwa). To identify and understand the challenges in material development, the study involved conducting interviews and using structured questionnaires with key stakeholders from both communities. Participants included members of the local literary societies such as Rabha and Tiwa Sahitya Sabhas, translators who work on textbooks, and experienced retired teachers who have worked in schools in these areas. The study reveals several significant linguistic issues, such as textbooks are not originally written in Rabha or Tiwa. Instead, they are translated from Assamese, which often leads to errors and inconsistencies. Logistically, there are major gaps, such as a lack of trained teachers who can effectively teach in the mother tongue. Existing local resources, such as *Tiwa Mat Sigaina Naw* and the *Rabha Language Learning Textbook*, offer foundational grammatical sketches but are insufficient for building a comprehensive school curriculum. Addressing these foundational challenges in material development is essential for ensuring the success of MTB-MLE and promoting educational equity for Tiwa and Rabha children.

Keywords: MTB-MLE, Tiwa, Rabha, Linguistic Challenges, Material Development.

1. Introduction

1.1. Background Study

[1] This paper is a part of the ICSSR-sponsored research project titled ***"Locating Implementational Hurdles of Mother Tongue-Based Multilingual Education (MLE) in the Indigenous Languages of Assam, viz. Tiwa and Rabha, and Creating Grammar Teaching Modules and Digital Corpora."*** Under the Project Director, Dr. Arup Kumar Nath, Tezpur University

Mother Tongue-based Multilingual Education (MTB-MLE) is an educational approach that incorporates multiple languages into the learning process, emphasizing the learner's mother tongue as the primary medium of instruction, particularly in the early years of schooling. MLE aims to foster cognitive development, improve educational outcomes, and promote social justice by providing equitable access to education for speakers of minority and indigenous languages. It also seeks to preserve and revitalize linguistic diversity by integrating indigenous languages into formal education, thus helping to maintain cultural identity and intergenerational language transmission (Skutnabb-Kangas, 2009). According to Mohanty, Panda, and Skutnabb-Kangas (2009), the primary goal of MTB-MLE is to ensure that children develop literacy and cognitive skills in their first language before transitioning to other languages. This approach enhances educational outcomes and fosters a sense of pride and identity among learners who speak minority languages. One fundamental argument for MTB-MLE is its potential to promote social justice. Skutnabb-Kangas and Panda (2010) emphasize that MTB-MLE addresses the educational inequities faced by indigenous and minority language speakers, who often struggle in monolingual educational systems dominated by majority languages. By providing education in the mother tongue, MTB-MLE helps level the playing field, enabling these learners to achieve academic success and fully participate in society. Furthermore, MTB-MLE challenges the hegemony of dominant languages and promotes linguistic diversity as a resource rather than a barrier. In their work, Skutnabb-Kangas and Heugh (2010) argue that MTB-MLE is essential for sustainable diversity, as it supports the maintenance and revitalization of indigenous languages. They highlight that language is not just a means of communication but also a repository of cultural knowledge, traditions, and worldviews. Therefore, maintaining cultural diversity depends on maintaining linguistic diversity through MTB-MLE.

1.2. A brief introduction to the Rabha and Tiwa languages

The Rabha people are one of the significant plains' tribes of Assam. While they are of Mongoloid origin, their language belongs to the Tibeto-Burman branch of the Sino-Tibetan language family.

Grierson has placed them in the Bodo sub-group of the Assam branch of the Tibeto-Burman language family. Many tribes and sub-tribes within this language group have lost their native language, and today, almost all of them have adopted Assamese as their primary language. The Rabhas are one such tribe, which is divided into different clans or branches, which are Pati, Rangdani, Maitori, Totla, Kocha, Bitolia, Dahuri, and Sunga, among others (Goswami, 2004). The Rabha language lacks its script; therefore, the Assamese script has been adopted for writing this language. The Rabha people understand and speak Assamese, but their language is more similar to Garo and Boro than to other Tibeto-Burman linguistic groups. The Rabhas primarily reside in the Goalpara district, in the northern areas of the Garo Hills towards the south bank of the Brahmaputra in the Kamrup district. However, the Rabha population is also found in Dhubri, Kokrajhar, Bongaigaon, and Darrang districts, although their numbers are relatively small (Goswami, 2004).

The Tiwa are also an Indo-Mongoloid tribe and descendants of the Bodo family of the Tibeto-Burman branch of the Sino-Tibetan speech family, which are settled in Nagaon, Morigaon, Kamrup, Jorhat, Karbi Anglong, and Dhemaji districts of Assam and Ri-Bhoi district of Meghalaya. The neighbouring groups, such as Khasi-Jaintia and Karbi, have been referring to the Tiwas as "Lalung" (Patar, 2018)

The Tiwa language is a member of the Sino-Tibetan language family's Bodo-Naga subgroup of the Tibeto-Burman group. Lalung was the name given to them in ancient linguistic literature. In his 1891 Linguistic Survey of India, G.A. Grierson noted that there were perhaps 40,000 Tiwa speakers. However, the 1961 census and the 1971 census showed that the number had dropped to 10,576 and 9954, respectively. The main reason for this declining trend is the Tiwas' assimilation into Assamese life and culture. Nonetheless, 26,481 people speak Tiwa overall, according to the 2001 census. (Muchahary, 2014)

2.

2.1. Aims & Objectives

- To identify the linguistic challenges and hurdles faced in translating educational materials into Tiwa and Rabha.
- To explore the logistics hurdles in developing teaching-learning materials for these languages.

2.2. Research Methodology

The research adopts a qualitative approach and focuses on fieldwork to gather primary data and gain insights into the challenges faced in translating and developing educational materials for Mother-Tongue based Multilingual Education (MTB-MLE) in the Tiwa and Rabha languages in Assam.

The data collection was carried out in two districts of Assam, viz. Goalpara and Morigaon districts, keeping in mind their significant populations of Rabha and Tiwa speakers, respectively. The participants included a range of stakeholders from the communities involved in the MLE program such as members of local literary societies, such as Rabha Sahitya Sabha and Tiwa Sahitya Sabha, translators involved in the translation process of the textbooks for the elementary level, and a few retired teachers who are also members of the literary societies.

A structured questionnaire was designed to gather quantitative data on the challenges faced during the translation process and the development of teaching and learning materials. Interviews were also conducted with stakeholders, including literary committee members, translators, and educators.

3. Literature Review

The preservation and revitalization of indigenous languages, such as Rabha and Tiwa, through Mother Tongue Based Multilingual Education (MTB-MLE) has been an area of

growing academic interest. Many researchers have examined the value of MLE programs in advancing social justice and linguistic variety, as well as the difficulties in putting these programs into practice in indigenous communities. This study of the literature looks at important publications that shed light on the difficulties of translating educational materials into indigenous languages and the function of MTB-MLE in protecting them. In their edited volume *Multilingual Education for Social Justice: Globalising the Local (2009),* Mohanty, Panda, and Skutnabb-Kangas emphasize the role of multilingual education (MLE) in fostering social justice, particularly for marginalized and indigenous communities. They argue that MLE empowers learners by using their mother tongues as a medium of instruction, which enhances cognitive development and cultural identity. *Social Justice through Multilingual Education (2010)* by Skutnabb-Kangas and Panda focuses on the transformative potential of MLE in promoting linguistic human rights and equitable access to education. The editors present case studies from around the world, illustrating how MTB-MLE can address the educational disadvantages faced by speakers of minority languages, advocating for inclusive policies that integrate local languages into education systems. They argue that well-implemented MLE programs can protect endangered languages, promoting both cultural preservation and educational equity. *Tiwa Mat Sigaina Naw* by D. Uphing Maslai, *Tiwa Mat Sigaiane Laishap; and Mekdo-Tiwa Matkhop Phera* by Tiwa Mathonlai Tokhra (TMT), and *Rabha Language Learning Textbook* by All Rabha Sahitya Sabha discuss the basic grammatical sketch of the Tiwa and Rabha languages, respectively. However, since these textbooks cover just the basic grammar of the languages, these books are not sufficient to be used as full teaching materials in school.

4. Hurdles faced during the translation process

The implementation of Rabha as a medium of instruction in education officially began on March 19, 1988, with the appointment of 70 Lower Primary School (LPS) teachers in the Rabha-dominated areas, including Goalpara, Kokrajhar, and Dhubri, where students were exclusively Rabha speakers. Following this, the need for educational materials in Rabha led to the ongoing process of translating textbooks from Assamese into Rabha. Although the Rabha language employs the Assamese script, significant linguistic differences create challenges in translation, particularly for young learners. The Rabha Sahitya Sabha has played an instrumental role in translation efforts, including notable works such as the *Bhagavad Gita*. However, fundamental linguistic differences at the phonological, morphological, and semantic levels present difficulties in accurately translating educational content. This paper explores these linguistic challenges and their impact on Rabha-speaking elementary school students.

Although the Rabha language adopts the Assamese script, it has modified the vowel inventory by dropping and introducing some new vowels, creating significant phonological challenges for learners.

Fig 1: Rabha Vowel Phonemes (Joseph, 2007)

	Front	Central	Back
High	i	ɨ	u
Close-mid	e		o
Open		a	

Vowels	Front	Central	Back
High	i		u
Higher-mid	e		o
Lower-mid	ɛ		ɔ
Low		a	ɒ

Fig 2: Assamese Vowel Phonemes (Sarma, 2014).

The Rabha language presents a vowel system that shares similarities with Assamese but also exhibits crucial differences. The generally accepted vowel phonemes in Rabha include /i/, /e/, /a/, /o/, /u/, /ɔ/ (Joseph, 2007), and some linguistic analyses suggest the presence of an additional vowel, phonetically realized as [ɨ] (Sarmah, 2009). A significant distinction is Rabha's lack of phonemic nasalization, a feature present in Assamese (Joseph, 2007). These phonological divergences can create challenges for translators in finding precise phonetic equivalents between the two languages, particularly concerning the unique /ɨ/ in Rabha and the presence or absence of nasalization. This also leads to confusion for young learners who may struggle to recognize and pronounce Assamese vowels absent in their

native language. The special /ɨ/ sound in Rabha makes it difficult to link the sounds of the two languages directly. Therefore, when adapting Assamese words for Rabha, the absence of certain vowels requires us to alter the pronunciation, and this can sometimes change the meaning or make the words feel unnatural. Moreover, the use of the Assamese script with a reduced vowel set in Rabha creates difficulties in writing and spelling for students exposed to both languages. Assamese proper names, loanwords, and formal terms, which retain their full vowel inventory, lead to inconsistencies in spelling and pronunciation for Rabha learners. This phonological gap presents a considerable challenge in bilingual education, requiring systematic interventions to support students in effectively navigating these linguistic differences.

Similarly, the Rabha language has also reduced the consonant inventory from the Assamese consonant inventory. This modification poses several phonological challenges for learners, particularly in pronunciation, spelling, and language adaptation.

	Bilabial	Labiodental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Plosive	p b p^h b^h		t d t^h d^h			k g k^h g^h	
Nasal	m		n			ŋ	
Fricative			s z s^h z^h	ʃ		x	h
Approximant		ʋ	ɹ		j		
Lateral Approximant			l				

Fig 3: Assamese Consonant Phonemes (Sarma, 2014)

	Bilabial	Labiodental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Plosive	p b p^h b^h		t d t^h d^h			k g k^h g^h	
Nasal	m		n			ŋ	

Trill			r				
Fricative			s z				h
Approximant		ʋ			j		
Lateral Approximant			l				

Fig 4: Consonant Phonemes of Rabha (Brahma, 2014)

The lack of the voiceless velar fricative /x/ in Rabha presents a potentially significant challenge, as this sound is relatively frequent in Assamese (Mahanta, 2012). Translators might need to substitute /x/ with the closest available sound in Rabha, which could be /h/ or /k^h/, depending on the phonetic environment and the specific Rabha dialect being targeted. Such a substitution could lead to alterations in the pronunciation of words and might impact the recognition of loanwords that have been adapted into Assamese using this sound.

Furthermore, proper names and loanwords from Assamese into Rabha retain the dropped consonants from the Assamese alphabet, creating inconsistencies in spelling. Students may face confusion when writing Assamese words that require consonants absent in Rabha, as they might substitute them with similar-sounding letters or omit them entirely. This can lead to spelling errors and an inability to recognize words across both languages.

Rabha-speaking children, especially in early education, experience cognitive overload when learning to read and write in Assamese. Since Rabha is used at home and in lower primary schools, while Assamese is dominant in public spaces and higher education, students frequently switch between the two linguistic systems. This constant shift, combined with the missing consonants, leads to slower learning progress, increased spelling mistakes, and pronunciation difficulties in Assamese.

Since the Rabha language has dropped a few consonants from the original Assamese script, it creates significant challenges for young learners. As they grow and attempt to master both languages, they often struggle with proper word formation and grammatical accuracy. The absence of these consonants leads to difficulties in spelling, pronunciation, and morphological consistency. This linguistic gap results in errors when children try to apply Rabha phonetic patterns to Assamese or vice versa. Consequently, their bilingual learning process is marked by confusion and structural inconsistencies in both written and spoken communication.

For example:

English – Assamese - Rabha

Lettuce - লাই /laɪ/ - লায় /laj/

Linguistically /laɪ/ in Assamese and /laj/ in Rabha are different because the /laɪ/ ends in a diphthong, while the /laj/ ends with a palatal approximant.

Radish - মূলা /mula/ - মুলা /mula/

Even though the word /mula/ is phonetically identical, it differs in the distinct orthographies that Assamese and Rabha have.

Pea - মটৰ /mɔtɔɹ/ - মতৰ /motor/

The vowel quality in Assamese /mɔtɔɹ/ and Rabha /motor/is different, where Assamese employs more open vowels and Rabha uses mid vowels.

Chickpea - বুট /but/ - বুত /but/

Even though the word /but/ is phonetically identical, it differs in the distinct orthographies that Assamese and Rabha have.

Mynah Bird - মইনা /mɔɪna/ - ময়না /mojna/

The Assamese /mɔɪna/ features a diphthong, whereas the Rabha /mojna/ uses a monophthong followed by a consonant cluster.

Vulture - শগুন /xɔgun/ - ছুগুন /sugun/

In Assamese /xɔgun/, the initial consonant is a voiceless velar fricative, but in Rabha /sugun/, it contrasts with the voiceless alveolar fricative at the beginning, which is because of the absence of the voiceless velar fricative in the Rabha phoneme inventory.

Sage - ঋষি /rɪxɪ/ - ৰিছি /risi/

The usage of the velar fricative /x/ differentiates between the Assamese /rɪxɪ/ from the Rabha /risi/, which is absent in Rabha.

Vehicle - গাড়ী /gaɹɪ/ - গাৰি /gari/

The Assamese /gaɹɪ/ and Rabha /gari/are different because of the specific articulation of the rhotic consonant and the quality of the final vowel.

Rocket - ৰকেট /ɹɔkƐt/ - ৰকেত /roket/

Assamese /ɹɔkƐt/ uses an alveolar approximant and more open vowel sounds in comparison to Rabha /roket/, which uses the trill and the mid vowels. (ARSS, 2020)

Additionally, in Assamese, the usage of explicator compounds conveys certain meanings, while in Rabha, sometimes a single-word verb is employed to express the same idea. This difference highlights the linguistic variation between the two languages in terms of verb construction.

For example:

English	Assamese	Rabha
Take a bath	গা ধো /ga dʰu/	ৰু /ru/
Wake up	সাৰ পা /xaɹ pa/	বিছিঙি /bisiŋi/
Heat it up	গৰম কৰ /gɔɹɔm kɔɹ/	কুতুং /kutuŋ/
Stand up	থিয় হ /tʰɪo hɔ/	চাপ /sap/
Caress	মৰম কৰ /mɔrɔm kɔɹ/	হাছা /hasa/
Shout	চিঞৰ মাৰ /sɪoɹ maɹ /	আও /au/
Share /kokdek/	ভাগ কৰ /bʰag kɔɹ/	ককদেক
Bring forward	আগবঢ়াই আন /agboɹhaɪ an/	আদ্রায় /adraj/
Choose	পছন্দ কৰ /pɔsɔnd kɔɹ/	নুছুক /nusuk/
Solve	সমাধা কৰ /xɔmadʰa kɔɹ/	তীপ্রান /tɨpran/

Rabha-speaking children learning Assamese may struggle with the extra verb components and may omit auxiliary verbs, leading to grammatically incorrect sentences. Since children are exposed to both languages, they may struggle with verb formation, leading to errors in grammar, sentence structure, and meaning interpretation.

Textbooks frequently incorporate loanwords from various languages that might be pronounced using sounds not native to Rabha. If these loanwords have been assimilated into Assamese using the consonant sounds that are absent in Rabha, their pronunciation in Rabha will require careful consideration. Translators will need to adapt these words to fit the Rabha phonological system, potentially by substituting the missing sounds with the closest available equivalents. Similarly, the introduction of new concepts in textbooks might rely on Assamese words containing these unique consonants. It will be necessary for translators to locate already existing Rabha words or coin new terms (neologisms) that are both semantically correct in expressing the intended meaning and phonetically acceptable within Rabha.

Due to the structural differences of the two languages, translating textbooks from Assamese to Rabha is quite challenging. The Rabha-speaking students find difficulty in understanding and spelling certain words correctly because of the difference in vowel and consonant phonemes. In addition to that, the verb formation processes in each language call for careful attention to make sure the meaning is clear. The translators need to know both languages and understand their unique features to do this job well.

Similar to the Rabha language also faces the translation main difficulties is script for writing some pronunciations accurately.

Vowels	Front	Central	Back
High	i		u
Higher-mid	e		o
Lower-mid	ɛ		ɔ
Low		a	ɒ

	front	central	Back
High	i		u
Mid	e		o
low		a	

Fig 5: Tiwa phonemic vowel inventory (Dawson, 2020)

language, the Tiwa challenges during process. One of the the use of Roman Tiwa, which makes harder to capture

Fig. 6: Assamese Vowel Phonemes (Sarma, 2014).

	Bilabial	Labiodental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Plosive	p b p^h b^h		t d t^h d^h			k g k^h g^h	
Nasal	m		n			ŋ	

Fricative			s z s^h z^h	ʃ		x	h
Approximant			ʊ	ɹ		j		
Lateral Approximant				l				

Fig 7: Assamese Consonant Phonemes (Sarma, 2014)

	Bilabial	Labiodental	Alveolar	Postalveolar	Palatal	Velar	Glottal
Plosive	p b p^h		t d t^h			k g k^h	
Nasal	m		n			ŋ	
Affricate				ʧ			
Fricative			s z	ʃ			h
Approximant		ʊ	ɹ		j		
Lateral Approximant			l				

Fig 8: Tiwa phonemic consonant inventory (TMT, 2023)

Since both Assamese and Tiwa have different scripts, it gets difficult to retain the exact sound for certain words while translating. Even though the Tiwa language uses the Roman script, the alphabets are used based on its sounds, unlike the English language, which means that each letter in the Tiwa alphabet represents a specific sound, which can differ from the same letter being pronounced in the English language. The elementary level textbooks are translated to Tiwa from the Assamese language, but certain phonemes in Assamese are not present in English. As a result, the Tiwa language has developed its writing system to accommodate these sounds.

For example, the চ, ছ, শ, ষ, স (/s/, /ʃ/, /x/) from the Assamese alphabet are pronounced as Ch- চ্য, S- ছ, Sh- ছ্য in the Tiwa language.

Again, some pronunciations of the Tiwa language should be the pronunciation between two phonemes. Since it is not easy to distinguish between the two such characters, the Tiwa people pronounce either of these two characters in different places. This is probably because Tiwa is a tonal language, and thus the pronunciation is between two sounds. (Bordoloi, 2023)

The alphabets with such pronunciations are ক/গ (K/G), চ্য/জ (Ch/J), ত/দ (T/D), প/ব (P/B).

For example:

ক/গ (K/G) - কই/গই (/koi/ or /goi/) = Betelnut;

ছকল/ছগল (/sokol/ or /sogol/) = all of them

চ্য/জ (Ch/J) - চ্যু/জু (/ʧu/ or/zu/) = alcohol;

ত/দ (T/D)- তাও/দাও (/taʊ/ or /daʊ/) = Red;

তই/দই (/toi/ or /doi/) = curd

প/ব (P/B)- পিয়াদি/বিয়াদি (/pijadi/ or /bijadi/) leprosy;

পান্থাই/বান্থাই (/pant^{h}ai/ or /bant^{h}ai/) = brinjal

These different pronunciations add complexity to the translation process, as the translator needs to find ways to convey these sounds accurately while ensuring they make sense in the Tiwa language.

One of the major challenges in translating textbooks from Assamese to Tiwa is the limited availability of native Tiwa terminology. The Tiwa language lacks many technical and academic terms, which forces speakers to borrow words from Assamese or English. As a result, the Tiwa vocabulary includes a significant number of loanwords, especially for modern concepts and objects that do not have traditional equivalents in the language. This heavy reliance on borrowed words creates difficulties for translators, as they must decide whether to retain the original Assamese or English terms, modify them to fit Tiwa phonetics, or attempt to create new words that Tiwa speakers can understand. For example, words such as *department, research, training, school, rose, glass,* and *toothpaste* do not have direct Tiwa equivalents, so they are often used as they are or slightly modified to fit the Tiwa pronunciation style. This may cause issues for the young students, as they might struggle to adapt to foreign words that do not fit into the Tiwa linguistic patterns. When loan words are taken from Assamese, it is more complicated because Assamese is a non-tonal language, and there are differences in pronunciation and writing systems.

According to the basic grammatical sketch of Tiwa, *Mekdo-Tiwa Matkhop Phera* by Tiwa Mathonlai Tokhra, the language does not have the /c/ sound. The letter c is always pronounced as /ch/, which has a chance of creating confusion when learning new words. The textbook translators thus decided to retain the /c/sound at the elementary level, and

later, the students will relearn to pronounce it as /ch/ according to Tiwa phonetics. This adjustment is believed to maintain consistency in spelling while gradually teaching the correct pronunciation, but it also has chances of creating confusion in regards to clarity and uniformity in translation.

For example:

Chána thâwa (চ্যানাথাৱা): tasty

Chóma (চ্যমা): wet

Chuwa (চুৱা): tall

Chidi (চ্যিদি): if

Chongréng (চ্যংৰেং): thin

ʧin	et^{h}ase	ʧiŋe	k^{h}rom-ko	moʧoi-doŋ
1PL	just now	our	work-CL	finish-PFV

We have finished our work just now

pe	pe-ne	ʧurazin	p^{h}eu
3SG	3SG-POSS	room	come

She comes to her room

na	koi	ʧai	t^{h}ai-doŋ
2SG	betel-nut	chew	be-PFV

You have chewed the betel nut

While translating the textbooks from Assamese to Tiwa, one of the major challenges faced by the translators is the difference in tonal structures, where Tiwa is a tonal language, and Assamese is not. In tonal languages like Tiwa, the tone or pitch that is used while pronouncing a word can change its meaning completely. Since Assamese is a non-tonal language, similar-looking words in Tiwa can have different meanings based on tones. The translators must be very focused and ensure that the intended meaning is provided without change in natural tone patterns in Tiwa.

For example, if an Assamese word is translated to Tiwa, it might not convey the required meaning unless the correct tonal variation is applied. This creates a major problem for young learners at the elementary level, as they might get confused if the tonal differences are not properly marked. There are numerous loan words in Tiwa taken from Assamese, which have a chance of not fitting naturally into the tonal system, creating inconsistencies in pronunciation. Therefore, the translators must be highly cautious, and need to modify words so that they fit with Tiwa's tonal structure without changing the original meaning of the text.

5. Discussion

In the context of education and the translation process for the elementary textbooks, both the indigenous languages of Assam, Rabha, and Tiwa face many linguistic and cultural challenges. Since the textbooks are translated from Assamese, the process involves navigating numerous semantic and syntactic hurdles. One such challenge is finding direct equivalents for certain terminologies that will fit culturally with the language. This leads to a complex process of adapting and coining new terms to retain the original meaning while keeping in mind that the translation remains culturally relevant for the communities.

For the Tiwa language, the situation is complex because of the presence of two distinct speech varieties, namely Hajowali and Datowali. The semantic precision, along with the syntactic consistency, must be balanced by the translators to ensure that the translated materials remain applicable across both varieties. Due to this, it is difficult to produce text materials that accommodate speakers from both varieties. In addition to this, the lack of trained teachers who are fluent in the language and limited funds and resources available for these languages make it difficult in the implementational process. Moreover, the gradual shift of some of the Tiwa and Rabha speakers towards dominant languages such as Assamese in the region adds to further difficulty.

Apart from all these issues, the translation process is further affected by time constraints, limited availability of resources, and the lack of advanced linguistic tools. Even though the government policies emphasize the usage of the mother tongue in early education, there are many challenges in practically implementing it. One of the major hurdles is the lack of grammatical books for the elementary level, and folk tale collections or grammatical sketches available in these languages. The lack of teachers who are fluent in the language and are also trained in teaching young children is a major challenge. There are people in the community who speak the language but are not trained or have any experience in teaching young children. The available trained teachers do not speak the language, and thus, this gap makes it difficult to produce teaching-learning materials that are suitable for children. Since the population of these indigenous languages is small, they often get less attention and monetary support from the government when it comes to creating educational resources. It is also very important for translators to ensure that the translated materials

accurately reflect the culture, traditions, and day-to-day life of the children. Translating and creating such materials requires an ample amount of time and resources, and quality time with the community elders. The lack of basic tools for language support also poses a huge challenge. Many indigenous languages do not have large language models (LLMs), parts-of-speech tagging (PoS tagging), etc, that will help in building digital content for educational purposes. The software that will make the language available online is also lacking, making it difficult to have a digital presence. These logistical issues directly impact the accuracy of the translations, thereby complicating the production of high-quality materials.

6. Conclusion

Even though there are numerous significant hurdles, the implementation of MTB-MLE programs in the two indigenous languages of Assam, Rabha and Tiwa, provides an opportunity to preserve the languages, as it can elevate the status of these languages and make them stand out in various social domains. However, to fully utilise this potential, there is a need for more prioritised actions, which will include additional training for textbook translators and other associated personnel, as well as focused linguistic research and the creation of a standardised lexicon.

Joint efforts by the community members along with the government are important to overcome these hurdles for a successful implementation. With more support from the centre, the Rabha and Tiwa communities can flourish as vibrant languages in Assam, with their continuous presence in the state's linguistic and cultural heritage for the coming generations. Resolving the challenges in the translation process and the other logistical issues will primarily secure their place in the multilingual setup of the state.

List of abbreviations:

MTB-MLE: Mother-Tongue Based Multilingual Education

TMT: Tiwa Mathonlai Tokhra

SG: Singular

PL: Plural

CL: Classifier

PFV: Perfective

POSS: Possessive

List of Figures:

Fig. 1: Rabha Vowel Phonemes (Joseph, 2007)

Fig. 2: Assamese Vowel Phonemes (Sarma, 2014).

Fig. 3: Assamese Consonant Phonemes (Sarma, 2014)

Fig. 4: Consonant Phonemes of Rabha (Brahma, 2014)

Fig. 5: Tiwa phonemic vowel inventory (Dawson, 2020)

Fig. 6: Assamese Vowel Phonemes (Sarma, 2014).

Fig. 7: Assamese Consonant Phonemes (Sarma, 2014)

Fig. 8: Tiwa phonemic consonant inventory (TMT, 2023)

References:

All Rabha Sahitya Sabha. 2020. *Rabha language learning textbook.*

Bordoloi, D. N. 2023. *Mekdo-Tiwa Matkhop Phera*. Tiwa Mathonlai Tokhra.

Joseph, U. V. 2007. *Rabha*. Brill.

Mahanta, S. 2012. Assamese. *Journal of the International Phonetic Association* 42(02). 217–224.

Maslai, D. U. 2023. *Mukhli*. Tiwa Mathonlai Tokhra.

Maslai, D. U. 2023. *Tiwa Mat Sigaina Naw*. Tiwa Mathonlai Tokhra.

Mohanty, A. K., M. Panda & T. Skutnabb-Kangas (eds.). 2009. *Multilingual education for social justice: Globalising the local*. Orient BlackSwan.

Sarma, M. & K. K. Sarma. 2014. Phoneme-based speech segmentation using hybrid soft computing framework. Springer. https://doi.org/10.1007/978-81-322-1862-3

Sarma, P. 2009. Tone systems of Dimasa and Rabha: A phonetic and phonological study. Doctoral dissertation, University of Florida.

Sarma, P., M. Mitra, M. P. Bhuyan, V. Deka, S. Sarmah & S. K. Sarma. 2019. Automatic vowel recognition from Assamese spoken words. *International Journal of Innovative Technology and Exploring Engineering* 8(10). 2235–2239. https://doi.org/10.35940/ijitee.J9351.0881019

Skutnabb-Kangas, T. & K. Heugh (eds.). 2010. *Multilingual education and sustainable diversity work: From periphery to center*. Routledge.

Skutnabb-Kangas, T. & M. Panda (eds.). 2010. *Social justice through multilingual education*. Multilingual Matters.

वाक् मंथन
ISSN: 2426-2149

Sociolinguistic Profile of Magahi

Sachin Singhaniya &
Hari Madhab Ray
Jawaharlal Nehru University, Delhi

Abstract

This paper presents a sociolinguistic profile of Magahi. Magahi is an Indo-Aryan language spoken mainly in the state of Bihar but it is also spoken in the states of Jharkhand, West Bengal, and Orissa. Magahi language holds special status in the history of Bihar due to its rich historical and cultural significance. Magahi language has originated from Magadhi Apbhransha which was once official language or rajbhasha of ancient kingdom Magadha. Presently it co-exists along with Bhojpuri Maithili, and Angika language in Bihar but it does not enjoy the privilege of constitutional recognition. The people have negative attitude towards their language and the census data suggests language attrition. But on the brighter side, there has been extensive research on Magahi language focusing on grammar, phonetics and phonology, morphology, syntax, semantics, and computational aspects of language. But sociolinguistic aspect of this language is yet to be explored. This paper presents a sociolinguistic profile of Magahi by discussing the language history, geographical distribution, language variation, socio-linguistic aspects of language status, identity, and attitude.

Keywords: Socio-linguistic profile, Magahi, Language, Indo-Aryan, Bihar.

1. Introduction

A sociolinguistic profile typically includes an assessment of speaker demographics and geographic distribution; patterns of bilingualism or multilingualism; language attitudes and ideologies; degree of intergenerational transmission; and the level of institutional support or policy recognition. These elements collectively provide insight into the language's vitality, endangerment status, and ongoing processes of language maintenance or shift (Fishman, 1991; Grenoble & Whaley, 2006). This paper presents a sociolinguistic profile of Magahi by discussing the language, its history of literary tradition, geographical distribution, demographic profile, language variation, issues of identity and attitude and status of language. Magahi is an Indo-Aryan language historically rooted in the ancient Magadha region of present-day Bihar and Jharkhand, considered a direct descendant of Magadhi Prakrit—the court language of the Mauryan Empire. Despite being spoken by over 12 million people, Magahi lacks official status in India and is often classified as a dialect of Hindi, which has contributed to its marginalization in educational, administrative, and literary domains. The language has a rich oral tradition, including folklore, epic ballads, and seasonal songs, and its earliest known literary expressions can be traced to the Siddha poets of the 8th–12th centuries.

The etymological significance of Magahi lies in its direct connection to the ancient kingdom of Magadha, one of the most prominent political and cultural centers of early Indian civilization. The term Magahi is derived from Magadhi, which itself originates from Magadhī Prakrit, the vernacular language spoken in Magadha during the Mauryan period. This Prakrit form was the court and administrative language of Emperor Ashoka, as evidenced by his rock and pillar edicts spread across the Indian subcontinent. Over time, Magadhi Prakrit evolved into Apabhramsha, and eventually into Magahi, retaining the name that denotes its geographic and cultural origin. As Pushpadanta (959–972 CE) used the term Magah for Magadha in his Nayakumarchariu, the linguistic identity of the region was clearly established. The name Magahi, therefore, carries more than just linguistic relevance—it symbolizes a continuity of historical, cultural, and political identity from ancient to modern times.

Contemporary Magahi is believed to have evolved from Vedic Sanskrit following the path:

Figure 1: Evolution chart of modern Magahi

2. Script and literary tradition

Magahi language was spoken in large geographical area which included the states of Bihar, Jharkhand, Orissa, and Bengal. The standard or central variety of Magahi is used in the districts of Patna, Gaya, Nalanda and Rajgir and the other places has different variety. Since the Magahi language had different variety, there used to be multiple script for writing in Magahi. Verma (2007) mentions four distinct scripts namely Kaithy, Bangla, Oriya, and Devnagari. "The traditional script for Magahi has been Kaithi, which is still used in personal communication and sometimes in semi-legal transactions. The Kaithi script gets its name from the word 'Kāyath' < 'Kāyastha', the caste of writers in Northern India, and has been current until recently in Bihar and Eastern Uttar Pradesh. It derives from early Nagari and is closely related to the current Gujrati script which replaced the Devanagari only in the nineteenth century." Today, Devnagari is the standard script which is widely accepted and mostly used in not only Magahi but many other Indo-Aryan languages like Maithili, Bhojpuri, Hindi etc.

Magahi literature, with its roots in the linguistic continuum of Vedic Sanskrit through Prakrit and Magadhi Apabhramsha, represents a significant yet often overlooked component of Indo-Aryan literary traditions. The earliest literary manifestations of Magahi appear during the Siddha period (8th–12th centuries CE), a golden age characterized by the mystical poetry of Vajrayana Buddhist saints. Sarahapa, regarded as the foundational figure of Magahi literature, authored aphoristic verses compiled in the *Doha Kosh*, a work later edited by Rahul Sankrityayan and indicative of the language's early literary potential. This period saw the emergence of over 80 Siddha poets, many of whom were based in the scholarly centres of Nalanda and Udantpuri. However, the subsequent medieval era (1200–1600 CE) marked a pronounced decline in written output due to socio-political disruptions, including the destruction of Nalanda and the rise of Persian, Urdu, and later English as dominant languages of administration and culture. Nevertheless, oral traditions such as epic ballads (*Lorikaen*, *Raja Dholan Rani Maruan*), seasonal songs (*Barahmasa*, *Jitiya*), and narrative

folklore (*Biraha*, *Khissa*) preserved Magahi's literary vitality. A renewed literary consciousness emerged in the 19th and early 20th centuries with figures such as Jainath Pati, who authored the first Magahi novel *Sunita* and translated contemporary political texts, and Krishnadeo Prasad, whose poetry and essays (*Baanchi Le More Patiya*, *Mantarang*) helped shape modern Magahi prose and verse. The post-independence era witnessed an institutionalization of literary efforts through organizations such as the Bihar Magahi Mandal and Akhil Bhartiya Magahi Sahitya Sammelan, which facilitated the publication of journals (*Bihan*) and over 70 books. Prominent authors like Dr. Ram Prasad Singh (*Loha Marad*, *Magahi Sahitya ka Itihas*), Kesari Kumar (*Kesari Kumar Ke Magahi Kavita*), and Dr. Janardan Mishra "Gopal" (noted for his Sanskrit-Magahi verse translations) expanded the genre spectrum of Magahi literature.

3. Geographic Distribution

Magahi language is spoken in the Indian state of Bihar, Jharkhand, West Bengal, and Orissa. It is the second official language of Jharkhand and one of the major languages of southern Nepal. In Bihar, the region around Patna district and the areas below ganga river is the region of central Magahi variety. The area is surrounded by different languages of Bihar as it has Bhojpuri in the west, Maithili in the north, Bengali in the east and Khortha in the south. Sinha (2023) provides the most recent details regarding the geographical distribution of Magahi language- "The current form of Magahi is spoken in the area comprising ten districts of South Bihar (Patna, Jahanabad, Arwal, Aurangabad, Gaya, Nawada, Nalanda, Shekhpura, Lakhisarai, and Jamui), ten districts of Jharkhand State (Devghar, Jamtara, Dhanbad, Bokaro, Giridih, Kodarma, Hazaribag, Chatra, Palamu, and Wn. Singhbhum), one district, Maldah, of West Bengal, and one district, Keonjhar (now Kendujhar) of Odissa. However, beyond the boundary of Bihar, the influence of the tribal languages caused some differences between the Magahi spoken in the core area and the Magahi spoken in the peripheries (especially in Jharkhand) and the outliers (in West Bengal and Odisha)"

Figure 2: Map showing Magahi speaking area. (Source: Sinha, R.N.P. (2023). Preservation of Magahi Language in India: Contemporary Developments. In: Brunn, S.D., Kehrein, R. (eds) Language, Society, and the State in a Changing World. Switzerland: Springer.)

4. Magahi Language Use in Social Domains/ Use of Magahi in various social domains

4.1 Educational domain

The status of Magahi in the educational domain remains limited, particularly at the primary and secondary levels, where Hindi and English dominate the formal curriculum across Bihar and Jharkhand. Despite its historical significance and rich oral traditions, Magahi has not yet been incorporated into mainstream school syllabus. A notable institutional advancement occurred in 2010 with the establishment of the Department of Magahi at Magadh University, Bodh Gaya. The department has emerged as a crucial academic space for the study and development of Magahi language and literature. Several important literary works have contributed to the academic corpus of Magahi. These include *Kesari Kumar ke Magahi Kavita*, a modern poetic collection that elevated Magahi to literary parity with contemporary Hindi poetry, and *Kalam Aa Sapna*, a collection of essays co-edited by Kesari Kumar and Ram Nandan. Additionally, Dr. Janardan Mishra "Gopal," a distinguished scholar and translator, has made substantial contributions by rendering classical Sanskrit texts, including the *Patanjal Yoga Sutra* and *Valmiki Ramayan*, into classical Magahi verse. A range of journals has historically supported Magahi literary activity, beginning with *Magadhi* (1950s), edited by Shrikant Shastri, and continuing through *Bihan*, the literary magazine of Bihar Magahi Mandal that has served as a vital platform for over five decades. The digital periodical *Magahi Manbhawan* has been issued on a monthly basis by its editor, Uday Kumar Bharati, since 2011. Narayan Prasad maintains a blog titled *magahi-sahitya.blogspot.com*, which functions as a comprehensive repository of Magahi literature and linguistic resources. In addition to curating content, he has translated the Kannada novel *Vijeta* and six Russian literary works into Magahi. Complementing his efforts are other valuable digital platforms, such as Mrityunjay Kumar's blog *magadhi-magadh.blogspot.com*, which offers additional insights into Magahi language and culture, and Dr. Lakshman Prasad's blog *lprasadjsr.blogspot.com*, presents various aspects of Magahi literature. Nevertheless, the absence of Magahi from formal educational curriculum and the lack of officially sanctioned textbooks underscores the continuing marginalization of the language within institutional education.

4.2 Religion & Cultural tradition

Magahi, derived from Magadhi Prakrit, was the vernacular language used by Gautama Buddha to spread his teachings, deliberately chosen over Sanskrit, which had become formal and exclusive to orthodox rituals. This made religious ideas accessible to the common people across regions like Magadh, Kashi, and Videh (Aryani, 1965). Today, Magahi's religious significance continues through its rich tradition of folk songs and dances performed during various festivals and life events. Songs like *Phaag* and *Chaita* mark seasonal celebrations in Phalgun and Chait months, while *sohar*, *khilona*, *kohbar*, and *sumangali* are sung during births, marriages, and other important ceremonies. These cultural expressions play a vital role in preserving the community's heritage and spiritual life.

4.3. Politics and Law

In politics, Magahi is often used by regional leaders in Bihar and Jharkhand to connect with local voters, especially during rallies and campaigns. Using Magahi helps politicians build trust and appeal to cultural identity at the grassroots level. However, Hindi and English dominate official government and legal settings, limiting Magahi's role in formal politics. Despite this, Magahi remains important for everyday political communication and local engagement.

4.4. Media

The presence of Magahi in media remains limited but notable, with efforts spanning print, film, and digital platforms. Historically, the weekly newspaper *Magahi Samachar* launched in Patna in 1978, though it lasted only a few months, marking an early attempt at Magahi journalism. In cinema, Magahi has produced a handful of films, beginning with *Bhaiya* (1964) and *Morey Man Mitwa* (1965), followed by more recent releases like *Ajgut Joddha* (2015), *Dewan Misir* (2018), and *Bidhna Nach Nachabe* (2018), reflecting a slow but steady growth in regional filmmaking. On digital platforms, YouTube channels showcase Magahi folk music, including songs by artists like Tilak Chadhave and Tiwari. In recent years, social media has emerged as a vibrant space for the informal promotion and preservation of Magahi. Platforms like YouTube, Facebook, and blogging websites have allowed speakers to express themselves freely in Magahi through videos, vlogs, poetry recitations, and cultural commentary. Creators use Magahi to produce folk song performances, local storytelling, and comedic sketches that reflect the socio-cultural life of the region. These media efforts contribute to preserving and promoting Magahi language and culture, though the language's overall visibility in mainstream media remains modest.

4.5. Tourism Industry

In the tourism sector, Magahi contributes to cultural authenticity, particularly in heritage sites like Bodh Gaya, Rajgir, and Nalanda. While formal communication occurs in Hindi and English, Magahi is commonly spoken by local guides and vendors, enriching visitor engagement. Regional specialties like *Magahi paan* serve as both culinary attractions and cultural markers, promoting local identity through language.

4.6. Health Care

In the healthcare domain, Magahi is primarily used in informal interactions between medical staff and patients, especially in rural and semi-urban areas of Bihar and Jharkhand where it is widely spoken. While official documentation and prescriptions are written in Hindi or English, healthcare workers often rely on Magahi to explain diagnoses, treatment procedures, and medication instructions to ensure better patient understanding. This vernacular use facilitates effective communication and trust, particularly among non-literate or elderly patients. However, the absence of Magahi in formal medical training, resources, and public health campaigns highlights its marginal status within institutional healthcare settings.

4.7. Public Transport

In the domain of public transport, Magahi is widely used as a functional vernacular, especially across the Magahi speaking region of Bihar. Bus conductors, auto-rickshaw drivers, street vendors, and passengers commonly use Magahi for everyday transactions, directions, and interpersonal interactions. Beggars use Magahi, often singing traditional folk songs to appeal emotionally and ask for help. While Hindi is used for signage and official announcements, Magahi facilitates smoother communication at the grassroots level, fostering a sense of familiarity and accessibility among local commuters. Despite its prevalent spoken use, Magahi has not been institutionalized within the formal public transport system.

4.8 Demographic profile of Magahi

The demographic profile of Magahi reflects its position as a major regional language in eastern India, primarily spoken in the states of Bihar, Jharkhand, and to a lesser extent, in parts of West Bengal and Odisha. According to the 2011 Census of India, approximately 12.7 million people identify Magahi as their mother tongue, though the actual number may be higher due to its frequent classification under Hindi in official records. The language is predominantly spoken in ten districts of South Bihar—including Patna, Gaya, Nalanda, Jehanabad, Arwal, Aurangabad, Nawada, Lakhisarai, Shekhpura, and Jamui—and in ten districts of Jharkhand, such as Hazaribagh, Chatra, Koderma, Giridih, Dhanbad, and Bokaro. Smaller pockets of Magahi speakers are also found in Malda (West Bengal) and Keonjhar (Odisha).

Magahi is most commonly used in rural and semi-urban areas, where it functions as the primary medium of informal communication, while Hindi dominates formal, educational, and administrative contexts. The language is primarily spoken by members of agricultural and artisan communities, but its use cuts across caste and economic divisions. Significant diaspora communities of Magahi speakers also exist in countries such as Fiji, Mauritius, Trinidad, Suriname, and Guyana, the result of 19th-century indentured labor migrations. Despite its large speaker base, the lack of official recognition and institutional support has led to concerns about language shift, especially among the younger generation.

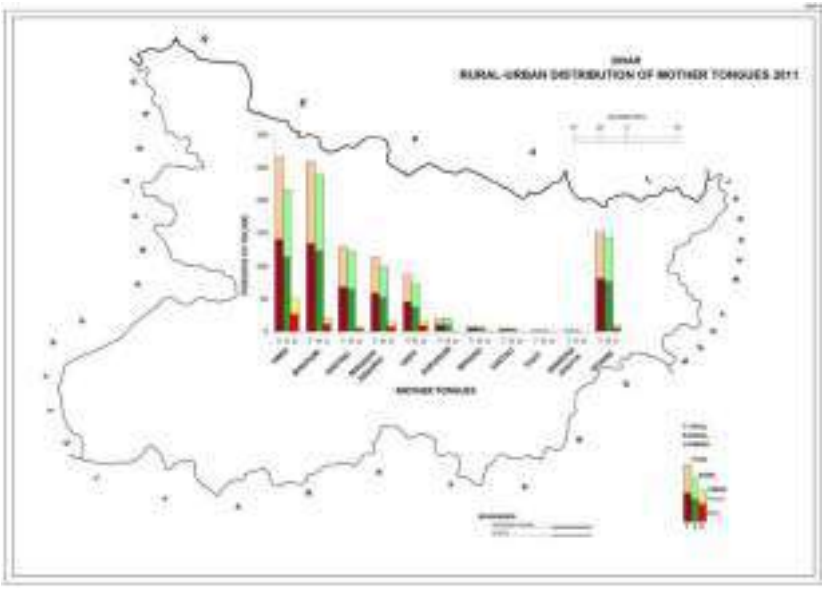


Figure 3: Map showing rural urban mother tongue distribution of different languages in Bihar (source: www.censusindia.gov.in)

5. Language variation

Magahi, as a regional language, displays considerable dialectal differentiation influenced by geographic, cultural, and administrative factors. Pandey (1980) classifies Magahi into four primary varieties: Standard, Eastern, Western, and Mixed forms, while earlier scholars such as Grierson (1927) and Aryani (1965) identified three: Standard, Eastern, and Mixed. Among these, Eastern Magahi is the most linguistically diverse and includes several distinct local dialects. Grierson records varieties such as Kurmali (spoken in Mayurbhanj and Bamra), Khontai (western Malda), Kurmali Thar and Khotta (Manbhum region), and Pargania or Tamaria (southeastern Ranchi). These dialects are primarily spoken along the linguistic borders

of Bihar with West Bengal and Odisha, and in adjacent regions of Jharkhand, where contact with Bengali and Odia has significantly influenced their phonology, lexicon, and script. Despite their divergent names and localized characteristics, these dialects are unified under the broader umbrella of Eastern Magahi, reflecting the language's adaptability and its deep entrenchment in the socio-linguistic landscape of eastern India.

Historically, Magahi has been classified in different ways by different scholars. While Grierson (1903) puts Magahi under the Eastern group of Outer sub-branch of Indo- Aryan languages, others like Turner have clubbed the 'Bihari' languages with Eastern and Western Hindi (Masica 1991). A classification given by Chatterji (1926) where Magahi is kept together with other languages of Eastern group which is separate from the Western Hindi. Jeffers (1976) gives a classification which is very similar to that of Grierson. In the present time, Magahi is spoken mainly in Eastern states of India including Bihar and Jharkhand, along with some parts of West Bengal and Orissa. There are three main varieties of Magahi spoken today (Verma, 1991).

• Central Magahi of Patna, Gaya, Hazaribagh (in Bihar)

• South-Eastern Magahi of Ranchi (in Jharkhand) and some parts of Orissa

• Eastern Magahi of Begusarai and Munger (in Bihar)

Variation in Kinship Terms in Magahi

The kinship system in Magahi reflects the broader patterns of Indian kinship, particularly the descriptive system, which is characterized by precise, differentiated terminology for various relatives. As part of a patrilineal and patrilocal social structure, Magahi employs a rich lexicon to distinguish between consanguineal and affinal relations. This kinship nomenclature is deeply embedded in the region's cultural fabric and is indicative of both social hierarchy and gendered roles within the family.One of the distinctive features of the Magahi kinship system is its variation across generations, particularly in terms of phonology, lexicon, and semantics. For instance, the same kin relation may be referred to using different terms or pronunciations by different generations or age groups.

- Phonological variations in Magahi are seen in substitutions such as /r/ with /l/ (e.g., *sara~sala* for wife's brother.
- Lexical variation includes multiple terms for the same relation. For example,

Father	/pɑːpɑː/, /pəppɑː/, /bɑːbʊ/
Mother	/mɑːj/, /məmmɪ/
Grandfather	/dɑːdɑː/, /bɑːbɑː/
Grandmother	/dɑːdɪ/, /mɑːmɑː/

- Semantic variation is also evident, with terms like *mama* ambiguously used for both 'mother's brother' and occasionally for grandmother, depending on context.

6. Identity and Attitude

The identity and linguistic attitudes of Magahi speakers are shaped by a complex interplay of cultural heritage, socio-political marginalization, and evolving regional consciousness. Despite being the mother tongue of millions, Magahi has historically lacked formal recognition, often subsumed under Hindi in censuses and administrative policies. This has led to a sense of linguistic invisibility among its speakers, who, while deeply connected to their language in cultural and emotional terms, frequently switch to Hindi or other dominant languages in formal or educational contexts. However, recent decades have seen a revivalist sentiment among the Magahi-speaking community, marked by increased literary activity, grassroots mobilization, and advocacy for inclusion in the Eighth Schedule of the Indian Constitution. This linguistic reassertion reflects not only a desire for institutional validation but also a broader reclamation of cultural identity. While older generations may associate Magahi with rurality or limited prestige, younger speakers and intellectuals increasingly view the language as a marker of regional pride and historical continuity, reflecting a gradual shift from linguistic self-doubt to cultural affirmation.

7. Status of Magahi language

The neglect of Magahi by the government is evident in its exclusion from educational and official recognition frameworks. Before independence, Magahi was included in school curricula, with poems featured in Patna University's 1943 matriculation syllabus. Post-independence, however, Magahi was removed from textbooks and denied recognition as a distinct language in the 1961 Census, being subsumed under Hindi. This erasure affected millions of native speakers by undermining their linguistic identity. Unlike Maithili—which was recognized by the Sahitya Academy in 1965 and included in the Eighth Schedule of the Indian Constitution in 2003—Magahi still lacks constitutional status. As a result, it is excluded from major academic and administrative opportunities, including competitive exams in Bihar. This unequal treatment has prompted the Magahi-speaking community to push for both scholarly representation and political recognition. Despite having over 12 million speakers (2011 Census), Magahi remains constitutionally unacknowledged and institutionally marginalized.

Grierson, referring to the 1881 census, estimated the number of Magahi speakers at around 6,504,817. By 1951, this number had drastically dropped to just 3,728. A modest rise was noted in the 1961 census, which recorded 3,792,447 speakers. In 2001, the speaker population reportedly increased to 10,566,842 (Abbi 2000:15). The steady reduction in speakers over time can be attributed to several factors. The growing influence of widely spoken and socially dominant languages in the region has significantly impacted Magahi. Economic hardship has also forced many native speakers to migrate to other parts of the country in search of jobs, food, housing, and improved living conditions. This relocation has distanced the younger generation from their linguistic heritage, leading to a gradual loss of fluency. Moreover, there is a strong tendency among Magahi speakers to shift toward more accepted and advantageous languages like Hindi and English, further contributing to the erosion of their native tongue. However, revitalization efforts in recent decades through literature, education, and media have begun to strengthen Magahi's demographic and cultural visibility.

8. Conclusion

This paper reflects detailed exploration of the various sociolinguistic facets of the Magahi language. It traces geographical distribution of Magahi across Bihar, Jharkhand, West Bengal, and Odisha and highlights Magahi's use across diverse social domains despite its marginalization. The exploration of linguistic variation reveals rich regional dialects that attest to the language's vibrant oral traditions. An overview of its literary history underscores Magahi's significant yet often overlooked cultural contributions. Analysis of its current status reveals challenges arising from the lack of constitutional recognition and institutional support. Furthermore, investigation into speaker identity and attitudes exposes a complex interplay between cultural pride and language stigma, which affects language maintenance. Together, these findings underscore the urgent need for policy interventions and community efforts to preserve and revitalize Magahi. By addressing these sociolinguistic dimensions, this paper fills an important gap in research and highlights the critical importance of supporting Magahi within India's multilingual framework.

References

Aryani, S. (1965). *Magahi Lok Sahitya*. Patna: Hindi Sahitya Sansar.

Aryani, S. (1976). *Magahi Bhasha aur Sahitya*. Patna: Bihar Rashtrabhasha Parishad.

Census of India. (2001). Table No.- C 16. Registrar General of Census of India.

Chatterji, S.K. (1926). *The Origin and Development of Bengali Language*. Calcutta: Calcutta University Press.

Chatterji, S.K. (1963). *Bhartiya Arya Bhasha aur Hindi*. Delhi: Rajkamal Prakashan.

Fishman, J. A. (1991). *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*. Clevedon: Multilingual Matters.

Grenoble, L. A., & Whaley, L. J. (2006). *Saving Languages: An Introduction to Language Revitalization*. Cambridge: Cambridge University Press.

Grierson, G.A. (1903-28). *Linguistic Survey of India*. Vol 1-11. Calcutta. Reprint: (1968). Delhi: Motilal Banarsidas.

Holmes, J. (2013). *An Introduction to Sociolinguistics* (4th ed.). New York: Routledge.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Pandey, R.B. (1976). *Magahi Sahitya va Sahityakal*. Gaya: Lok Sahitya Sagar.

Pandey, R.B. (1980). *Magahi Bhasha ka Itihas*. Gaya: Lok Sahitya Sagar.

Sanskrityanan, R. (1993). *Pali Sahitya ka Itihas*. Delhi: Vani Prakashan.

Sinha, R.N.P. (2023). Preservation of Magahi Language in India: Contemporary Developments. In: Brunn, S.D., Kehrein, R. (eds) Language, Society, and the State in a Changing World. Switzerland: Springer.

Verma, S. (2007). Magahi. In G. Cardona and D. Jain (Eds.), The Indo-Aryan Languages, pp 547-565. London: Routledge.

वाक् मंथन
ISSN: 2426-2149

Case Marking in Pagro Mising

Normoda Doley
JNU, Delhi
doleynormoda@gmail.com

Abstract

Case marking in the Pagro dialect of Mising shows overt and covert markers closely following Blake's (2001) definition of Case being a system that marks dependent nouns with respect to their heads. Case used to be limited to inflectional markers but is now applied to wider syntactic functions. Pagro Mising shows case marking on proper nouns, kinship terms, pronouns, and common nouns, although there is variation in overt and covert realization. It has an elaborate case system with *accusative, dative, instrumental, ablative, genitive, locative, benefactive, allative,* and *vocative* cases (Taid, 2016). These markers indicate the noun's syntactic and semantic function within the clause or phrase, marking relationships to verbs, postpositions, or other nouns. This study explores the distribution and function of these case markers, contributing to the understanding of case systems in Tibeto-Burman languages.

Keywords: Pagro Mising, Case, Tani language, Trans-Himalayan, Morphology.

1. Introduction

Mising (ISO 639-3: mrg), a member of the Eastern Tani branch of the Tibeto-Burman (Trans-Himalayan) language family, was recorded as the native language of 687,836 speakers according to the Census of India (2011). Sun (1993:12) noted that the Mising community, historically called the "Plains Miris," has traditionally inhabited the northern regions of Assam. Their sustained interaction with the Indosphere has resulted in considerable linguistic influence from Indic languages, setting Mising apart from other Tani languages that have remained relatively insulated from such external influences. This linguistic divergence underscores the complex interplay of linguistic and cultural dynamics within Assam, a region characterized by its rich mosaic of language communities. The Mising-speaking population is primarily concentrated in the Upper Assam districts of Dhemaji, Lakhimpur, Tinsukia, Dibrugarh, Sivasagar, Jorhat, Majuli, Golaghat, Biswanath, and Sonitpur.

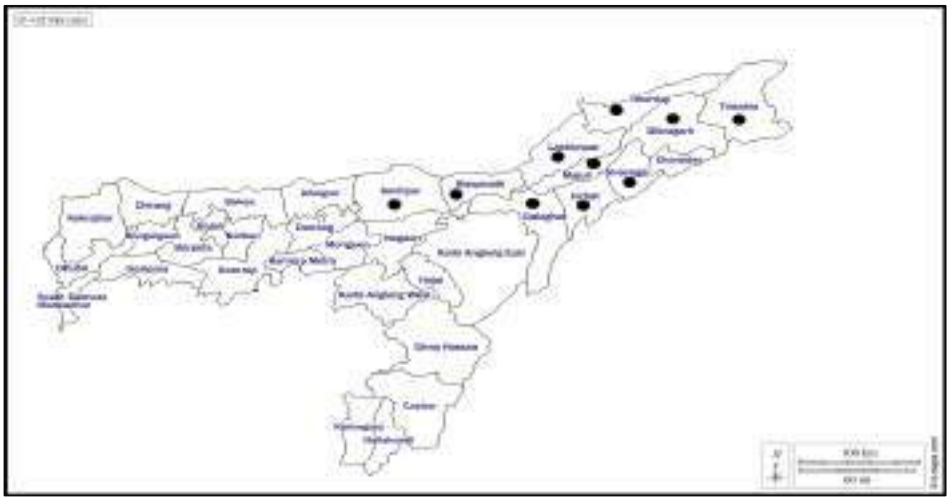


Figure 1: Map of the Mising population in the state of Assam.

This study examines case marking in Pagro Mising, building on previous detailed analyses conducted for a specific variety of Mising. The language exhibits a range of case markers, including *accusative, dative, instrumental, ablative, genitive, locative, benefactive, allative, and vocative* (Taid, 2016). The data is analyzed in such a way that the Roman script is transcribed using IPA, with each entry followed by a phonetic transcription and an English translation.

2. Literature Review

Case marking plays a crucial role in shaping the grammatical structure of languages, helping to indicate how nouns relate to other elements in a sentence. Blake (2001:1) describes *case* as "a system of marking dependent nouns for the type of relationship they bear to their heads." Traditionally, the term was used to refer specifically to inflectional markers, but case marking can also function at different linguistic levels—linking nouns to verbs in clauses or to prepositions, postpositions, or other nouns within phrases.

In the Mising language, particularly in its Pagro dialect, case markers appear in both overt and covert forms. Taid (2016:137) provides a detailed account of case marking in Mising, identifying ten cases: **nominative, accusative, dative, instrumental, ablative, genitive, locative, benefactive, allative, and vocative**. These markers attach to various noun types, including proper nouns, kinship terms, pronouns, and common nouns, each showing distinct case-marking patterns. Some markers are explicitly expressed through morphological inflections, while others are inferred through syntactic structures.

The study of case marking in Mising reveals its intricate system and offers insights into how the language organizes grammatical relationships. Given its complex patterns and typological significance, further research could shed light on its role within the broader Trans-Himalayan language family.

3. Methodology

This study adopts a descriptive linguistic approach to analyze *Case Markers* in Mising, with a specific focus on the Pagro variety. Given the absence of dedicated research on case marking in Pagro Mising, this study aims to provide a systematic account of its case system based on primary data collection and linguistic analysis.

Data Collection

- **Primary Data:** The study relies on fieldwork (Author, unpublished manuscript) conducted with native speakers of Pagro Mising. Data is collected through elicitation sessions, narratives, and spontaneous speech to comprehensively represent case usage in natural discourse.
- **Secondary Data:** Existing literature on Mising (e.g., Taid 2016) and related Trans-Himalayan languages is reviewed to contextualize the findings within a broader linguistic framework.

4. Subject Marking

Recent research in the typology of case marking and grammatical relations has increasingly focused on **variable, optional, or pragmatically conditioned marking** of Agent, Subject, or A arguments (Chelliah & Barðdal, 2009; Chelliah et al., 2011). In Tibetan, for instance, subject marking is not syntactically obligatory but is strongly influenced by pragmatic factors (Lehman, 1973). Similarly, Coupe (Genetti et al., 2008) observes that in Mongsen Ao, agents in generic statements of habitual activity must be explicitly marked.

In the Pagro variety of Mising, subject marking is predominantly **covert**, with overt markers appearing **optional** in generic references but **obligatory** in specific references. Within the **Tani language group**, to which Mising belongs, object marking follows an **accusative pattern**, while subjects typically remain unmarked (Walling, 2010). However, certain markers, such as those indicating **generic reference, topic, and third-person agreement**, are attested in proper nouns (restricted to persons). Notably, the **third-person marker** occurs with subjects in transitive and intransitive constructions, further highlighting the interaction between subject marking and discourse-pragmatic factors in Pagro Mising.

Taid (2016) examines the generic and specific markers of Mising within the framework of the nominative case. This study builds on that analysis by presenting a revised account of **subject marking** in Pagro Mising, offering a more nuanced perspective on its distribution and function.

1.1 **Common Nouns**

In Pagro Mising, common nouns may be marked with the generic marker /jə/ or /ə/ when used in a *generic reference;* however, this marking is **not obligatory**. In contrast, when a noun is used in a *specific reference*, it obligatorily takes the specific subject marker /də/.

1.1.1 Generic reference

1. pəttaŋ-**ə** dəː-dag [elicited]
   birds-GEN fly-DECL
   "Birds fly."

2. pəttaŋ dəː-dag [elicited]
   birds fly-DECL
   "Birds fly."

In example (1), the subject marker /ə/ denotes a generic marker, which is not obligatory as seen in (2). A list of common nouns where the subject marking is optional is given in Table 1.

Mising	English	Obligatory/Optional
pəttaŋ	bird	Optional
tani	people	Optional
ikiː	dog	Optional
pɔtin	book	Optional
məːnam	love	Optional

Table 1: Subject Marking in Common Nouns with Respect to Generic Reference

1.1.2 **Specific reference**

3. pətta-**də** dəː-dag [elicited]
   bird-SPEC fly-DECL
   "The bird is flying."

4. Tani-**də** jub-duŋ [elicited]
   man-SPEC sleep-PROG
   "The man is sleeping."

In examples (3) and (4), the speaker possesses prior knowledge of the "bird" and the "man," making the use of the subject marker obligatory.

In Pagro Mising, **generic markers** may be optionally omitted without altering the semantic interpretation. However, the addition of **/ə/** to the subject in generic reference serves to enhance emphasis. In contrast, the **specific reference marker /də/** is **obligatory** and must always be present.

5. Object Marking or Accusative Marking

In Pagro Mising, objects are marked by adding *-m, -əm, or -mə* to both common and proper nouns. Objects known to the speaker or hearer obligatorily take an object marker, as illustrated in examples (5)–(8).

5. Mukkaː-bɨ situm-**dəm** səgap-tɔ [elicited]
   Mukkaː-3PS bear-ACC catch-PAST
   "Mukkang caught the bear."

6. Mukkaː-bɨ Appun-**mə** kaː-tɔ [elicited]
   Mukkaː-3P appun-ACC see-PAST
   "Mukkang saw Appun."

7. ɔ-kɔ aːm-**əm** jɨg-duː-n
   inanim-which paddy-ACC separate-PROG-Q
   "Which paddy are you separating?" [PM_AA_03] (Author,unpublished data)

8. aːm-sə-**m** ɔkkai-dɔː-bɔ lɛ-ka.. jɛːt..aːɾ-dɔː-bɔ lɛ-ka
   paddy-PROX:LOC-ACC inanim-TMP-DISP sow-PERF May..July-TMP-DISP sow-PERF
   "the paddy was sown...during May-June-July.." [PM_AA_08] (Author,unpublished data)

6. Dative Case

The Dative Case shows who is getting something, who benefits from an action, or who the action is directed towards. The dative case markers in Pagro Mising, which are marked in the indirect object in ditransitive verbs, are marked using *-mé*. In Mising, case syncretism is evident in the overlap between the accusative *-mé* (10) and dative *-mé* (13) (14) (15), a phenomenon observed in various languages, including Latin and other Indo-European

languages. In Latin, syncretism manifests in multiple paradigms, such as the neutralization of the nominative and accusative in neuter nouns and many plural forms, as well as the merging of the dative and ablative cases in all plurals and certain singular declensions. A strictly form-based approach to case distinction would necessitate a highly complex system with separate rules for each variation, complicating the analysis of case marking. In contrast, the traditional approach allows for a more systematic and generalized description by recognizing overarching case functions despite surface-level syncretism. Similarly, in Mising, the convergence of the accusative and dative case marking indicates that distinct morphological markers for these cases are not required in all contexts. This supports the broader linguistic observation that case syncretism serves as a natural mechanism for structural simplification while maintaining functional distinctions within the grammatical system. Blake (2001:20)

9. mukka-bɨ ajaŋ-**mə** appun-əm bɔm-bi-tak-ku [elicited]
   Mukkang-3PS Ayang-DAT flowers-ACC bring-give-do-PAST
   "Mukkang has brought flowers for Ayang."

10. bau-bɨ ɔu-**mə** rɛdiɔŋ-kɔ rəbi-tag [elicited]
    father-3PS mother-DAT radio-IND buy-DECL
    "My father bought a radio for my mother."

11. ɔi-bɨ kaŋkan-**mə** potin-də-m bi-ka [elicited]
    Oi-3PS Kangkan-DAT book-DEF-ACC give-PAST
    "Oi gave the book to Kangkan."

7. Instrumental Case

The *instrumental case* is a grammatical category that denotes the means, tool, or agency by which an action is carried out. It typically answers *"by what means?"* or *"with what?"*, marking an entity that facilitates or enables an action. This case is widely attested across languages and serves various syntactic and semantic functions, including indicating the instrument used in an action (*e.g., writing with a pen*), expressing accompaniment (*e.g., walking with a friend*), and denoting causation or manner. The instrumental case markers in Mising are *-kokki* and *-kki*.

12. pɔɾɔtaŋ **kɔkki** dɔ-tɔ
    paratha INST eat-IMP
    "Eat with paratha."

13. ɨgiŋ ɔlɔ-**kki** ɨs:i-dəm zeɾ-tɔ
    axe LOC-INST tree-acc tear-imp
    "Tear the tree with that axe."

8. Ablative Case

The *ablative case* is a grammatical category that primarily denotes movement away from or separation from a source. It typically answers questions such as *"from where?"*, and *"from whom?"*.

In Pagro Mising, the ablative case is marked by the morpheme *-kké*. These markers are used across different nominal categories, including common nouns (15), proper nouns (16), and pronouns (17), to indicate source, origin, or cause within a sentence. Different ablative markers can be used, as seen in (17) for pronouns.

14. telam sɔ-**kkə** pasigat tɔ:-pə
telam LOC-ABL pasighat to-LOC
"From Telam to Pasighat."

15. ɨsiŋ dɔ-**kkə** ɔkum sɔ:-pə
tree LOC-ABL house PROX-LOC
"From the tree to the house."

16. ɾɔmɛn kɔlɔ-**kkə** bɔm-tak-ku
Romen LOC-ABL bring-do-PAST
"Brought it from Romen."

17. bɨ-kkə tɔ-**kkə**/tɔlɔ-**kkə**/bɔ-**kkə**/bɔlɔ-**kkə**
3PS-GEN LOC-ABL
"It is from his/her place."

In **Pagro Mising**, the **ablative case** seems to have evolved from a combination of **locative and genitive markers**, much like in **Galo**- /lə+kkə/. Post (2007:738) The **genitive and ablative postpositions** often function in similar ways, especially at the **clause level**. Its ability to convey **extra spatial or deictic meaning sets the ablative apart**, adding a sense of direction or movement away from a source.

9. Genitive Case

The *genitive case* is chiefly employed to mark noun phrases as dependents of other nouns, serving an adnominal function that essentially links modifiers to the nouns they describe. Most notably, it conveys relationships of possession or ownership, imbuing the noun it modifies with a sense of connection or belonging. In Pagro Mising, *ké/ kké* is used to mark the genitive case. *-ké* is used for names of persons (18). *-kké* /kkə/ is used after high vowel sounds (19) like bɨ-kkə "his", bu-lu-kkə "theirs," and bɨ-ŋi-kkə "their-dual."

18. mukkaŋ-**kə** mətɔɾsaikɛl
Mukkang-GEN motorcycle
"Mukkang's motorcycle."

19. ..ŋɔ-lu-**kkə**-m ə-də-m-pə-na i-dag-nə
1PS-PL-GEN-ACC that-DEF-ACC-SUBR-DECL do-DECL-NZR
"...we do it that way." [PM_AA_14] (Author,unpublished data)

10. Locative Case

The locative postposition **/lɔ/** or **/lə/** in Pagro Mising serves as the primary and most frequently occurring marker for indicating spatial locations. It plays a fundamental role in encoding semantic locations, specifying where an event occurs. Additionally, **/lɔ/** functions as an etymological component in distal locative demonstrative postpositions, such as **/tɔlɔ/** or **/tələ/** ('downward distant location') and **/bɔlɔ/** or **/bələ/** ('upward distant location'). In contrast, the proximal locative markers are realized as **/sɔ/** or **/sə/**, indicating locations closer to the speaker.

This postpositional system reflects a rich spatial deixis in Pagro Mising, distinguishing between proximal and distal locations while incorporating vertical distinctions. These markers not only indicate static spatial reference but also contribute to a broader system of deictic spatial relations within the language.

20. nɔ…ɔdɔkkə **h(s)ə** a:m-**sə**-m ədɨ-lɔ lɛ-ka-n
    2P…then PROX.LOC paddy-PROX.LOC-ACC when-TMP sow-PERF-Q
    "…when did you sow the paddy?" [PM_AA_07] (Author,unpublished data)

21. arɨg-**tələ** duŋ [elicited]
    field-DIST.LOC.UP PROG
    "It is in the field(agriculture)."

22. ...zɔ:-mɔ-gɛ:-la kumsu:-**dɔ** bɔrai-dag ɔdɔkkə la:ji la:ji-gɛ:-la
    ...carry-CAUS-COMT-and farm-LOC put-DECL SEQ bring down-REDUP-COMT-and

    jɨg-dag
    separate-DECL

    "...they carry them…then it is put inside the farmhouse…then it is brought down when needed and separated (grains from stem)" [PM_AA_10] (Author, unpublished data)

11. Benefactive Case

The *benefactive case* in Pagro Mising is morphologically marked on the benefactor (23), indicating that an action is performed for the benefit of a specific entity. The morpheme *-képé* functions as the dedicated benefactive marker, attaching to the noun denoting the benefactor. There is another marker /bɔ/ which represents accompaniment, followed by the benefit of the entity. (24) This marking plays a crucial role in distinguishing events where an action is carried out on behalf of, for the sake of, or in service to another individual.

23. dərmi **kəpə** appun bɔm-bi-tak-ku [elicited]
    dermi BEN flower bring-give-do-PAST
    "I have brought flowers for Dermi."

24. ambɨn ti:n-n(l)ɔ bɔrai-jɛ-ku .. bɔrai-gɛ:-la ɔdɔkkə
    uncooked rice rice tin-LOC put-FUT-CMPL.. put-COMT-and SEQ

    apin-pə mɔ-la dɔ-**bɔ**-dag-bɔ-nə-na
    cooked rice-MANN.ADVZ make-ABIL eat-BEN-DECL-REDUP-NMZ-DECL

    "put them inside the rice basket..then cook the rice and feed people" [PM_AA_18] (Author,unpublished data)

12. Allative Case

Allative Case indicates the motion *to a location*. It is marked by following the combination with the locative case marker where the locative is lengthened in the process. The allative marker in Pagro Mising is *-pé*.

25. dilli tɔː-**pə** gɨ-la duŋ [elicited]
    Delhi LOC-ALL go-ABIL PROG
    "Going to Delhi."

13. Vocative Case

The *vocative case* is a well-documented feature in the Tani languages, and it is prominent in Pagro Mising, where it serves the primary function of direct address. Unlike other grammatical cases that signal relationships between nouns and their syntactic heads, the vocative does not establish such dependencies within a sentence. Instead, it operates independently, often appearing outside the main syntactic structure or as an inserted element in discourse.

26. karmuk-**a**.. ɔkɔ-lɔ du-n
    karmuk-VOC where-LOC stay-Q
    "Karmuk..where are you?"

14. Conclusion

This study has explored the case marking system in the Pagro variety of Mising, offering insights into its structure, functions, and typological significance. The analysis reveals that case markers in Pagro Mising play a fundamental role in encoding grammatical relations and semantic roles, distinguishing between subjects, objects, indirect objects, and various oblique functions. Case marking in Pagro Mising is characterized by both overt and covert markers, with their usage being influenced by syntactic, semantic, and discourse-pragmatic factors. The presence of optional case marking, particularly in generic or indefinite contexts, reflects a broader tendency observed in other Tani languages.

A key finding of this study is the syncretism observed across different case markers, particularly in the accusative, dative, instrumental, ablative, and genitive cases. Such overlaps suggest an underlying structural economy, where a limited set of markers serves multiple grammatical functions, a phenomenon commonly found in Trans-Himalayan languages. This syncretism also highlights the role of context in determining case interpretation, as speakers rely on syntactic and semantic cues to disambiguate meanings.

The study also underscores the interaction between case marking and discourse factors such as definiteness, specificity, and topicality. In many instances, case markers are omitted when the referent is indefinite or generic, reinforcing the view that case marking in Pagro Mising is not purely syntactic but is also shaped by pragmatic considerations. Additionally, historical and areal influences appear to have played a role in the development of the case system, as seen in the structural similarities between Pagro Mising and other Tani varieties, as well as neighboring languages from different language families.

By providing a detailed account of case marking in Pagro Mising, this study contributes to a broader understanding of case systems within the Tani language family and the Trans-Himalayan linguistic landscape. Given the limited existing research on Pagro Mising, these findings serve as a foundation for further comparative studies, particularly in relation to case syncretism, alignment patterns, and language contact effects. Future research could examine dialectal variation within Mising, explore morphosyntactic change over time, and investigate how case marking interacts with other grammatical domains, such as verbal agreement and information structure.

Ultimately, this study highlights the richness and complexity of case marking in Pagro Mising, demonstrating its significance not only for the study of Tani languages but also for broader linguistic typology. The findings contribute to ongoing discussions on the nature of case systems in Tibeto-Burman languages and provide a basis for further inquiry into the morphosyntactic diversity of the region.

Abbreviations

1=First Person
2=Second Person
3=Third Person
ABL=Ablative
ABIL=Abilitative
ADVZ=Adverbalizer
ALL=Allative Case
ACC=Accusative Case
BEN=Benefactive Case
CAUS=Causative
COMT=Comitative
DAT=Dative Case
DECL=Declarative
DEF=Definite
DISP=Disposal Perfect
GEN=Genitive
IMP=Imperative
IND=Indefinite
INST=Instrumental Case
LOC=Locative
MANN=Manner
NZR=Nominalizer
P=Person
PAST=Past
PERF=Perfect
PROG=Progressive
PROX=Proximal
REDUP=Reduplication
Q=Question
S=Singular
SUBR=Subordinate
SPEC=Specific
TMP=Temporal
VOC=Vocative Case

References

Blake, B. J. (2001). *Case*. Cambridge University Press. https://doi.org/10.1017/s0022226700000827

Chelliah, S. L. and Barðdal, J. (2009). The role of semantic, pragmatic, and discourse factors in the development of case. *The Role of Semantic, Pragmatic, and Discourse Factors in the Development of Case*, Studies in Language Companion Series https://doi.org/10.1075/slcs.108.01bar

Chelliah, S. L., Hyslop, G., et al. (2011). *Introduction to special issue on optional case marking in Tibeto-Burman. Linguistics of the Tibeto-Burman Area*, 34(2):1–7. https://doi.org/10.32655/ltba.34.2.01

Genetti, C., Coupe, A. R., Bartee, E., Hildebrandt, K., and Lin, Y.-J. (2008). *Syntactic aspects of nominalization in five Tibeto-Burman languages of the himalayan area.* Linguistics of the Tibeto-Burman area, 31(2). https://doi.org/10.32655/ltba.31.2.04

Lehman, F. K. (1973). *Remarks on the grammar of topic and comment in Burmese.* In Sixth International Conference on Sino-Tibetan Languages and Linguistics, San Diego.

Post, M. W. (2007). A Grammar of Galo (Unpublished doctoral dissertation), La Trobe University Bundoora.

Taid, T. (2016). *An Introduction to Mising Phonology and Grammar*. Anundoram Borooah Institute of Language, Art and Culture (ABILAC), Assam.

Walling, P. (2010). *Semantic agent in Tibeto-Burman languages.* Bulletin of the Deccan College Research Institute, 70:305–311.

वाक् मंथन
ISSN: 2426-2149

The preponderance and cultural hegemony of the standard languages and the dominant elites

Hari Madhab Ray
Jawaharlal Nehru University
hmray@jnu.ac.in

Abstract

Language is intertwined with social, cultural and political aspects of power. Language is not only a system of communication but it is also a medium for thought, a vehicle for literary expression. It is considered to be a psychological phenomenon, an identity marker, a social institution, a matter for political controversy, a factor in state or nation building, a unique gift to human beings, a source of human life and power and so on. The notion of 'language' has always intrigued human minds. The perception created by the ruling elites for the standard language created a division and hierarchy in the society. This paper attempts to throw some lights on the preponderance and cultural hegemony of the standard languages and the dominant elites in Indian context.

Keywords: Standard language, dialect, power, ideology, cultural hegemony, scheduled languages

1. Introduction

The Linguistic Survey of India (1894-1928) compiled and edited by George A. Grierson ended with technical descriptions of 179 languages and 544 dialects. In addition to the grammatical sketches of those languages and dialects, the perennial question of the distinction between language and dialect is solidified by the survey. As for example, Hindi, in its totality, refers to a dialect continuum spoken mainly across northern India. This continuum is usually divided into two forms: Eastern and Western Hindi. Eastern Hindi is mainly made up of Awadhi, Chhattisgarhi and Bagheli dialects, while Western Hindi consists of Haryanvi, Brajbhasha, Bundeli, Khariboli and Kanauji dialects etc. Many dialects once again pushed into the periphery and some dialects became standard language as for example Maithili and Dogri (2004). Lewis, Simons & Fennig (2013) mention that Kanauji is a language with very low identity based on their research. In independent India, the eight scheduled languages or the notion of standard languages has led to the emergence of a regional ruling class whose cultural hegemony over the larger masses is not very different from that of the English-speaking elite or that of the dominant elite with their cultural and religious ideologies. According to D.P. Pattanayak (1999:149), "Inequality has many faces. Giving recognition to a single language variety as standard creates a cadre of people who through various controls gain from the acquisition, processing, storage and transmission, retrieval and other manipulations of the language as the language of education, administration and mass communication in a plurilingual society bestows advantages on the speakers of that language." In the following, the notion of language, standard language, dialect and the cultural hegemony of the standard languages and the dominant elites will be explored in the Indian context.

2. The notion of 'language' and 'standard language'

Humans are different from the rest of the beings, from the birds, animals and also from the closest evolutionary ancestors like the intelligent apes, the bonobos, chimpanzees and the gorillas. The communication systems of the humans are entirely different from the rest of the beings. The humans claim that they fairly know about the communication system of the rest of the beings but in reality, they know very little about their communication system. The so-called intelligent homo sapiens i.e, humans have 'language' and the discovery of FOXP2, is the first gene known to be involved in the development of speech and it made us believe that humans are up in the ladder in the evolutionary stages (1). The possession of 'Language' or the 'language faculty' is undoubtedly the biggest gift to humans. According to Lewis Thomas (1978), "The gift of the language is the single human trait that marks us all genetically, setting us apart from the rest of life". In Chomsky's words (2006) "when we study human language, we are approaching what some might call ' human essence', the distinctive qualities of mind that are so far as we know unique to man".

Language is many things—a system of communication, a medium for thought, a vehicle for literary expression, a psychological phenomenon, an identity marker, a social institution, a matter for political controversy, a factor in state or nation building (2), a unique gift to human beings, a source of human life and power and so on. We express our ideas, thoughts, emotions, desires, passions by language. The use of language is central and inevitable in speech and also in the classical texts, epics, sermons, writings of the great minds, fictions, novels, stories, dramas, comedies and tragedies—all the genre of literature and in every aspect of our human existence. We cannot think of our lives without language, even our thinking takes place in language, though some levels of perception and cognition can take place without language. Though language is ubiquitous in nature, it is very difficult to provide a ready-made definition of language. According to Kracht (2007), "Languages are sets of signs. Signs combine an exponent (a sequence of letters or sounds) with a meaning. Grammars are ways to generate signs from more basic signs. Signs combine a form and a meaning, and they are identical with neither their exponent nor with their meaning." Edward Sapir (1921:8) opines that "Language is a purely human and non-instinctive method of communicating ideas, emotions and desires by means of voluntarily produced symbols." But in our daily lives we communicate much more than ideas, emotions and desires. Ideas are inherently imprecise. Many voluntarily produced symbols are not counted as language. In Sapir's definition body language i.e, gestures, postures, eye gaze and so on would be counted as language. According to Bloch and Trager (1942:5), "A language is a system of arbitrary vocal symbols by means of which a social group co-operates". This definition of language makes no appeal directly to the communicative function of language. It has all the emphasis on social function. It takes a narrow view of the role that language plays in society. It brings a property of arbitrariness and explicitly restricts language to 'spoken language' but written language or sign language used by the deaf community is also considered as language. Hall (1968:158) opines that Language is "the institution whereby humans communicate and interact with each other by means of habitually used oral-auditory arbitrary symbols". Both communication and interaction are introduced in this definition. Interaction is broader and better than 'cooperation' which was used in Bloch and Trager definition (1942). Oral auditory symbols equivalent to vocal signals were introduced but it refers to hearer as well as to the speaker (receiver/sender).

Hall (1968) and Sapir (1921) treat language as purely human institution i.e, Language that is used by a particular society is part of society's culture. The property of arbitrariness is also mentioned in their definitions. Robins (1979:9-14) does not give a formal definition of language. For him it is trivial, uninformative and it presupposes many things rather he is interested in talking about salient facts about language. Languages are "symbol systems… almost wholly based on pure or arbitrary conventions…". He emphasizes adaptability and flexibility. Following Behaviorist understanding language is considered as a set of behaviors but as we know habit is not adaptable behavior but language is, as for example new words enter in the language, language is always dynamic, English which was spoken hundred years ago have changed in many ways if we compare it with modern English of the 21st century.

It is interesting to note that in America during the 1950s, linguists and psychologists were strongly influenced by Stimulus and Response theory of the Behaviorist. Language is considered as a set of behaviors in response to stimuli. Children learn language by imitation, reinforcement and analogy (B. F. Skinner 1957). Many behaviorists have adopted this theory and applied it in language use and second language teaching. The stimulus and response theory has very restricted applications in linguistics and in psychology, and many modern linguists completely abandon this theory after the advent of Generative transformational grammar propagated by Noam Chomsky (1957, 1965, 1981, 1995). According to Chomsky (1957:13); "From now on I will consider a language to be a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements." In this definition, nothing is said about communicative functions of either natural or non-natural languages, nothing is said about the symbolic nature of the elements or sequences of them. It gives emphasis to the structural properties of languages and the formal properties can be investigated from a mathematical point of view. Chomsky gives emphasis to structure dependence rules and the Universal Grammar. For him, humans have an innate language faculty. He argues that children can acquire language from an unsystematic and small amount of data in a relatively small amount of time without any formal training. Language acquisition involves very little imitation and with little stimulus children make innovative sentences that they might not have heard before. They also understand hundreds of sentences they have never heard before. He proposes the poverty of the stimulus in which UG facilitates language acquisition for the children. If we were not biologically predisposed for acquiring language then this speed of acquisition from a small amount of data would not be possible. Children from all over the world irrespective of their social and geographical background develop an adult-like command over their language in 4 to 6 years. A cross linguistic investigation of linguistic universals shows that features found in one language are also found in totally unrelated languages. Chomsky (1957, 1965, 1981, 1995) argues that we are born with a set of rules of grammar that he calls the 'Universal Grammar'. These are hardwired in our brain. All human languages are built on this Universal Grammar; it is a set of principles and parameters that constrain all human languages.

India is a land of diverse communities, cultures, ethnic groups, and foreign settlers. India has five recognised language families; Indo-Aryan, Dravidian, Tibeto-Burman, Austro-Asiatic and Andamanese. In addition, Tai Kadai languages, other languages of Andaman and Nicobar Islands, language isolates namely Burushaski and Nihali, Indian Sign languages (ISL) deserve special mention. The Constitution of India recognises 22 languages of India as the Eighth Schedule languages. These are Assamese, Bangla, Bodo, Dogri, Gujarati, Hindi, Kashmiri, Kannada, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Tamil,

Telugu, Sanskrit, Santali, Sindhi, and Urdu. The exact number of unwritten languages is hard to determine in the Indian context. Ethnologue (25th edition) has data to indicate that of the currently listed 7,151 living languages, 4,169 have a developed writing system. An additional 154 languages are signed.[3] The fundamental differences between 'Language' and 'language' are made by the linguist. The Language in uppercase L(anguage) is the human language capacity and the language in lower case l(anguage) is the instantiation of the human language capacity as for example Rajbanshi, Bengali, Hindi, Spanish, French, English etc. Irrespective of the differences across languages in the surface level and the language are alike in the underlying levels. They are governed by the same set of UG principles. The Generativist looks at the formal properties of language and ignores the functional aspects of language use. Humans use 'language' as an effective communicative tool which helps them to express their ideas, thoughts, create literature, debate, discuss national, international philosophical matters and so on. The Language used by humans is not the same across a region, country or in the world. It cannot be denied that languages are alike in the underlying level but language differences within a same speech community is a norm rather than an exception. The form of language varies across regions, social class and according to the age, education status, gender, religion of the person. All of us speak differently even if we speak a common language. It is a misconception that all members of the same 'language community, speak exactly the same language' (Lyons 2011:24). There are systemic differences in pronunciation, use of grammar, selection of words and discourse markers in the speech of its members. They do not consider these differences as barriers of communication rather some of these factors solidify and strengthen the community members. It helps them to identify the members, its social status and class, and place them into the hegemonic scale of 'we' vs 'they' relationship. The perception created by the ruling elites for the standard language created a division and hierarchy in the society. In the following, the notion of 'standard language' and 'dialect' will be discussed. The main objective is to find out the layman's understanding of the terminologies and how these terms are understood and interpreted by the linguists.

3. Language and Dialect: An unending saga

When a common man talks about 'standard language' i.e. *bhasha,* he refers to the inherently superior variety of a language. It is understood that the superordinate variety has a script, a well-written literary tradition. There is availability of well written grammar and dictionary in the language. It is used in a large geographical area. The language is used as a medium of instruction in schools, colleges or universities. It is also used in print media and also the language of administration. It is perceived as prestigious by the community members and in the hegemonic power relation with other varieties of the region it has a dominant role. The state showers support for the development and functioning of the language with its cultural and religious baggage. The people of the region want to learn the language for social mobility as speaking in the language is considered prestigious. On the other hand, the opposite is true for a dialect i.e, *upabhasha*. It is perceived as an inferior variety of a language. It is misunderstood that the degrader variety has no script and does not have a well-written literary tradition. There is non-availability of written grammar and dictionary in the variety. It is used in a smaller geographical area. The variety is not used as a medium of instruction in schools, colleges or in universities. It is also not used in print media and also not used as the language of administration. It is perceived as non-prestigious, impure and lower by the community members and in the hegemonic power relation with other varieties of the region and it has a

subordinate role to play in the society. The state does not fully support the development and functioning of the variety; rather it tries to stop its growth by imposing the standard variety in the name homogeneity, cultural uniformity or national integration. The people who use the variety, majority of them do not feel like speaking it in public places. The majority of the people of the variety also do not want it to pass on to their next generation because of several factors as for example negative attitude of the people towards their language and also for social, psychological, economic and practical reasons.

There are many commonly held views about 'language' and 'dialect'. In this view, 'Dialect is regarded as a degraded version of a language' and 'the use of a dialect should be avoided'. The views based on perception of a community is also very common as for example 'Bengali is a sweet language', 'French is a romantic language', 'Sanskrit is mother of all languages', 'Hindi is the national language of India' etc. People have different misconception about linguists and linguistics as well. Many people think that linguists know many languages, they can read and write in multiple languages and they are grammarians. The misconception about Linguistics is that it is often thought to be a sub branch of English (language) study. However, all these are treated as myths or misconceptions in Linguistics study. It is argued by linguists that a standard language is also a dialect. It is a question of power and recognition, status and institutional support for that variety which can make it a standard language. According to Max Weinreich (1940), 'a language is a dialect with an army and navy'.

The colonial philologists addressed the problems of the origins, classification and interrelation between various vernacular languages. John Beams addresses the problem of classification of Indian languages in his 'Outlines of Indian Philology (1867). It also sets the norms of philological study of the North Indian languages. He produced a subsequent three volume work titled "On Dialects", 'Comparative Grammar of Modern Indian Languages'. Beams attempted to treat the commonly held standards that were used to determine whether a language was a dialect of another language or an independent language in its own rights. It Challenged the Commonly held view of "mutual intelligibility" i.e, if the speakers of two different languages could understand each other then, the two tongues were dialects of the same language. Beams argued that such a test was unsuitable for Indian languages because many languages either shared similar words (like Hindi and Bengali) or the same grammatical structure (like Hindi and Punjabi). Beams warned that these pitfalls could result in misclassification of Indian languages. These pitfalls, he argues, could be counteracted by supplementing the rule of 'mutual intelligibility' with another set of parameters. These influences are, geographical position, civilization, political and physical accidents, religion, differences of pronunciation and education. Beams called for a juxtaposition of observations of contemporary speech patterns, vocabulary and socio-cultural context along with the study of the historical context for the development of languages in India. John Beams (1867) states in the 'Outlines of Indian Philology' that colonial philologists agreed that even though eventually the aboriginal peoples were enslaved and colonized by the invading Aryans, they left their mark on the resultant Aryan dominated culture of India. In both of his books, the 'Outlines of Indian Philology' as well as 'Comparative Grammar of Modern Indian Languages', Beams painstakingly proved that Odia drew heavily from tribal languages.

The colonial philologists and structural linguists applied the criterion of 'mutual intelligibility' to distinguish a 'language' from a 'dialect'. This is not a cure-all criterion to solve the issue of language and dialect. There are many exceptions to the applied criteria of mutual intelligibility

where both the varieties are intelligible but considered as two different languages as for example Hindi and Urdu. Even when two varieties are not mutually intelligible, the varieties in question are declared as dialects of one of the standard varieties. Many regional varieties of Chinese are not mutually intelligible but these are not considered as languages but rather dialects of Mandarin.

The district collector of Cuttack (1854) opines, "The Ooriah of this district, whether it may originally have been, is not but a dialect of Bengalee, from which it differs chiefly in pronunciations and in its written character...I would submit as a measure of general policy, it is desirable that the Ooriah should cease to exist as a separate language within the British territories (Collector of Cuttack, the capital of the Orissa Division of the Bengal Presidency,1854). In the mid of the 19th century the existence of Odia as a separate language was often questioned by the erstwhile dominant elite of Bengal. They consider Odia as a dialect of Bengali not as an independent language. According to Kantichandra Bhattacharya, a Bengali school teacher from the Odia-speaking district of Balasore, Uriya Swatantra Bhasa Naye (Odia is not an independent language). He wrote extensively to support his claim that why is Odia not an independent language?

Linguists are aware of the fact that whether X (a variety) is a language or a dialect and it is not considered to be a pure linguistic question. In order to answer the question, they discuss the criteria of 'mutual intelligibility', differences in various levels of grammar between the said varieties, and also differences in the History, Culture, Tradition, written script, literary tradition of the given linguistic community. It is clear that none of these applied criteria qualified to call the said two varieties as two different languages. The recognition of language is determined by socio-political forces. Politics and polity often decide the fate of languages. When a particular linguistic community is volatile and voices their opinion for their language and demands the recognition of their language, the said variety may be recognised as a separate language by the state or the Centre. The dominant elite in power looking at the vote-banks and local politics either give recognition to that language or it may sabotage the prospect of the language by dividing the linguistic communities. As for example, the people of North Bengal still are not sure what should be the name of their language? Is it Kamtapuri? Is it Rajbanshi? Is it Koch-Rajbanshi? Is it Goalparia? Is it Rangpuri? Is it Koochbehari, Northern Bengali, Surjapuri, Desi, Mui Tui Bangla, Kamta, Kamtapuri, Kamtabihari or Koch Rajbanshi? The politics of divide and rule will continue as long as the people want it and the state will always seek the opportunity to subjugate the toiling masses in various means of exploitation. After assuming the chief minister's office, Mamata Banerjee formed an academy to preserve and propagate the Rajbanshi language and Rajbanshi Bhasha Academy was created. Later, she also announced the formation of Kamtapuri Bhasha Academy in 2017. It is the same language but there are many names for the language, at least two different academies, it is the politics of divide and rule, it is the politics of cultural subjugation in the name of standard language for Bengali cultural homogeneity According Weinreich (1945) 'A language is a dialect with an army and a navy'. For linguists, no dialect or language is better, more correct or more logical than any other. Even the standard language is also a dialect or a variety and in majority of the cases it is the variety of the dominant elite.

'Dialect' refers solely to the way in which the language is pronounced and carries no implication whatsoever with respect to grammar and vocabulary. In literacy, spoken language or dialect has no role to play, or still worse, it is only a hindrance. It encourages fissiparous, anti-national tendencies and only signifies backwardness of the people. During the left regime in West Bengal, the sons of the soil in North Bengal who voiced their opinion in their local language namely Rajbanshi or Kamtapuri, were targeted by the erstwhile dominant ruling class of Bengal. Many of them were sent to jail because they spoke of their mother tongue, mother land, exploitation by the ruling class and overall development for the son of the soil. The people were tortured for speaking in their mother tongues, for voicing their opinion in their mother tongues, they were forbidden to create written literature in their mother tongues. The ruling class felt that by giving prominence to Rajbanshi or Kamtapuri may encourage disintegrative forces and, hence, their work should be checked and it should not be brought into the mainstream language and literature.

The teaching and learning of local language or dialect should be avoided because of the existence of different types of frivolous notions in the society. The people think that education in standard languages only or in Indian context Eight schedule languages will liberate the people socially, economically. The education in the Eight schedule languages brings national integration. Though the Hindi orthodox camp tried to push Hindi as the national language, they were not successful in declaring Hindi as a national language. While setting the standard Hindi language as a national language, a language for unity and cultural assimilation, many spoken languages (Awadhi and Brajbhasha, Angika, Magahi etc.) of Northern India lost its glory. If a dialect spoken in a geographical area and that area for historical and political reasons acquire the status of a state or a nation, the dialect in question may be recognised as a standard language.

3. The Preponderance of Standard language and cultural hegemony

The concept of 'standard' language has somehow reduced the spoken languages or dialects to a symbol of backwardness. The whole educational system has a very hostile and demeaning attitude towards the spoken varieties. The teachers prefer to talk in the standard language and majority of them are hostile towards the student's regional variety. The damage done to the self-respect, self-confidence of the child at the beginning stage is irreparable. Even the word '*desi'*, *'dehati'* has acquired a derogatory connotation. The people of the ruling class, the dominant section of the society who are proficient in the standard language use humiliating expressions towards the speakers of the dialects—'they cannot speak', 'they do not have a language', 'they do not have a script' etc. Lack of knowledge of a particular standard language is many times reduced to not having a language. It has been mentioned by many informants that many surveys conducted by the government officers in the rural areas, the amount of information collected by the officers not visiting every household premises of the community members. The officers prefer to collect the information from the local Panchayats, Pradhan, sarpanch, or influential person in the village. According to the District Census Handbook Coochbehar; "Mother tongue and language are the two most important characteristics of the population. As per Census definition, mother tongue is the language spoken in childhood by the person's mother to the person. The Table below shows the population by mother tongue for Koch Bihar district since 1961. It is observed from the Table that as per Census 2001, Bengali is the mother tongue for 90.3 per cent of the population of the district followed by Hindi by 1.6 per cent. Only 0.6 per cent of the population of the district

have returned to their mother tongue as Telugu." The figures of the Census for the number of speakers of Rajbanshi is alarming in the District Census Report, though the majority of the population in the districts of North Bengal speak Rajbanshi. Here is the table, Population by mother tongue in Coochbehar District (1961-2001).

Mother Tongue :

Table 11 : POPULATION BY MOTHER TONGUE IN KOCH BIHAR DISTRICT (1961-2001)

Year	Bengali	Hindi	Telugu	Sadan/ Sadri	Bhojpuri	Kurukh/ Oraon	Nepali/ Gorkhali	Santali
	1	2	3	4	5	6	7	8
1961	980330	31884	164	0	1	857	805	321
	(96.1)	(3.1)	(0.0)	(0.0)	(0.0)	(0.1)	(0.1)	(0.0)
1971	1375948	31427	33	---	---	936	1218	1493
	(97.3)	(2.2)	(0.0)			(0.1)	(0.1)	(0.1)
1981	1720444	41281	89	---	---	1182	1227	1053
	(97.1)	(2.3)	(0.0)			(0.1)	(0.1)	(0.1)
1991	2117062	37018	---	2007	401	1869	1492	1054
	(97.5)	(1.7)		(0.1)	(0.0)	(0.1)	(0.1)	(0.0)
2001	2239155	38604	14122	4154	1923	1777	1680	1305
	(90.3)	(1.6)	(0.6)	(0.2)	(0.1)	(0.1)	(0.1)	(0.1)

Table 11 : contd....

Year	Marwari	Rajbangsi	Munda	Rabha	Rajasthani	Assamese	Koch	Other Mother Tongues	Total Population
	9	10	11	12	13	14	15	16	17
1961	254	80	---	1595	357	813	---	2345	1019806
	(0.0)	(0.0)		(0.2)	(0.0)	(0.1)		(0.2)	
1971	---	---	208	1291	---	365	---	1264	1414183
			(0.0)	(0.1)		(0.0)		(0.1)	
1981	---	---	422	1017	---	277	---	4651	1771643
			(0.0)	(0.1)		(0.0)		(0.3)	
1991	459	1685	644	1043	---	426	---	5985	2171145
	(0.0)	(0.1)	(0.0)	(0.0)		(0.0)		(0.3)	
2001	1138	1122	770	712	589	586	566	170952	2479155
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(6.9)	

Table.1 Population by mother tongue in Coochbehar District (1961-2001)

It is strange that the mother tongue speakers in Rajbanshi are only a few in numbers when the Rajbanshis are the majority in Coochbehar district. They are more than 70% of the total population of the districts and it is not only the Rajbanshis but also the local Muslims who speak the Rajbanshi language. In spite of that, it is reported in the Kochbihar District Census Handbook 1685 (1991) speakers and 1122 (2001) speakers. The Rajbanshi speakers are also decreasing in numbers from 1991 to 2001 as shown above. The Rajbanshi community comprised about 76 percent of the total rural scheduled caste population in Koch Bihar in 2011. Koch Bihar district comprised a large percentage of scheduled caste population, which was about 50.17 percent of its total population in 2011 and ranked first when compared with all other districts of West Bengal. The concentration of scheduled caste population increased between 1971 (47.03 percent) and 1991(51.76 percent). It slightly diminished between 1991 and 2001 (50.17 percent) and remained constant in 2011. The Rajbanshi community comprised about 75 percent, 78 percent and 77 percent of the total scheduled caste population of the district in 2011, 2001 and 1991 respectively.

In India, minority languages are hardly taught in schools and colleges. These languages are not the medium of instruction in primary schools. The Constitution of India safeguards the minority languages, enlist 22 scheduled languages and linguistic rights are enshrined in article 343-351. The article 350 and 351 categorically mention about the linguistic minority community and medium of instruction in mother tongue in primary schools.

English language is not to be found in the list of 22 languages, but English enjoys its elite position as an official language. The 8th Schedule languages are always seen as a boost for political, socio-economic upliftment, as an epitome of power and prestige. Languages having less than 10000 speakers are not listed in the Census after 1971. The Constitution of India does not provide any definite criteria for inclusion of languages in the 8th Schedule, but it is often stated that the languages should have (a) literary tradition (b) sizeable population (c) contiguous geographical area. According to Census (2001), there are almost no Sanskrit speakers in the country's north-east, the eastern States beyond Madhya Pradesh, Jammu and Kashmir, Tamil Nadu, Kerala and Gujarat. Tamil Nadu is unique because according to the 1921 Census, Madras province had the highest number of Sanskrit speakers in the country, 315 of a total of 356. Sanskrit is also the only scheduled language that shows wide fluctuations, rising from 6,106 speakers in 1981 to 49,736 in 1991 and then falling dramatically to 14,135 speakers in 2001. Sanskrit is one of the scheduled languages not because of the number of speakers it has but purely for religious and political reasons, it does not have a sizable number of mother tongue speakers nor a contiguous geographical area where it is spoken. But Pali is not listed in the Eight Schedule though it has enormous literature and it spread out of India to countries such as Thailand, Burma, Indonesia, Srilanka and South East Asian countries with the spread of Buddhism. Sindhi also does not have a corresponding geographical base. Many languages of India fulfill the requirements to be listed in the Eight Schedule Languages but there are other political considerations for not allowing all these languages to enter the Eight Schedule (See the list of 38 languages below).

4. Language, Power and Hegemony: the cultural hegemony of the ruling elites

Language is intertwined with social, cultural and political aspects of power. But what is Power? There are two traditions found in the literature on power ; (a) Power through dominance (b) power by consent. In the first tradition (a) power through dominance, power comes from the privileged access to social resources such as education, knowledge and wealth. Access to these resources provides authority, status and influence. It is an enabling mechanism for the domination, coercion and control of subordinate groups. In Weber's study (1914), we find the corrective power of the state and its institutions, judicial and penal institutions. Power also resides in other sovereign organizations, such as businesses and the church. In democratic society, power needs to be seen as legitimate by the people in order to be accepted and this process of legitimation is generally expressed by means of language and other communicative systems.

In the second stream tradition (b) power by consent in which power is jointly produced (See Scott 2001). Power has been mainly concerned with the significance of its persuasive influence. Gramsci's (1971) concept of *hegemony* describes the mechanisms through which dominant groups in society succeed in persuading subordinate groups to accept the former's own moral, political and cultural values and institutions. Within this framework, discourse constructs hegemonic attitudes, opinions and beliefs. It makes these beliefs appear 'natural' and 'common sense'. An important factor in this process is 'consent'. Subordinate groups are said to consent to the existing social order because it is effectively presented by the state and its institutions as being universally beneficial and commonsensical. The reason why the concept of hegemony as power is especially important is that it operates largely through language. People consent to particular formations of power because the dominant cultural groups generating the language, tend to represent them as natural or common sense. Gramsci (1971) also points out that

dominant groups have to work at staying dominant. They attempt to secure domination, firstly, by constructing a ruling group through building and maintaining political alliances and secondly, by generating consent or *legitimacy* among the population; and, thirdly, by building a capacity for coercion through institutions such as the police, the courts and the legal system, prisons and the military in order to create authority. Foucault (1977, 1980) sees the concept of power as *productive*, as a complex and continuously evolving web of social and discursive relations. Foucault (1980) argues that power is more a form of action or relation between people that is negotiated and contested in interaction and is never fixed or stable. Foucault (1980) does not regard power as an already given entity. Power is maintained through the ideological operations of society.

Intertwined with our understanding of power, *ideology* refers to the ways in which a person's beliefs, opinions and value-systems intersect with the broader social and political structures of the society in which they live. Language is influenced by ideology. All texts, whether spoken or written, and even visual language, are inexorably shaped and determined by a web of political beliefs and socio-cultural practices as for example 'liberal' view of language where texts are seen simply as natural outcomes of the free communicative interplay between individuals in society, uninhibited by political or ideological influence. In our preliminary reading we can find that the texts are anything but neutral or disinterested but in close examination, close linguistic analysis or critical discourse analysis of the text can help us understand how ideology is embedded in language, how the reflexes of 'dominant' or 'mainstream' ideologies are sustained through textual practices. Ideology, for Karl Marx (1933) is the belief systems which are held either individually or collectively by social groups. Marx's original conception (1933), ideology captures means by which dominant forces in society, such as royalty, the aristocracy or the bourgeoisie can exercise power over subordinated or subjugated groups, such as the industrial and rural proletariat. Marx's best-known axioms: 'the ideas of the ruling class are in every epoch the ruling ideas' (1965: 61).

Althusser (1971) was one of the first to describe power as a discursive phenomenon, arguing that ideas are inserted into the hierarchical arrangement of socially and politically determined practices and rituals. Althusser (1971) highlights the significant roles of ideologies in reproducing or changing political relations through so-called *ideological state apparatuses*, such as the church, the legal system, the family, the media and the educational system. The ideological state apparatus (ISA) is at work, and it constructs the citizens as 'Consumers'. The ISA constructs readers as consumers who should take personal responsibility for their health through proper lifestyle choices.

Though linguistic rights are enshrined in the Constitution of India, implementation of the constitutional promises is a far-fetched reality. The dominant state language or the scheduled languages hinder the prospect of the state minority languages because the education in school or colleges are only carried out in those dominant languages of the elite. The children of the other community who do not know the state dominant languages have to face various challenges. For example, in North Bengal (West Bengal), the Santhali, Toto, Rabha or Rajbanshi children will be forced to get their education in Bengali, because Bangla is the state language and it is the medium of instruction in school. Many students do not understand Bengali and fail to understand what is being taught in the class. It gradually leads to unwillingness to come to the class. The children also develop a negative attitude and apathy towards education and they finally drop out of the class. While making a field visit in North

Bengal, a teacher in a primary school shared an anecdote. “It was the first day of his school, he went to the class and started discussing the cultivation of sugar cane in Bangla. The students were in class IV standard. The word ‘sugar cane’ in Bangla is called [akʰ] and the cultivation of sugar cane is translated as [akʰer caʃ]. Since it is Rajbanshi populated area and more than 95% students speak in Rajbanshi, the students had hard time to grasp the word [akʰ] ‘sugar cane’. He could look at the face of the students and having realised that they are not able to understand what was he speaking, he said that [akʰ] *sugar cane* is called [kuʃair] or [kuʃiar] in Rajbanshi. The students immediately understood what he was talking about and he switched over to Rajbanshi instead of Bangla and all the students learned the lesson with great enthusiasm and interest.” In addition to Bengali, there has been attempt to introduce the minority and other languages like Santhali, Toto, Rajbanshi, Kurux in the primary school as a medium of instruction in North Bengal but in reality, it has less impact on the language community because of the step motherly treatment towards these languages for its implementation by the ruling elites of the state.

Similar is the story of the tribal areas and also other areas of India where the state dominant languages have major impact on the lesser- known languages and their communities. Most of the teachers speak in the standard language and they do not speak the regional or other languages in the class, translanguaging is never practiced in the classroom. Rather, speaking of a certain variety has never been appreciated and in some cases, it is being ridiculed by the teachers or the elite peer groups. It has been noted from the above anecdote that when the teacher repeats the instruction in the spoken Rajbanshi variety, the children were able to follow him very well. There used to be perfect communication between him and his students. But most of the teachers do not speak in Rajbanshi either they are not competent in the language or they have the assumption that the children will remain backward if they teach them in the local language i.e, Rajbanshi. They are also of the opinion that the textbooks have to be in the standard Bengali. A textbook written in one of the spoken dialects would be of no use in other parts of Bengal.

The parents and guardians also held the view that teaching in their local languages may not have prospects for the children and they may not have bright careers in the future. The parents prefer to send their children either in a government Bengali Medium School or a private English Medium school available in the vicinity. Since Rajbanshi is not introduced as a subject in the school or colleges or in the university curriculum in North Bengal, the majority of the parents shy away from teaching their children in Rajbanshi. Though, recently there has been an attempt to introduce the Rajbanshi language in the schools and there are some colleges and universities that offer Rajbanshi as a certificate course. The state government and local politicians for the development of the language and community distribute SOPs efore the election to woo the voters and once the election is held, a new set of promises will be made for them. The community members of the minority languages are entangled in the labyrinth of promises, and bigger promises, the more they try to come up with their own set of strategy, the more they find themselves in the marshy land or *daldal* planned by the ruling elites.

Even for standardization of a language or recognition of a language in the Eight schedule, the cultural hegemony of the ruling elites is clearly visible. Standardization is a kind of prescriptivism. The arguments put forward in favour of prescriptivism (Mesthrie et.el, 2009) are; (a) One form is more logical than another like the principles of mathematics (b) Appeal to classical forms or a preference for older forms of the language or the language used by the

renowned authors, scholars, influential people to be followed and (c) Injunction against the use of foreign words or another dominant language in the vicinity. James Milroy and Lesley Milroy (1985) locate the origins of prescriptivism in what they call 'the Complaint Tradition', that is, a long-standing tradition of complaints about the adequacy of the English language compared to others. Einar Haugen (1966) discussed a few processes of standardization in a given language, which may not be chronological in order. Here are the processes; (a) **Selection:** a variety among many must be chosen as the standardized variety (b) **Codification:** the variety will be laid down in the form of grammar books, dictionaries, spellers, manuals, styles, texts etc. (c) **Elaboration**: Putting the variety into whole range of language functions which may be called upon to discharge, including intellectual and abstract functions (d) **Acceptance:** the variety is being pushed through language policies, promotion, spread, establishment and enforcement etc. Later, the said variety which is standardized becomes a part of the hegemonic process of power and domination. It is the hegemony of the standard language of the state or a nation which becomes the cultural hegemony of the ruling elite. In the process of language standardization, the chosen variety needs support from the community and also strong support from the state/central government for its proper language policy formation and implementation in different domains of language use.

India is a multilingual country. According to the Census (1961), there are 1652 languages spoken in the country. In the Eight Schedule of the Constitution of India only 22 languages are listed. Here is the list of languages and its family wise distribution;

1. Assamese. 2. Bengali. 3. Bodo. 4. Dogri. 5. Gujarati.6. Hindi. 7. Kannada. 8. Kashmiri. 9. Konkani. 10. Maithili. 11. Malayalam. 12. Manipuri. 13. Marathi. 14. Nepali. 15. Odia 16. Punjabi. 17. Sanskrit. 18. Santali. 19. Sindhi. 20. Tamil. 21. Telugu. 22. Urdu

Language Family	**Languages**	**No of Languages**
Indo-Aryan	Assamese, Bengali, Dogri, Gujarati, Kashmiri, Konkani, Maithili, Marathi, Nepali, Odia, Punjabi, Sindhi, Urdu, Sanskrit, Hindi	15
Dravidian	Kannada, Malayalam, Tamil, Telugu	04
Tibeto-Burman	Bodo, Manipuri	02
Austro-Asiatic	Santali	01

Table 2. Family wise distribution of Eight Schedule languages

Out of the twenty two languages, fifteen languages are from the Indo-Aryan language family, four from the Dravidian language family, two from the Tibeto Burman family, one from the Austro-asiatic family as shown above. The Eight Schedule languages show Indo-Aryan supremacy. There is a special place for Hindi and its promotion as an official language in India. The Hindi orthodoxy could not become successful to announce Hindi as a national language. The majority of the languages of the Tibeto Burman, Austro-Asiatic, Dravidian and also Andamanese language families do not find its place in the prestigious Eight Schedule languages. The languages which are ignored will figure out in the list of endangered languages. There are many constitutional amendments for the Eight Schedule Languages. In 1967, 21

amendments Sindhi were included in the Eight Schedule and in 1992, through 71 amendments Konkani, Manipuri and Nepali were included in the 8th Schedule Languages. In 2003, Bodo, Santhali, Maithili and Dogri were included in the 8th Schedule Languages through 92 amendments act. In 2011, 96 amendments Act the spelling 'Oriya' was replaced by Odia. From different nooks and corners of the country demands for inclusion in 8th Schedule are increasing. Here is a list of 38 languages given below;

1.Angika, (2) Banjara, (3) Bazika, (4) Bhojpuri, (5) Bhoti, (6) Bhotia, (7) Bundelkhandi (8) Chhattisgarhi, (9) Dhatki, (10) English, (11) Garhwali (Pahari), (12) Gondi, (13) Gujjar/Gujjari (14) Ho, (15) Kachachhi, (16) Kamtapuri, (17) Karbi, (18) Khasi, (19) Kodava (Coorg), (20) Kok Barak, (21) Kumaoni (Pahari), (22) Kurak, (23) Kurmali, (24) Lepcha, (25) Limbu, (26) Mizo (Lushai), (27) Magahi, (28) Mundari, (29) Nagpuri, (30) Nicobarese, (31) Pahari (Himachali), (32) Pali, (33) Rajasthani, (34) Sambalpuri/Kosali, (35) Shaurseni (Prakrit), (36) Siraiki, (37) Tenyidi and (38) Tulu.

One will have to wait and see when these languages will be included in the Eight schedule or is it a mere promise to the people of these languages? In order to sustain cultural hegemony, the ruling elites may not easily allow other languages to be included in the prestigious list of Eight Schedule languages. Whatever is the reason for the inclusion of the languages in the Eight schedule, it is certain that local politics and politics of the regions will decide the future of these languages.

The standard language of the elite has its different types of 'standards' in various domains of language use whether it is used in literature, in law and court, land and revenue documents, insurance policy documents, or in the charge sheets filed by the police or in law enforcements, print media, electronic media etc. The local variety or the spoken language which is easily understood by the masses cannot be replaced in the above-mentioned domains. It is beyond the scope of the ordinary person and they would find it difficult to comprehend the language of the elite. The dominant elites know all the nuances of the hegemonic relations. They do not want to sabotage the bright prospect of their future generations and continue their dominance in language and cultural domains. The dominant elites plan a set of criteria for the standard language which they are well aware of, well versed and majority of the cases, it is their own spoken variety which they proclaim as the standard variety. The dominant elites of India used Hindi with Devanagari Script as a standard language and heavily borrowed words and vocabulary from Sanskritized sources and on the other hand, the dominant elites of Pakistan used Urdu with Perso-Arabic Script with heavy borrowings from Perso-Arabic sources. In both the scenarios, the standard official language of India and Pakistan used their lingua-cultural and religious hegemony in the regional languages of both the countries. The newly created standard language of the countries enjoyed their cultural and religious support. These standardized languages remain in control of the dominant elite. The toiling masses and their language will gradually deviate from the standard prescriptive language norms and cultural practices. If these practices are challenged by another language group or a dialect group, the dominant elite, in power and authority, the custodian of language and culture will clearly use their command in the standard language to repress the challenge raised by the group. The toiling masses will be subjugated and the ruling elites will dominate in all spheres of the society.

The dominant elites are only a few in numbers but the machinery they possess is gigantic. If the toiling masses of the local or regional languages come with the demands for inclusion in the Eight schedule or their languages to be used in schools, colleges, universities and in offices, the dominant elites in power and position will shatter their confidence by various means. The language group in question can be termed as anti-national elements, terrorist organizations or the people who do not want national integrity or harmony. The attitude of the dominant elites affects the poor, the speakers of the local or regional varieties and pushes them into a perpetual cycle of exploitation and oppression. It is also difficult for them to explain the pain and suffering to the dominant elites who do not understand their language and come from a different cultural background. In the words of D.P. Pattanayak (1999:149), " Inequality has many faces. Giving recognition to a single language variety as standard creates a cadre of people who through various controls gain from the acquisition, processing, storage and transmission, retrieval and other manipulations of the language as the language of education, administration and mass communication in a plurilingual society bestows advantages on the speakers of that language."

5. Conclusion

For the elite, the standard language unites the society. It is the language of cultural and national identity. The education and administration will have to be carried out in the standard language. For the downtrodden, standard languages divide the society. They feel that the standard language is the language of exploitation and gradually annihilate the local languages along with their cultures. According to Saxena and Mahendore (2011:149), the standard language could be used "as effective tools to maintain the status quo in power relations and control over information flow. And in all this, the spoken languages of the people, which they are proud of and in which they are confident and articulate, have naturally to be subdued and depreciated." In independent India, the eight scheduled languages or the notion of standard languages has led to the emergence of regional ruling class whose cultural hegemony over the larger masses is not very different from that of the English speaking elite or that of the dominant elite with their cultural and religious ideologies. It is true that people's social, cultural and political identity is always linked with the language that they speak.

Multi-ism (multilingualism, multiculturalism) is the backbone of India and it should be flourished in different social and cultural settings. The children of minority languages should be given their due to learn or acquire knowledge in their mother tongues. This is guaranteed in article 350 (a) in the Constitution of India. We can hope that what is being prescribed and preached in the Constitution should also be practiced in reality. Multilingual Education (MLE) is the key to success in the tribal areas and also in the areas where the state dominant language is different from the community language. Bilingual class at the initial level or incipient stage and gradually switch over from the spoken variety to the state standard language should be practiced. The appointment of a local teacher who knows both the languages; local variety and standard language, the teachers who are aware of the language dynamics of the area should be followed. The teachers should be properly trained not only in their subjects but also in the language and cultural milieu of the communities in the area before sending them in the classroom to teach the young minds. The state should maintain a balance and harmony in saving the minority language and culture.

The Eight Schedule of the Constitution lists only 22 languages and does not list many others. There should be more languages in the Eight schedule. The Sign Language of India (ISL), languages of the Andaman and Nicobar Islands, many languages of Tibeto-Burman, Dravidian, Austro-Asiatic and Indo-Aryan families should also find their places in the prestigious Eight schedule. Inclusion of language in the Eighth Schedule languages is perceived as a hallmark of its status, positive attitude and prestige. Once included in the Eight Schedule, the status of the language is beyond conflict, beyond doubt. Though many languages of India fulfill all the requirements to be listed in the Eighth Schedule Languages, there are other considerations for not allowing all these languages to enter the Eight Schedule. The time has come to revisit our strong prejudices deeply rooted in the prescriptive notions of standard language and mainstream culture. It should be looked at through a different lens, from the perspective of the minority communities, for their presence and survival in the age of globalization and Anglicization. Otherwise, at the end of the 21st century, we may not boast about the diverse languages and cultures of India.

Footnote

(1) The FOXP2 protein is extremely conserved among mammals, it acquired two amino-acid changes on the human lineage, at least one of which may have functional consequences. This is an intriguing finding, because *FOXP2* is the first gene known to be involved in the development of speech and language--https://www.nature.com/articles/nature01025

(2) Bangladesh, Serbia, Croatia many new nations emerged where language played a central role. In India, many states are linguistically organised as for example Bangla for Bengal, Gujarati for Gujarat, Marathi for Maharashtra etc.

References

Adamson, Sylvia; Ayres-Bennet, Wendy. 2011. Linguistics and Philology in the 21st Century: Introduction. *Transactions of the Philological Society Volume 109:3* (2011) 201–206.
Agnihotri, Ramakant. 2018. *Multilingualism in India.* Module for EPG Paathshala for UGC.
Althusser, L. (1971) 'Ideology and ideological state apparatuses', in L. Althusser (ed.), *Lenin and Philosophy and Other Essays*, London: New Left Books, pp. 121–176.
Bhattacharya, Kantichandra. 1989. Odiya Swatantra Bhasa Nuhe, in *Odiya Bhasa Andolana*, ed., Bansidhar Mohanty, Cuttack.
Bloch, B & Trager, G. L. 1942. *Outline of Linguistic Analysis*. Baltimore: Linguistic Society of America. Waverly Press.
Bloomfield, L. 1935. *Language*. London: Allen & Unwin.
Beams, John. 1966. *A Comparative Grammar of the Modern Aryan Languages of India*. Delhi.
Beams John., "On the Distribution of Indian Languages" in Beams, Chatterji, and Grieson., Outlines of Indian Philology and Other Philological Papers, p.24-25.
Chomsky, N. 1957. *Syntactic Structures.* The Hague: Mouton.
Chomsky, Noam. 1959: Review of B. F. Skinner, Verbal Behaviour. In *Language:* 35, 26-58.
Chomsky, Noam. 2006. *Language and Mind*. Third edition. Cambridge University Press. Cambridge, New York.
Fairclough, N. (1989) *Language and Power* (2nd edn 2001), London: Longman.
Fairclough, N. (1992) *Discourse and Social Change*, Cambridge, UK: Polity.
Fairclough, N. (2010) *Critical Discourse Analysis: the Critical Study of Language*, 2nd edn, London: Routledge.
Fairclough, N. and Wodak, R. (1997) 'Critical Discourse Analysis', in T. van Dijk (ed.), *Discourse as Social Interaction*, London: Sage, pp. 258–285.
Foucault, M. (1980) *Power/Knowledge: Selected Interviews and other Writings 1972– 1977*, New York: Pantheon.
Fowler, R., Hodge, R., Kress, G. and Trew, T. (1979) *Language and Control*, London: Routledge.

Fromkin, Victoria & et al. 2011. *An Introduction to Language.* 9th edition. Wadsworth Cengage Learning.
Gramsci, A. (1971) *Selections from the Prison Notebooks*, ed. and trans. Q. Hoare and G. Nowell-Smith, London: Lawrence & Wishart.
Grierson, J. A. 1886-1927. *Linguistic Survey of India*. Reprinted, 1977. Motilal Benarsidas, Delhi.
Gupta, R. S.; Abbi, Anvita; Aggarwal, Kailash S. 1995. *Language and the State: Perspectives on the Eight Schedule.* (edited). Creative Books. New Delhi.
Hall, R. A. 1968. *An Essay on Language.* Philadelphia and New York. Chilton Books.
Kracht, M. 2007. Introduction to Linguistics. Department of Linguistics, UCLA
http://www.linguistics.ucla.edu/people/Kracht/courses/ling20 - fall07/ling-intro.pdf
Lehmann, Winfred P. 1984. Philology to Linguistics: Constructive to Literary Study. *South Central Review*, Spring - Summer, 1984, Vol. 1, No. 1/2 (Spring - Summer, 1984), pp. 131-140. The Johns Hopkins University Press.
Lewis, Paul M. and Gary F. Simons and Charles D. Fennig. 2013. *Ethnologue: Languages of the world.* In Lewis, M. Paul and Simons, Gary F. and Fennig, Charles D. (eds.) 17th edn. Dallas, TX: Dallas: SIL International.
Lyons, John. 1981 (2001). *Language and Linguistics: An Introduction.* Cambridge University Press. South Asian edition.
Marx, K. [with Frederick Engels] ([1933] 1965) *The German Ideology*, ed. and trans. S. Ryazanskaya, London: Lawrence & Wishart.
Pattanayak, D. P. 1991. Multilingualism in India. *Multilingual Matters*, 61.
Radford, Andrew & et al.2009. *Linguistics: An Introduction.* Second Edition. CUP: Cambridge.
Robins, R. H. 1979. *General Linguistics: An Introductory Survey.* Third edition. London: Longman.
Sapir, Edward. 1921. *Language.* New York: Harcourt Base.
Saussure, Ferdinand de. 1915. *A Course in General Linguistics. Philosophical Library. New York.*
Skinner, B. F. 1957. *Verbal Behaviour*. New York. Appleton Crofts.
Skinner, B. F. 1957. *Verbal behavior.* Appleton-Century, Crofts. https://doi.org/10.1037/11256-000
Sreevatsan, Ajai. 2014. Where are the Sanskrit Speakers? *The Hindu.* 2014. August. 10.
Saxena, Sadhana and Mahendru, Kamal. 1995. Politics of Language. In Gupta, R. S.; Abbi, Anvita; Aggarwal, Kailash S. 1995. *Language and the State: Perspectives on the Eight Schedule.* (edited). Creative Books. New Delhi.
Sen, Sukumar. 1992. *History of Bengali Literature*. New Delhi: Sahitya Akademi.
Sacks, H., Schegloff, E. and Jefferson, G. (1974) 'A simplest systematics for the organization of turn-taking in conversation', *Language 50* (*4*): 696–735.
Simpson et al. 2019. Language and Power: A Resource Book for Students. 2nd edition. Routledge English Language Intriduction. Routledge, Taylor and Francis Group. London and New York.
Thomas, Lewis. 1978. *The Lives of a Cell*. Penguin Books.
Thompson, Hanna Ruth. 2012. *Bengali.* Amsterdam: John Benjamins Publishing Company.
Turner, James. 2014. *Philology: The Forgotten Origins of the Modern Humanities*. Princeton, New Jersey: Princeton University Press.
Van Dijk, T. (2008) *Discourse and Power*, Basingstoke, UK: Palgrave Macmillan.
Wodak, R. and Meyer, M. (2015) *Methods of Critical Discourse Studies*, London: Sage.
THE CONSTITUTION OF INDIA [As on 31st July, 2014] 2014, GOVERNMENT OF INDIA MINISTRY OF LAW AND JUSTICE LEGISLATIVE DEPARTMENT
District Census Handbook Kochbihar. 2011. Census of India. West Bengal. Series 20. Part XII-A. Directorate of Census Operations West Bengal.
https://www.mha.gov.in/sites/default/files/EighthSchedule_19052017.pdf
https://www.ethnologue.com/ethnoblog/mpl/languages-world-numbers
https://en.wikipedia.org/wiki/A_language_is_a_dialect_with_an_army_and_navy
https://www.mha.gov.in/sites/default/files/EighthSchedule_19052017.pdf
https://en.wikipedia.org/wiki/Max_Weinreich

SEL

Society for Endangered
and lesser known Languages

www.selindia.org

Society for Endangered and Lesser known Languages
Bharat Ashram, Raja Bazar, Lucknow-226003

Email: contactselindia@gmail.com | Phone: +91 9940138120