Proceedings of Formal Approaches to South Asian Linguistics 15

Held in 2025 at University of Texas at Austin

April 2026

Edited by Sampada Deshpande, Ahmad Jabbar, Jonathan Paramore, & Shaunak Phadnis

Preface

The 15th annual meeting of the (Formal) Approaches to South Asian Languages was held at the University of Texas, Austin, hosted by their Linguistics Department. The conference was a three-day event (April 11-13), starting on an overcast Friday and concluding on a sunny Sunday. Like the variety in weather, the program covered a range of topics and sub-fields. This compiled volume serves to tell part of the story.

The papers from the main session featured in this volume explore left-adjoined correlatives in Bangla, honorification in Assamese and Hindi-Urdu, the aspectual system in Hindi, paradigm gaps in Urdu, information structure in Munda languages, pronominal enclitics in Sindhi and Brahui, prosodic licensing in Katki and Sambalpuri, geminate stops in Malayalam, a focus-sensitive scalar enclitic in Tamil, a degree modifier in Tiwa, Bangla hyper-raising, and *tough*-movement in Vedic Sanskrit.

The recent boom in research on discourse particles in South Asian languages also provided an opportunity for *Discourse Dynamics*, a focused workshop held on the second day of the conference. The papers from the workshop contained in this volume explore rich empirical and theoretical issues, such as negative meaning without negation, referring to and confirming propositions, and incorporation of mirativity within the Table model of conversation.

Invited talks provided another great venue for deep engagement with linguistic problems. While Sadhwi Srinivas talked about Kannada bare singulars in Episodic Contexts, Aditi Lahiri explored Phonological Transparency and Opacity, with a focus on Bengali data. Shakuntala Mahanta kicked off the final day with her work on Focus Prosody in Boro and Dimasa. Each talk was as interesting as it was exemplary.

The conference would not have been possible without Ashwini Deo, who spearheaded the organization, and was aided by the Linguistics Faculty, Staff, and student volunteers at the University of Texas, Austin. More specifically, the editors would like to thank David Beaver, John Beavers, and Steve Wechsler from the faculty, Benjamin Rapstine and Erica Alvarez from the staff, and the student volunteers, Sampada Deshpande, Gabriel Gallinate, Youn-Gyu Park, and Asher Zheng. We also thank the Program Committee for their help in reviewing the abstracts. Finally, we gratefully acknowledge the support of the South Asia Institute at UT Austin.

We hope that this volume serves as a record of the present exciting moment in South Asian linguistics, inviting further inquiry into the many open problems.

Sampada Deshpande
Ahmad Jabbar
Jonathan Paramore
Shaunak Phadnis

Contents

A Paradigm Gap in Urdu

FARAH ADEEBA, *University of Massachusetts Amherst and University of Engineering & Technology, Lahore*
RAJESH BHATT, *University of Massachusetts Amherst*

ABSTRACT

In this paper, we document a paradigm gap in the combinatorial possibilities of verbs and aspect in Urdu: the perfective form of the *-ya: kar* construction (e.g., *ro-ya: ki:* 'cry-Pfv do.Pfv') is sharply ungrammatical in modern Urdu and Hindi, despite being freely attested in 19th-century literature. We investigate this diachronic shift through historical text analysis, a large-scale corpus study—which confirms the stark absence of perfective forms—and subjective evaluation tasks with native speakers, who judge perfective examples as highly unnatural. We argue that this gap arose from a fundamental morphosyntactic conflict: the construction's requirement for a nominative subject and an invariant participle clashes with the core grammatical rule that transitive perfectives assign ergative case. This conflict rendered the perfective form unstable, and its functional replacement by other constructions allowed the gap to become entrenched in the modern grammar.

1 Introduction

Human languages are dynamic systems that continually evolve, resulting in the emergence, change, and sometimes complete disappearance of morphological and grammatical structures. Within this diachronic landscape, the phenomenon of *paradigm gaps*—systematic absences of expected word forms or constructions—presents a particularly intriguing puzzle for linguistic theory (Albright 2003, Sims 2006 , Bermel & Knittl 2012). Such gaps challenge models of language production, acquisition, and change, as they represent a failure of the grammar to generate a logically possible form. A notable example is provided by İleri & Demirok (2022), who identify a gap in the third person plural paradigm of desiderative constructions in Turkish, a pattern which speakers consistently avoid or deem unacceptable.

This paper investigates the existence of a gap in the combinatorial possibilities of a verb and aspect in Urdu. In general, in Urdu all verbs combine with the full range of aspectual/tense options i.e. given an unfamiliar verb, an Urdu speaker can use it in the perfective, the imperfective, the subjunctive, the future and so on. Given this, it comes as a surprise that the *-ya: kar* '-Pfv do' construction cannot combine with the perfective.

(1) Vo pehrõ ro-ya: kar-ta: tha:
Dem hours cry-Pfv do-Impfv.MSg be.Pst.Msg
He used to cry for hours.

(2) *Vo pehrõ ro-ya: kiya:
Dem hours cry-Pfv do.Pfv.MSg
Intended: He cried for hours.

The puzzle deepens with the finding that this gap is a historical innovation. Through analysis of 19th-century Urdu literature, we show that the perfective *-ya: kar* construction was not only possible but actively used (3).

(3) Vo raat bhar ro-ya: ki: (Umrao Jaan Ada, Mir Hadi 'Ruswa', 1899)
Dem night all cry-Pfv do.Pfv.F
'She cried/kept crying all night.'

This diachronic shift raises two central research questions:

1. What factors led to the loss of the perfective *-ya: kar* form in modern Urdu?
2. Why has this specific gap persisted, and how is it learned by speakers, given that it does not create an expressivity problem?

We propose that the gap is motivated by a fundamental *morphosyntactic conflict*. The *-ya: kar* construction exhibits two invariant properties: the subject appears in the nominative case, and the participial verb *-ya:* does not agree with the subject. This configuration clashes with a core rule of Urdu grammar: transitive verbs in the perfective aspect typically assign ergative case to their subject. We propose that this conflict rendered the perfective form unstable, leading speakers to avoid it over time. Its semantic function was seamlessly absorbed by other available forms, such as the simple perfective or the *-ta: rah-* construction, thus preventing an ineffability problem and allowing the gap to become entrenched.

To support this argument, we present evidence from multiple sources: a qualitative analysis of historical texts establishing the construction's prior existence, a quantitative corpus study demonstrating its stark absence in modern Urdu, and native speaker judgment tasks confirming its contemporary ungrammaticality.

The paper is structured as follows: Section 2 outlines the two case studies of lost generalizations in Urdu, focusing on the *ca:h-* and *-ya: kar* construction. Section 3 details the syntactic and semantic properties of *-ya: kar* and its historical attestation. Section 4 introduces the *-ta: rah-* construction, which shows no such gap. Section 5 presents our analysis of the case-based motivation for the gap. Section 6 discusses evaluation of *-ya: kar* construction by native speakers. Section 7 concludes. This case study offers a clear model for how paradigm gaps can emerge and stabilize diachronically, driven by syntactic conflict and resolved by competition from existing grammatical forms.

2 Loss of generalization

This section examines how a once broadly applicable grammatical construction becomes restricted in its usage over time. A general pattern loses its ability to combine with a wider

range of linguistic elements (like verbs or aspects), becoming specialized or even ungrammatical in previously acceptable contexts. This process of reduced applicability demonstrates that language change involves not only innovation but also the narrowing and specialization of existing grammatical structures.

2.1 Loss of a subcategorization frame

The verb *ca:h-* 'want' provides an example of semantic and syntactic change over time in Hindi-Urdu. The following examples, attested up until the early 20th century, illustrate an earlier usage of *ca:h-* where it conveyed an 'about to' or 'imminent' meaning when combined with a perfective verb form:

(4) a. ghaṛi: 12 baj-a: ca:h-ti: hai
clock.F 12 ring-Pfv want-Impfv.F be.Prs.Sg
'The clock is about to strike 12.'

b. (Small, 1895)
vo kal aya: ca:h-te hain
Dem tomorrow come-Pfv want-Impfv.M be.Prs.Pl
'They will come tomorrow.'

c. asar ki: nama:z=ka: waqt hu-a: ca:h-ta: hai
Asar Gen.f prayer=Gen time be-Pfv want-Impfv.M.Sg be.Prs.Sg
'It is about to be time for the Asar prayer.'

(5) a. infinitive:
vo dubai ja:-na: ca:h-ti: hai
Dem Dubai go-Inf want-Impfv.F be.Prs.Sg
'She wants to go to Dubai'

b. subjunctive:
vo ca:h-ti: hai ki mEN dubai ja:-ũ
Dem want-Impfv.F be.Prs.Sg that I Dubai go-Sbjv.1Sg
'She wants that I go to Dubai'

Now, however, only the desire meaning of *ca:h* remains and it takes an infinitival/subjunctive complement.

2.2 Loss of a particular aspectual form

Consider the *-(y)a: kar* construction in contemporary Hindi-Urdu. This construction combines with a range of aspects.

(6) a. Imperfective:

vo ro:z sku:l ja:-ya: kar-ti: hai
Dem everyday school go-Pfv do-Impfv.F be.Prs.Sg

'She goes to school every day'

b. Subjunctive:

mEN ca:h-ta: hũ: ki vo ro:z sku:l ja:ya:
1Sg want-Impfv.M be.Prs.1Sg that Dem everyday school go-Pfv
kare
do.Subjv.3Sg

'I want that she/he go to school everyday'

c. Future:

agle mahi:ne=se vo ro:z sku:l ja:-ya: kare-gi:
next month=from Dem everyday school go-Pfv do.Subjv.3Sg-Fut.F

'From next month, she will go to school everyday'

d. Infinitive:

us-ka: yahã: ro:z a:-ya: kar-na: mujhe qat'ii pasand nahĩ
3Sg-Gen here everyday come-Pfv do-Inf me.Dat absolutely like not

'I don't like his/her coming here everyday one bit'

e. Polite Imperative:

a:p roz exercise zaru:r kiya: ki:jiye
you.Hon everyday exercise definitely do.Pfv do.Imp.Pol

'Please do exercise every day'

f. Plain Imperative:

waqt se sku:l ja:ya: kar
time from school go.Pfv.M do.Impfv.2Sg.Imp

'Go to school on time'

g. Future Imperative:

ro:z sku:l ja:-ya: kar-na:
daily school go.Pfv.M do-Inf

'To go to school daily'

However, it does not combine with the perfective and progressive.

(7) *Perfective:

*vo ro:z sku:l ja:ya: ki:
Dem everyday school go.Pfv.M did.Pfv.F.Sg

'She used to go to school everyday.'

(8) *Progressive:

*vo ro:z sku:l ja:-ya: kar rahi: hai
Dem everyday school go-Pfv do Prog.F be.Prs.Sg

'She goes to school everyday (as a routine).'

The ungrammaticality of the perfective form is a recent development in the history of the language. McGregor (1972)'s excellent Hindi grammar, published in 1972, reports instances of the *-ya: kar* construction in the perfective but notes that they are rare (p. 137). In the 50 years since, the combination has gone from rare to non-existent for contemporary speakers, while the construction remains alive and well in other tenses. A further reason for not entertaining a semantic incompatibility argument for the ungrammaticality of the perfective is that the semantically close *-ta: rah-* '-Impfv stay-' construction freely combines with the perfective, as shown in Section 4.

The impossibility with the progressive can perhaps be explained on semantic grounds. However, the impossibility of the combination with the perfective remains to be explained and is the subject of this talk.

To understand the current constraints on the *-ya: kar* construction, particularly its incompatibility with the perfective aspect, we must consider its historical evolution. The examples (with book title, author, and year of publication indicated in parentheses) presented below, drawn from this period, offer concrete instances of the perfective aspect occurring with the *-ya: kar* construction, raising the question of why it is no longer possible.

(9) *ja:-ya: kĩ:* (Khurshid Bahu, Mirza Hadi 'Ruswa')

tum aur pya:-ri: sa:-ji-dah is maka:n meN a:-ya: ja:-ya:
2Sg.Pron.Inf and dear.F.Sg Sajida this house in come-Pfv.M go-Pfv.M
kĩ:
did.Impfv.Pl

'You and dear Sajida used to come and go in this house.'

(10) *rakh-a: kĩ* (Umrao Jan Ada, Mirza Hadi 'Ruswa', 1899)

mere na:m=ka: ta'ziyah kha:-num mar-te dam tak rakh-a:
my name=GEN taziyah Khanum die-Impfv.M breath until keep-Pfv.M.Sg
kĩ
do-Impfv.Pl

'Khanum kept the taziyah of my name until her last breath.'

(11) *taR-pa: ki:* (Fasana-e-Ajaib, Rajab Ali Baig Suroor, 1844)

yeh sun kar tama:m shab taR-pa: ki:
this hear having all night writhe-Pfv.M.Sg did.Pfv.F.Sg

'Having heard this, she writhed all night.'

(12) *dekha: kiye* (Urdu-e-Mualla, Mirza Ghalib, 1869)

agast=ke mahi:ne=ka: ha:l kuchh ma'lu:m nahĩ kal sha:m=ko
August=GEN month=GEN state some known not yesterday evening=GEN
do do moonDhe rakh kar ka-i: a:d-mi: de-kha: kiye
two two stool keep having several person see.Pfv.M did.Impfv.Pl

'The situation of the month of August is unknown-just last evening, several people were seen stacking two stools on top of each other.'

(13) *soca: ki:* (Ayama, Deputy Nazir Ahmad, 1891)

akel-i: paR-i: kuchh socha: ki:
alone-F fall-Pfv. some think.Pfv.M did.Pfv.F

'Lying alone, she was thinking something.'

The contrast between the modern ungrammaticality in (7) and the historical attestations in (9-13) establishes a clear case of diachronic change resulting in a paradigm gap. This shift requires an explanation. The following sections will argue that this gap arose from a fundamental morphosyntactic conflict within the construction itself, specifically related to case assignment and the invariant nature of the *-ya:* participle.

3 The *-ya: kar* construction

3.1 Semantics

The historical usage of the perfective *-ya: kar* construction reveals a distinct and specific semantic profile, characterized by two core features: (i) its strong association with temporal adverbials that define a bounded timeframe, in many—though not all—cases, the construction co-occurs with durational adverbs like *ra:t bha:r* 'all night', *der tak* 'until late', and (ii) its strong preference for atelic predicates. The convergence of these features created a specific aspectual niche that we term *bounded atelicity* (Depraetere, 1995).

As illustrated by the examples presented below (with book titles, authors, and years of publication indicated in parentheses), this construction frequently appears alongside durational adverbials in literary texts.

(14) *ra:t bhar* (*Umrao Jan Ada*, Mirza Hadi 'Ruswa', 1899)

meN khud ra:t bhar ro-ya: ki:
1Sg self night full cry-Pfv.M.Sg did.Pfv.F.Sg

'I myself cried all night.'

(15) *der tak* (*Ahmaqullazin*, Munshi Sajjad Hussain, 1906)

der tak dukh dard ka-ha: kiye
long until sorrow pain say.Pfv.M.Pl did.Impfv.Pl

'They kept talking about sorrow and pain for a long time.'

(16) *ghantõ* (*Firdaus-e-Barin*, Abdul Halim Sharar, 1899)

aur khud bhi: te:re: sa:th ghaNton khaR-ii: ro-ya:
and self also.Part your.2Sg.Poss with.Postp hours stand-Pfv.F cry-Pfv.M.Sg
ki:
did.Pfv.F.Sg

'And I myself also stood and cried with you for hours.'

A list of adverbs used in the construction is provided below.

1. *chay sa:l tak*
   'for six years'
2. *der tak*
   'until late'
3. *is samay*
   'at this time'
4. *Ek mint*
   'one minute'
5. *barsõ*
   'years'
6. *badi: de:r*
   'for a very long time'
7. *bohat de:r tak*
   'for a very long time'
8. *pa:nc pehar*
   'for 5 pehar i.e. 15 hours'
9. *tama:m shab*
   'all night'
10. *ca:r din*
    'four days'
11. *cand lamhõ*
    'a few moments'
12. *cand mint*
    'a few minutes'
13. *din bhar*
    'all day'
14. *din caRhe*
    'until sunrise'
15. *do sa:l tak*
    'for two years'
16. *ra:t bhar*
    'all night'
17. *ra:t ko*
    'at night'
18. *sa:re: sa:re: din*
    'all day'
19. *shab ko*
    'at night'
20. *sadiyõ talak*
    'for centuries'
21. *kuch dinõ*
    'a few days'
22. *kuch der*
    'some time'
23. *kal sha:m*
    'last evening'
24. *guzray huay bars*
    'in years gone by'
25. *ghaRiyõ*
    'moments'
26. *ghantõ*
    'hours'
27. *marte: dam tak*
    'until dying'
28. *mahi:nõ*
    'months'

This specific semantic niche—the quantification of an atelic activity within a bounded timeframe—is key to understanding the construction's eventual loss. Its functional space overlapped with two more general and stable constructions: 1. The simple imperfective for habituality (e.g., *ja:-ti thi* 'used to go'). 2. The simple perfective for bounded atelic events, often with a durational adverb (e.g., *vo ra:t bhar roi* 'she cried all night').

We posit that this very specific semantic role made the perfective *-ya: kar* form functionally redundant. When the morphosyntactic conflict (Section 3.2) placed pressure on the construction, its narrowly defined meaning provided no functional incentive for speakers to preserve it. The language already had simpler, unproblematic ways to express the

same concept of a bounded atelic event, leading to the construction's disappearance from the perfective paradigm.

3.2 Syntax

The historical data reveals two critical and consistent syntactic properties of the perfective *-ya: kar* construction that are fundamental to explaining its eventual loss. These properties directly conflict with the core rules of Urdu grammar governing perfective transitives, creating an unstable configuration for speakers.

1. Nominative Case Assignment. Despite containing a transitive verb and a perfective auxiliary, the construction consistently requires its subject to be in the **nominative case**. This violates the standard rule of Urdu grammar where a transitive verb in the perfective aspect assigns **ergative** case to its subject (Butt & Deo, 2017). This holds true even when the embedded verb is clearly transitive, as shown in examples (17) and (20) below.

2. Invariant Participle. The *-ya:* element, a perfective participle, remains morphologically **invariant**. It is frozen in the default masculine singular form and does not agree in gender, number, or case with the subject or the object. This contrasts sharply with simple perfective verbs, which must agree with the object in gender and number.

The following examples from literary texts illustrate these properties. Critically, the subjects (e.g., *ja'fari:*, *Xa:-nam*) are nominative, and the participles (*ka-ha:*, *rakh-a:*) are invariant masculine singular forms, even when the context or the auxiliary's agreement suggests a feminine or plural subject.

(17) *ki:-ya: ki:* (*Fasana-e-Ajaib*, Rajab Ali Baig Suroor, 1844)

der tak akhlaaq o muhabbat malika:=ka: mazku:r kiya: ki:
long until ethics and love Queen=GEN mention do.Pfv.M.Sg do.Pfv.F.Sg

'For a long time, she kept mentioning Queen's ethics and love.'

(18) *jhap-ka:-ya: kiye* (*Aag ka Dariya*, Qurat-ul-Ain Haider,1959)

dono kuch der tak cup ca:p andhere=mẽ palkẽ jhap-kaya: kiye
both some delay till silently darkness=in eyelids blink-Pfv do.Pfv.MPl

'Both blinked their eyelids for a while in the darkness in silence.'

(19) *ka-ha: ki:* (*Akhteri Begum*, Mirza Hadi 'Ruswa')

ja'fari: ap-ni: Dhi-Ta:-i: se jhu:T ka-ha: ki:
Jafari self-Poss.F stubbornness-F from lie say-Pfv did-Pfv.F.Sg

'Jafari, with her stubbornness, kept telling lies.'

(20) *rakh-a: kĩ:*(*Umrao Jan Ada*, Mirza Hadi 'Ruswa', 1899)

mere na:m=ka: ta'ziyah kha:-num mar-te dam tak rakh-a:
my name=GEN taziyah Khanum die-Impfv.M breath until keep-Pfv
kĩ:
do-Pfv.F.Pl

'Khanum kept the taziyah of my name until her last breath.'

(21) *dekh-a: kiye* (*Urdu-e-Mualla*, Mirza Ghalib, 1869)

agast=ke mahi:ne=ka: ha:l kuchh ma'lu:m nahĩ kal sha:m=ko
August=GEN month=GEN state some known not yesterday evening=GEN
do do moonDhe rakh kar ka-i: a:d-mi: dekh-a: kiye
two two stool keep having several person see-Pfv do.Pfv.M.Pl

'The situation of the month of August is unknown-just last evening, several people were seen stacking two stools on top of each other.'

(22) *pi:s-a: kiya:* (*Safeena-e Gham-e Dil*, Qurat-ul-Ain Haider, 1953)

do saal tak mEN cakki: pi:s-a: ki-ya: aur meN-ne navaaRen
two year until 1Sg grinding.stone grind-Pfv do.Pfv.MSg and 1Sg-Erg tapes
bu-nĩ
weave-Pfv.F.Pl

'For two years, I ground the millstone and I wove the yarn.'

These two properties — obligatory nominative subjects and an invariant participle — define the syntactic signature of the construction. However, they also created a conflict for speakers by requiring violation of the wide-ranging generalization that a perfective transitive clause have an ergative subject. This placed the construction under pressure. We propose that this underlying instability is the primary cause for its disappearance from the paradigm; speakers resolved the conflict by avoiding the construction entirely in the perfective, where case assignment paradox created a conflict.

4 The *-ta: rah-* construction: A robust alternative

A crucial question arising from the paradigm gap in *-ya: kar* is whether it led to a loss of expressive power in the language. This can be examined by analyzing the *-ta: rah-* ('keep V-ing') construction, which exhibits no such gap and serves as a functional alternative for expressing sustained or continuous action. Its robustness across the entire aspectual paradigm is demonstrated below using the verb *ja:-* ('to go').

(23) a. Perfective:

vo cup ca:p sku:l ja:-ti: rah-i:
Dem silent silently school go-Impfv.F stay-Impfv.F

'She kept going to school silently.'

b. Imperfective:

vo sku:l ja:-ti: rah-ti: hai
Dem school go-Impfv.F stay-Impfv.F be.Prs.Sg

'She keeps going to school .'

c. Subjunctive:

meN ca:h-ta: hũ: ki vo ek sa:l tak sku:l ja:-ti: rah-e
1Sg want-Impfv.M be.Prs.1Sg that Dem one year until school go-Impfv.F stay-Sbjv.3Sg

'I want that she keeps going to school for one year.'

d. Future:

vo das sa:l tak sku:l ja:-ti: rah-e-gi:
Dem ten years until school go-Impfv.F stay-Sbjv.3Sg-Fut.F

'She will keep going to school for ten years.'

e. Infinitive:

sku:l ja:-te reh-na: us=ki: a:dat hai
school go-Impfv.F stay-Inf 3Sg-Poss habit be.Prs.Sg

'Going to school is her habit.'

f. Polite Imperative:

ro:za:na: sku:l ja:-ti: rah-iye-ga:
daily school go-Impfv.F stay-Sbjv.Pol-Fut

'Please keep going to school daily.'

g. Plain Imperative

cup kar aur sku:l ja:-ti: rah
silent do.Imp.Inf and school go-Impfv.F stay.Imp

'Be quiet and keep going to school.'

h. Future Imperative

tum ek sa:l sku:l ja:-ti: rah-na:
2Sg one year school go-Impfv.F stay-Inf

'You keep going to school for one year.'

The stark contrast between the two constructions is revealed in a minimal pair. The meaning intended by the lost perfective *-ya: kar* form is readily expressed by the perfective form of *-ta: rah-*.

(24) a. Form not acceptable now

vo ro:z sku:l ja:-ya: ki:
Dem everyday school go.Pfv.M did.Pfv.F.Sg

‘She used to go to school everyday.’

b. Counterpart with *-ta: rah-*

vo ro:z sku:l ja:-ti: rah-i:
Dem everyday school go-Impfv.F stay-Pfv.F.Sg

‘She used to keep going to school everyday.’

The *-ta: rah-* construction, unlike *-ya: kar*, readily combines with the perfective aspect to express a continuous or sustained action in the past, similar to the intended meaning of the lost *-ya: kar* form in the perfective. This suggests that the semantic function associated with a past habitual or continuous action is still expressible in the language through an alternative grammatical construction.

5 Why the gap?

İleri & Demirok (2022) propose that the observed paradigm gap within Turkish desiderative constructions in the third-person plural arises from a confluence of factors. Firstly, the near-absence of attested instances of the *-AsI* + 3PL suffix combination in corpus data suggests a paucity of positive evidence available to learners. This lack of direct input, while not explicitly negative, may effectively function as implicit negative evidence, hindering the acquisition and production of the expected forms. Secondly, the presence of alternative, competing forms for the expression of third-person plural desideratives further contributes to the marginalization and eventual paradigm gap of the *-AsI* construction. The availability of these alternative forms reduces the necessity and, consequently, the frequency of the targeted construction, ultimately leading to its diminished usage and potential grammaticalization of the gap. Following the approach, we argue that the loss of the perfective *-ya: kar* form in Urdu can be attributed to a fundamental **syntactic conflict**. This conflict made the construction unstable, and its functional replaceability (shown in Section 4) allowed it to be abandoned without loss of expressivity.

5.1 Motivating the gap in Urdu

The diachronic absence of the perfective *-ya: kar* construction is best explained by a morphosyntactic conflict. The construction’s internal requirements directly contradict a core rule of Urdu grammar.

1. **General Rule:** A transitive verb in the perfective aspect (*kar*) requires its subject to be in the **ergative** case.

2. **Construction-Specific Rule:** The *-ya: kar* construction, even in its perfective historical form, consistently requires its subject to be in the **nominative** case, as evidenced by the examples presented earlier (e.g., the subject *ja’fari:* in (19), or *Xa:nam* in (20))s.

This conflict is not merely theoretical. It is exacerbated when the embedded verb (*-ya:*) is itself transitive. In such cases, **both verbs semantically imply an agentive subject**, yet the construction forbids the ergative case that would typically mark that agent. For instance, in the Example 17 the embedded verb *mazku:r kiya:* 'mention' is transitive, yet its implied subject (the narrator) remains nominative with the *-ya: kar* frame. This creates a unresolvable tension for the speaker. This clash in case assignment expectations for a single underlying subject likely rendered the perfective *-ya: kar* construction grammatically unstable, leading to its avoidance by speakers. The availability of alternative means to express habitual past actions (as seen in the modern avoidance and potential use of simple perfective with adverbs) further facilitated this disuse, resulting in the observed paradigm gap in contemporary Hindi-Urdu.

Our analysis of historical attestations provides quantitative support for this point of tension. The construction appeared with both transitive and intransitive verbs as shown in Table 1, Table 2 and Table 3.

Table 1: Frequency and Transitivity of Verbs with *kar* Forms

Verb	***kar* Form**	**Count**	**Transitivity**	**Counter Example**
a:-ya:	ki:	1	intransitive	
baj-a:	kiye	1	intransitive	
badl-a:	kiya:	1	intransitive	
bars-a:	ki:	1	intransitive	
bhar-a:	kiye	1	intransitive	it can be transitive e.g *mEn ne gla:s men pani: bha-ra:* 'I filled water into the glass'
bhunbhana:-ya:	ki:	1	intransitive	
bah-a:	kiya:	1	intransitive	it can be transitive e.g *taiz hawa dhu:l ba-ha: le gayi:* 'The fast wind blew away the dust'
phir-a:	ki:	1	intransitive	
phir-a:	kiya:	1	intransitive	
taRp-a:	ki:	1	intransitive	
taRp-a:	kiya:	2	intransitive	
ter-a:	kiya:	3	intransitive	
ter-a:	kiye	1	intransitive	
Takra:-ya:	ki:	1	intransitive	
TamTama:-ya:	kĩ	1	intransitive	
Tehl-a:	ki:	1	intransitive	
Tehl-a:	kiye	1	intransitive	

Continued on next page

Table 1: Frequency and Transitivity of Verbs with *kar* Forms

Verb	***kar* Form**	**Count**	**Transitivity**	**Counter Example**
ja-ya:	kĩ	1	intransitive	
jag-a:	ki:	1	intransitive	
jhilmila:-ya:	ki:	1	intransitive	
carcara:-ya:	kĩ	1	intransitive	
cal-a:	ki:	1	intransitive	
cal-a:	kiya:	1	intransitive	
cal-a:	kĩ	1	intransitive	
cal-a:	kiye	1	intransitive	
ro-ya:	ki:	7	intransitive	it can be transitive e.g. *wo ap-ni: qismat ko roya:* 'He cried about his misfortune.'
ro-ya:	kiya:	1	intransitive	
ro-ya:	kĩ	1	intransitive	
so-ca:	kĩ	1	intransitive	
so-ya:	ki:	1	intransitive	
so-ya:	kiya:	1	intransitive	
kha:-ya:	ki:	1	intransitive	
khel-a:	ki:	1	intransitive	it can be transitive e.g *ham ne krikaT ka mac khel-a:* 'We played a cricket match'
kiya:	kiye	3	intransitive	
ga-ya:	ki:	1	intransitive	
ga-ya:	kiye	1	intransitive	
guzr-a:	kiye	1	intransitive	
ghũj-a:	kĩ	1	intransitive	
la:-ya:	kiye	1	intransitive	it can be transitive e.g. *wo seb la-ya:* 'He brought apples.'
muskara:-ya:	kĩ	1	intransitive	
na:c-a:	ki:	1	intransitive	
hãs-a:	ki:	1	intransitive	
hu-wa:	kiya:	1	intransitive	
hu:-wa:	kĩ	4	intransitive	

Continued on next page

Furthermore, two verbs appeared in both transitive and intransitive frames as shown

in Table 3 (e.g., *soc-a:* ‘thought (about something)’ vs. ‘thought’), demonstrating that the conflict was not limited to a few lexical items but was a pervasive structural feature of the construction.

Table 2: Frequency and Transitivity of Verbs with *kar* Forms

Verb	***kar* Form**	**Count**	**Transitivity**	**Counter Example**
baja:-ya:	kiya:	1	transitive	
baha-ya:	ki:	1	transitive	
behla:-ya:	ki:	1	transitive	
paRh-a:	ki:	1	transitive	
phũk-a:	kiye	1	transitive	
pher-a:	kĩ	1	transitive	
pehn-a:	kĩ	1	transitive	
pes-a:	kiya:	1	transitive	
jhapka-ya:	kiye	1	transitive	
cheR-a:	kiya:	1	transitive	
dekh-a:	ki:	7	transitive	
dekh-a:	kiya:	8	transitive	
dekh-a:	kiye	8	transitive	
rakh-a:	kĩ	1	transitive	
sun-a:	ki:	3	transitive	
sun-a:	kiya:	4	transitive	
sun-a:	kĩ	3	transitive	
sun-a:	kiye	1	transitive	
suna:-ya:	kiya:	1	transitive	
soc-a:	ki:	2	transitive	
kah-a:	ki:	1	transitive	
kah-a:	kĩ	1	transitive	
kah-a:	kiye	2	transitive	
ki-ya:	ki:	1	transitive	
ghur-a:	kiye	1	transitive	
laR-a:	kĩ	1	transitive	
laR-a:	kiye	1	transitive	
laga:-ya:	ki:	1	transitive	
laga:-ya:	kiye	1	transitive	
ma:r-a:	kiye	1	transitive	
mil-a:	kiye	1	transitive	

Continued on next page

Table 3: Verbs appeared as both transitive and intransitive

Verb	"Kar" Form	Transitive Count / Intransitive Count
soc-a:	kiye	2 /2
soc-a:	kiya:	1 /1
kha:-ya:	kiye	1/1

6 *dil kar*: an exception that survives

Our explanation for the loss of *-ya: kar* in the perfective implicates the absence of the ergative in a transitive perfective context. However in contemporary Hindi-Urdu, we find the complex predicate *dil kar* 'heart do', which does not license ergative in the perfective.

(25) a. mera: dil kiya: ke mEN abhi: cala: ja:-ũ:
1Sg.Poss heart did that 1Sg now move go-Subjv.1Sg
'My heart desired that I leave right now.'

b. us waqt mera: ba:har ja:ne=ko bahut dil kiya:
that time 1Sg.Poss outside go.Inf=GEN very heart do.Pfv.M.Sg
'At that time, I really wanted to go out.'

c. sardi: itni: zi-ya:-da: thi: ke mera: garm ca:e pi:-ne=ko dil kiya:
cold so.much much was that 1Sg.Poss hot tea drink-Inf to heart
do-Pfv.M.Sg
'The cold was so much that my heart wanted to drink hot tea.'

d. us ka: dil kiya: ke vo zor se hãse
3Sg.Poss heart did that 3Sg force with laugh-Subjv.3Sg
'His/Her heart desired that he/she laugh loudly.'

In all above examples ending with *dil kiya:*, the grammatical subject of the verb *kiya:* is *dil* 'heart'. This subject is not marked with the ergative postposition *ne*. Why might this construction survive in the perfective when the *ya: kar* did not make it? We present two speculations.

1. The *dil kiya:* construction lacks a prototypical external argument (agent). As ergative marking in Urdu is typically licensed by agentive subjects of transitive verbs in the perfective aspect, the absence of such an argument in these constructions means the conditions for ergative assignment do not arise.

2. Perhaps it is relevant that *dil kiya:* construction, functioning as a more lexicalized Noun+Verb compound to express desire, exhibits less syntactic articulation compared to the *-ya: kar* structure, which is a more transparent syntactic combination of a perfective verb stem and an auxiliary to mark habitual aspect. This difference in structural composition and grammatical function likely contributes to why *kar* in *dil kiya:* does not license an ergative subject.

7 An implication for learning

Our diachronic account raises a significant synchronic question: how is this paradigm gap sustained in the mental grammar of contemporary Hindi-Urdu speakers? For simple verbs, speakers can productively apply aspectual morphology to novel or nonce words (e.g., they know that the perfective of a hypothetical verb *zimb* would be *zimb-aa*), suggesting that paradigm gaps for simplex verbs are rare and perhaps learned by negative evidence.

However, the *-ya: kar* construction is a complex, periphrastic pattern. We propose that the combinations of such complex constructions with specific aspects are likely learned not by productive rule but from **positive evidence** in the input. The modern gap persists because learners are exposed to no positive evidence for the perfective form. The absence of **-ya: ki:* in the input is interpreted as a systematic gap, not as an accident of frequency. This is bolstered by the presence of fully grammatical alternatives (*-ta: rah-*, simple perfective), which provide no functional motivation for learners to hypothesize the missing form.

Thus, the gap is sustained because the necessary triggering evidence for this specific combination is absent from the primary linguistic data, and the grammar has stabilized in a state where the construction is simply undefined for the perfective aspect.

8 Human evaluation

8.1 Experiment

This study was conducted in accordance with the principles outlined in the Declaration of Helsinki. The primary aim of this evaluation was to obtain direct, quantitative evidence from native speakers on the acceptability of the perfective *-ya: kar* construction in modern Urdu. Our hypothesis was that sentences containing the perfective form (e.g., *ro-ya: ki:*) would be judged as sharply unnatural compared to their modern counterparts expressing habitual past action.

8.1.1 Participants

Twenty-one native speakers of Urdu (11 female, 10 male) were recruited for the study. Participant age ranged from 20 to 48 years . All participants reported having completed higher education. Participation was voluntary.

8.1.2 Design and procedure

A forced-choice acceptability judgment task was designed. The stimulus set consisted of 20 unique sentences historically attested in 19th-century literature, featuring the perfective *-ya: kar* construction. For each historical sentence, a modern counterpart was created by replacing the perfective *-ya: kar* form with a grammatical alternative, most commonly the imperfective habitual form (*-ta: tha:/ti: thi:*).

Participants were presented with these 20 minimal pairs in random order. For each pair, they were asked the question: “Which sentence sounds more natural to you?” . They were forced to choose one of the two options. The experiment was administered digitally.

8.1.3 Results

The results were unequivocal and aligned perfectly with our hypothesis. Across all 20 items and 21 participants, generating 420 total data points, **not a single participant** selected the sentence containing the perfective *-ya: kar* construction as the more natural option. This represents a 0% acceptance rate for the historical form. All 420 responses (100%) favored the modern grammatical alternative.

This stark, categorical rejection of the perfective form provides powerful convergent evidence that supports our corpus findings and confirms that a robust paradigm gap exists in the modern grammar of Urdu speakers. The complete absence of any selection for the target construction indicates that its ungrammaticality is not a subtle or gradient effect but a definitive and stable feature of the contemporary linguistic system.

9 Conclusion

This study has demonstrated that the contemporary absence of the perfective *-ya: kar* construction in Hindi-Urdu is a paradigm gap resulting from grammatical change. Historical evidence from the late 19th century attests to its productive use, a usage characterized by its unique morphosyntactic signature: a nominative subject despite transitive embedded verbs and an invariant *-ya:* participle. We have argued that this very configuration created a fundamental syntactic conflict with the core ergative alignment of the perfective aspect in Hindi-Urdu. The convergence of evidence from historical texts, large-scale corpus data, and controlled native speaker judgment tasks strongly supports the hypothesis that this conflict led to the construction’s avoidance and eventual disappearance.

A critical part of our proposal is that the functional replaceability of *-ya: kar* by other constructions allowed this gap to become entrenched without any loss of expressivity. To further validate this mechanism of change, a crucial direction for future work is to conduct a comparative diachronic frequency analysis of the alternative *-ta: rah-* construction. Tracking the frequency of *-ta: rah-* in its habitual sense across the same historical corpus would allow us to test the hypothesis that its rise coincided with the decline of the perfective *-ya: kar* form. This would provide powerful quantitative evidence for the role of constructional competition and functional replacement in the creation of this paradigm gap.

As it stands though, our explanation is incomplete as we do not have an understanding as to what changed between 19th century Urdu and contemporary Urdu that led to the loss of *-ya: kar* in the perfective. The case conflict that we have implicated as being responsible for the loss existed in the 19th century as well. This unresolved puzzle we leave for future work.

References

Adeeba, Farah & Rajesh Bhatt. In progress. UMass Urdu Corpus. Work in progress, University of Massachusetts Amherst.

Albright, Adam. 2003. A quantitative study of Spanish paradigm gaps. In *Proceedings of the 22nd West Coast Conference on Formal Linguistics*, 1–14. Somerville, MA: Cascadilla Press. https://web.mit.edu/albright/www/papers/Albright-WCCFL22.pdf.

Bermel, Neil & Luděk Knittl. 2012. Corpus frequency and acceptability judgments: A study of morphosyntactic variants in czech. *Corpus Linguistics and Linguistic Theory* 8(2). 241–275. doi:doi:10.1515/cllt-2012-0010. https://doi.org/10.1515/cllt-2012-0010.

Butt, Miriam & Ashwini Deo. 2017. Developments into and out of ergativity: Indo-aryan diachrony. In *The oxford handbook of ergativity*, Oxford University Press. doi:10.1093/oxfordhb/9780198739371.013.22. https://doi.org/10.1093/oxfordhb/9780198739371.013.22.

Depraetere, Ilse. 1995. On the necessity of distinguishing between (un)boundedness and (a)telicity. *Linguistics and Philosophy* 1.

Mcgregor, R. 1972. *Outline of Hindi grammar*. doi:10.1093/oso/9780198700074.001.0001.

Sims, Andrea D. 2006. Minding the gaps: Inflectional defectiveness in a paradigmatic theory. *Journal of Greek Linguistics* 7(1). 205–209. doi:10.1075/jgl.7.16sim.

Small, George. 1895. *A Grammar of the Urdu or Hindustani Language in Its Romanized Character*. Calcutta: Thacker, Spink, and Co.

İleri, Muhammed & Ömer Demirok. 2022. A paradigm gap in Turkish. In *Proceedings of the 7th workshop on turkic and languages in contact with turkic*, .

Honorificity and Number Features are Distinct Processes: A Case in Assamese Pronouns

JOYNAL ALI, *Indian Institute of Technology, Delhi*

ABSTRACT

The focus of this paper is the relationship between honorificity and number in Assamese. The language shows a three-tiered system of honorific pronouns, while verbs mark honorific distinctions only in the second person (Masica 1991; Palash et al. 2019). Number, by contrast, is marked exclusively in the pronominal system and has no effect on verbal agreement. In particular, plural suffixes never change the honorific value of a pronoun, and verbal honorific markers remain unchanged regardless of number. These patterns indicate that honorificity and number are computed by separate processes in the grammar. I propose in Assamese, within a Minimalist framework (Chomsky 1995, 2000, 2001), that honorificity is encoded in its own functional projection (HonP) and checked via Agree (Alok 2021), while number is introduced in a higher NumP (Ritter 1991, 1992). The findings here raise a possibility that [HON]-[NUM] independence holds more widely than hitherto assumed.

1 Introduction

In theoretical syntax, agreement remains a central issue, since it shows how languages organize formal features like person, number, and gender. Although φ-features are traditionally understood to include person, number, and gender (Chomsky 1995; Adger 2003), recent work has raised the possibility that honorificity should be treated as an independent feature with its own morphosyntactic processes (cf. also Alok 2020, 2021; Kaur & Yamada 2022; Niinuma 2003). Indo-Aryan languages provide especially useful evidence for this claim, since they employ a wide range of strategies for encoding politeness, from the use of honorific suffixes on nominals to specialized pronominal forms and agreement patterns on verbs (Bhatt & Davis 2023; Wang 2023).

Within this view, Assamese presents an especially notable case. The language shows an elaborate three-tier honorific system in the pronominal paradigm (non-honorific, honorific, and high-honorific) while simultaneously showing restrictions on verbal agreement. Verbs in Assamese inflect for person and honorificity in second-person contexts, but they do not mark number. Instead, plurality is encoded exclusively in the pronominal system, through suffixes (Palash *et al*. 2019; Masica 1991) that are themselves sensitive to the honorific status of the stem. Thus, plural morphology operates independently of verbal agreement, but still interacts with the honorific system in systematic ways. This separation of honorificity and number raises a broader theoretical question: should honorificity and number be modeled as distinct syntactic features, or is one dependent on the other (Bhatt & Davis 2023)?

Earlier work has shown that in some languages, politeness distinctions rely directly on number- such as the use of plural pronouns to mark singular politeness in French or Spanish (Wang 2023), or the plural agreement triggered by the Hindi-Urdu honorific suffix *-jī* (Bhatt & Davis 2023). Assamese, however, stands apart from this cross-linguistic tendency. Plural suffixes leave the honorific value of a pronoun intact, and verbal honorific distinctions do not vary with number. This pattern points to [HON] as an independent feature, separate from [NUM], yet still involved in syntactic Agree relations within the functional spine.

The following discussion develops this claim in detail. It begins in §1 with a typological sketch of Assamese and its pronominal system. §2 reviews the theoretical background in the Minimalist framework and the role of Agree. §3 then situates Assamese within the broader Indo-Aryan context, comparing it with related languages- Magahi and Hindi-Urdu. §4 presents the core analysis, showing that honorificity is encoded on separate functional projections, HonP, and operates independently of NumP in syntactic derivations. §5 discusses the broader theoretical and typological implications of this account, and §6 closes the paper with concluding remarks.

1.1 Typology of Assamese

The canonical word order in Assamese is SOV. But speakers often scramble the order or topicalize arguments. The tense system is past vs. future, while aspect is conveyed through separate morphological markers. The language has a mixed morphological typology which shows both agglutinative and fusional morphology (Masica 1991: 326). Nouns inflect for number, classifiers, and case, but grammatical gender is not marked (Palash *et al*. 2019: 353). Verbs, however, agree with person as well as honorificity using the same morpheme in 2P form, but not with number, so the same form is used regardless of whether the subject is singular or plural. For example, in (1-3), all second-person subject pronouns (singular or plural) take the same second-person verb agreement, as illustrated by the forms of *"*come along*"* in imperative:

(1) tɔi/ tɔ.hɔt bɔl
2P:NH.SG/2P:NH.PL come.along
'You (non-honorific) come along!'

(2) tumi/tuma.lʊk bɔl-a
2P:NH.SG/2P:NH.PL come.along-2P:H.IMP
'You (honorific) come along!'

(3) apuni/apuna.lʊk bɔl-ɔk
2P:HH.SG/2P:HH.PL come.along-2P:HH.IMP
'You (high honorific) come along!'

In (1-3), the subject-verb agreement (in imperative form) *bɔl-ø*, *bɔl-a*, and *bɔl-ɔk* reflects the speaker's attitude as NH, H, and HH towards the 2P, respectively. Again, this is the same regardless of whether the pronoun is singular or plural. Likewise, declarative sentences in Assamese show person and honorificity agreement, where *-i* is NH, *-a*: is H, and *-e* is HH, but no number agreement, shown in (4-6).

(4) tɔi/tɔ.hɔt bʰa:t kʰa-l-i
2P:NH.SG/2P:NH.PL rice eat-PST-2P:NH.AGR
'You (hon-honorific) ate rice.'

(5) tumi/tuma.lʊk bʰa:t kʰa-l-a
2P:H.SG/2P:H.PL rice eat-PST-2P:H.AGR
'You (honorific) ate rice.'

(6) apuni/apuna.lʊk bʰa:t kʰa-l-e
2P.HH.SG/2P:HH.PL rice eat-PST-2P:HH.AGR
'You (high honorific) ate rice.'

Thus, Assamese uses pronominal morphology (rather than verb morphology) to encode plurality. But the three tiers of honorificity are morphosyntactically reflected on the verb. In Assamese, subjects agree with the verb in person as well as honorificity only in the presence of 2P pronouns. Otherwise, in 1P and 3P, the subject and the verb only agree in person.

Assamese also distinguishes honorific levels explicitly in the pronominal system. Following Alok (2020), I assume three honor tiers: NH < H < HH. These tiers are apparent in 2P and 3P pronouns (1st person has no honor contrast). For example, the second-person pronouns are roughly: *tɔi* (2:NH), *tumi* (2:H), and *apuni* (2:HH). The plural form of each of them is formed by adding a suffix, but crucially different suffixes are used for NH vs. H/HH. I will discuss below that *-hɔt* suffixes to NH pronoun stems and *-lʊk* to H/HH stems.

In summary, Assamese (i) is SOV (with fluid word order), (ii) has no number agreement, but has person agreement on verbs, and (iii) has an elaborated honorific system in personal pronouns. This serves as the setting for examining how honorific and number features interact.

1.2 Assamese Pronominal Paradigm

This section discusses the relevant pronouns in Assamese, glossed as *2P* for 2nd person, *3P* for 3rd person, *PL* for plural, *SG* for singular, *NH* for non-honorific, *H* for honorific, and *HH* for high honorific in Table 1.

Person	Honorificity tiers	SG form (IPA)	Gloss	PL form (IPA)	Gloss
2P	Non-Honorific (NH)	tɔi	2P:NH.SG	tɔ.hɔt	2P:NH.PL
	Honorific (H)	tumi	2P:H.SG	tuma.lʊk	2P:H.PL
	High honorific (HH)	apuni	2P:HH.SG	apuna.lʊk	2P:HH.PL
3P	Non-Honorific (NH)	ɦi/tai (MASC/FEM)	3P:NH.SG	ɦi-hɔt	3P:NH-PL
	Honorific (H)	teu	3P:H.SG	teu-lʊk	3P:H-PL
	High honorific (HH)	tekʰet	3P:HH.SG	tekʰet-hɔkɔl	3P:HH-PL

Table 1: Assamese pronouns

Table 1 shows, for instance, that the 2P non-honorific singular is *tɔi*, which becomes *tɔhɔt* in the plural form (suffix *-hɔt*). The honorific singular form *tumi* (2P:H.SG) pluralizes to *tumalʊk* (*-lʊk* suffix). The singular high honorific *apuni* (2P:HH.SG) changes to *apunaluk* (with, again *-lʊk* suffix as the plural suffix). Note the suffix *-hɔt* vs. *-lʊk* distinction: *-hɔt* is used with NH stems, whereas *-lʊk* appears on H/HH stems. The 3rd person follows similar patterns. *ɦi/tai* (3P:NH:MSC.SG/3P:NH:FEM.SG) → *ɦi-hɔt* (3P.NH-PL) (note that gender distinction is only in 3P non-honorific singular form), *teu* (3P.H.SG) → *teu-lʊk* (3P:H-PL), *tekhet* (3P:HH.SG) → *tekʰet hɔkɔl*.[1]

Thus, Assamese allows certain combinations of honorific roots and plural suffixes but not others. In particular, pluralization does not change the honorific level. For example, *tɔi* (2P:NH.SG) pluralizes as *tɔhɔt* (2P:NH.PL) without adding honorific distinction. The form **tɔlʊk* (the would-be plural if *-lʊk* could attach to non-honorific *tɔi*) is ungrammatical. Conversely, adding an NH plural suffix to an honorific stem is illicit. Native judgments confirm these patterns, as shown in Table 2 for 2P.

Consider the honorificity statuses of the pronominal stem and the plural suffix in columns (ii) and (iv), respectively, in Table 2. The honorificity mismatch between the pronoun form and the plural suffix results in the ungrammaticality of the plural forms of pronouns. The same is true for 3P pronouns. These illustrate that *plural markers do not introduce honorificity status*. Instead, the honorific level of the singular root persists in the plural, and the plural markers themselves carry honorificity status: *-hɔt* (NH.PL), and *-lʊk* (H/HH.PL). I will analyze this asymmetry in §4.

[1] *hɔkɔl* in 3P high honorific plural form is not a plural marker, rather a quantifier construction like 'they all' in English. This is only true for 3P high honorific pronouns. This comes from Assamese NPs like *hɔkɔl-(u)e hukhi (hoi)* 'everyone/all-EMP happy (exist)'. In brief, the 3P:HH pronoun does not follow the paradigm of *typical* plural formation of pronouns in Assamese; hence, I exclude *hɔkɔl* from this analysis. Note that 3P:HH pronoun tekhet 'they' is singular. To refer to a 3P.HH plural subject, the speaker must use *tekhet-hɔkɔl* . (EMP= emphatic)

(i)	(ii)	(iii)	(iv)	(v)	(vi)
SG stem form	Stem honorificity status	Plural suffix form	Suffix honorificity status	SG → PL	Matching honorificity (ii & iv)
tɔi	NH	*-hɔt*	NH	*tɔi* → *tɔhɔt*	Yes
tɔi	**NH**	*-lʊk*	**H**	*tɔi* → ****tɔlʊk***	No
tumi	H	*-lʊk*	H	*tumi* → *tumalʊk*	Yes
tumi	**H**	*-hɔt*	**NH**	*tumi* → ****tumahɔt***	No
apuni	HH	*-lʊk*	HH	*apuni* → *apunalʊk*	Yes
apuni	**HH**	*-hɔt*	**NH**	*apuni* → ****apunahɔt***	No

Table 2: Mismatch of honorificity between pronoun root and plural suffix

Additionally, Assamese verbs in declarative sentences show honorificity agreement with the subject in three tiers. But this subject-verb agreement in honorificity is restricted only to 2P subjects (see 4-6). Examples (7-8) show that 3P subjects with NH, H, and HH levels of honorificity are not sensitive to subject-verb honorificity agreement.

(7) ɦi/tai/ɦi-hɔt — bʰa:t — kʰa-l-e
3P:NH:MASC.SG/3P:NH:FEM.SG/3P:NH-PL — rice — eat-PST-3P.AGR
'He/she/they (non-honorific) ate rice.'

(8) teu/teu-lʊk — bʰa:t — kʰa-l-e
3P:H.SG/3P:H-PL — rice — eat-PST-3P.AGR
'He/she/they (honorific) ate rice.'

(9) tekhet/tekʰet-hɔkɔl — bʰa:t — kʰa-l-e
3P:HH.SG/3P:HH-PL — rice — eat-PST-3P.AGR
'He/she/they (high honorific) ate rice.'

In the third person, the verbal suffix *-e* does not change: it appears with both singular and plural subjects, and it remains the same no matter what the honorific level of the subject is (7–9). What stands out is that exactly this morpheme also shows up with second-person subjects when they are high-honorific. This is illustrated in (10), a repetition of (6).

(10) apuni/apuna.lʊk — bʰa:t — kʰa:-l-e
2P:HH/**2P**:HH.PL — rice — eat-PST-**3P**:HH.AGR
'You (high honorific) ate rice.'

At first glance, the 3P agreement morpheme *-e* appears to be an agreement mismatch, since a 2P.HH subject co-occurs. A closer look at the imperative paradigm (1-3) points in a different direction. In these forms, the verb shows regular agreement with the subject in both person and honorific status. The dedicated 2P.HH imperative marker *-ɔk*, for instance, attaches only when the subject is second person high-honorific, and it never appears with third-person subjects. If the form in (10) really involved a mismatch between subject and verb, we would expect a similar anomaly to surface in the imperative domain. No such

anomaly is found. Instead, the facts indicate a case of syncretism: the morpheme *-e* expresses 3P agreement, and it is also used to index 2P high honorific subjects.

2 Theoretical Background

I situate the analysis within the Minimalist Program (Chomsky 1995), following later developments in feature-checking theory (Chomsky 2000; 2001). On this view, lexical and functional heads carry bundles of features. Not all features come fully specified; some features are unvalued. The mechanism that matches these valued and unvalued features is Agree: a probing head searches its c-command domain for a suitable goal and, once a match is found, any unvalued features on the probe are specified with values from that goal. Formally (Adger, 2003: 169), in any configuration

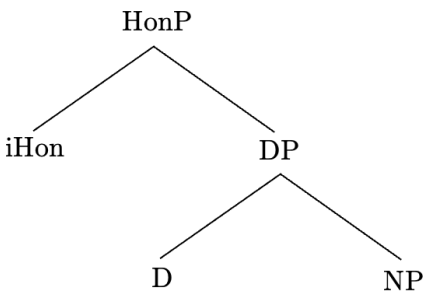
(11) X[F:**val**] … Y[*u*F: __]

> Where, if "…" represents c-command, then valued interpretable feature F with the value "val" checks and values the unvalued uninterpretable feature *u*F with "val", resulting in:
>
> X[F: **val**] … Y[*u*F: **val**]

That is, where X c-commands Y, the valued probe feature F on X can check against the unvalued feature *u*F on Y, resulting in Y[*u*F:val]. Unvalued uninterpretable features [*u*F] must be checked and deleted by this mechanism (Chomsky 1995; Adger 2003) for a successful derivation in the syntax.

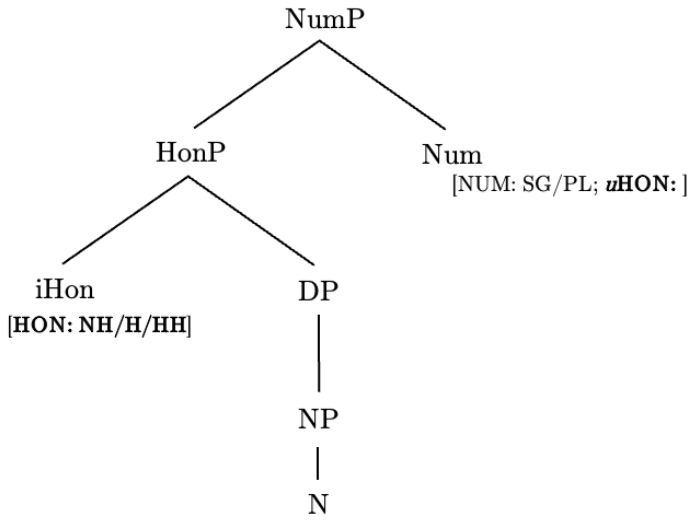
Following much recent work, I posit dedicated projections for honorificity and number features. Building on Alok (2020), I posit a HonP above the DP (Abney, 1987) for Assamese that hosts semantic feature [iHON] on a functional head: it takes a DP as argument and yields a relation indicating whether the speaker stands above or below the addressee in honorificity/social status. This is schematically represented in (12).

(12)



Also, following Ritter (1991, 1992), I assume a projection NumP above the DP that hosts plural number features for Assamese in (13).

(13)



In syntax, this is implemented by a Hon head with features [HON: NH/H/HH] (where either NH, H, or HH are its values or "val"). I adopt a similar approach for Assamese. The valued interpretable [HON: NH/H/HH] feature probes a corresponding unvalued uninterpretable [*u*HON:__] on another element, which, in the case of Assamese, is on the Num head. Here, the Num head carries two features- a valued interpretable [NUM] feature with its values either as [NUM: SG] or [NUM: PL], and an unvalued uninterpretable [*u*HON:__]. In the structure I present here, the [HON] feature on Hon head engages in an Agree relationship with the [NUM] feature on the Num head that bears [NUM: SG/PL; *u*HON: __].

When we extend the schema in (11) to Assamese, the Hon head already bears a valued [HON] feature. This value is then transferred to the Num head through Agree, so that the unvalued [*u*HON] feature on Num is properly specified. In effect, the honorific property of the root is first fixed in Hon and subsequently passed upward to Num, ensuring that the plural form receives the correct honorific interpretation. Thus, adding plural does not automatically change the [HON] value; it simply changes to [NUM:PL] from [NUM: SG].

In sum, the analysis here uses (i) a HonP above DP to introduce the interpretable honorific feature, (ii) a NumP (or Num head) above DP (subsequently above HonP) that has [NUM: SG/PL; *u*HON:__], and (iii) feature-checking via Agree (Chomsky 1995, 2000, 2001; Adger 2003) to value [*u*HON: __] on the Num head. This setup explicitly treats honorificity and plurality as separate syntactic elements.

3 Comparative Evidence from Magahi and Hindi-Urdu

In Magahi, honorificity is encoded in two domains- (pro)nominal and verbal. Honorificity is seen in two types within the verbal domain, involving verbal agreement- subject honorificity and addressee honorificity. Subject honorificity is when the subject agrees in

honorificity with the verb. The subject agrees with the verb in person and honorificity. In Magahi, addressee honorificity is a social relationship between speaker and hearer that is optionally encoded via a morphosyntactic presence on the verbal stem, even when the hearer (or addressee) is not the subject of the sentence. This is called allocutive agreement (Miyagawa 2012, 2017; Miyagawa & Saito 2008; Oyharçabal 1993; Portner *et al*. 2019) or *addressee agreement* (Verma 1991). In short, finite verbs encode both person and honorificity features of the subject, and optionally the honorificity status of the addressee (Alok 2020, 2021). Building on this observation, Alok (2021) proposes a formal account for allocutive agreement in which the clause contains a covert second-person addressee, written as *Hr*, in a designated Hearer position located at "a specifier of FinP". The Finite head, written as *Fin*, bears an uninterpretable honorific feature [*u*HON] (*i.e.,* ε). There is a corresponding interpretable honorificity feature on Hr as [iHON], which is a semantic honorificity feature on every DP that represents the addressee's honorific status. In this system, the uninterpretable [*u*HON] on Fin probes the interpretable [iHON] feature on Hr. Fin[*u*HON] Agrees with Hr[iHON] in every finite clause, thereby matching honorificity features on both heads- Fin and Hr. This explains allocutive agreement in Magahi. Note that in the syntactic processes involved here, the number feature [NUM] does not play any role.

In the same proposal, to explain the subject honorificity, Alok proposes a system where the subject DP carries an interpretable honorificity feature [iHON]. Often, this feature is a result of T/V distinction in pronouns, projected as [iHON] in the syntax on DPs. The tense head T carries a corresponding uninterpretable honorificity feature [*u*HON]. When the uninterpretable feature [*u*HON] on T Agrees with interpretable [iHON] feature on DP, the system attains at a successful derivation, which results in subject and verb agreeing in honorificity. In the same way, subject and verb agree in person where T carries [*u*P(erson)] and the DP a [iP] features. Again, note that the number feature does not play any syntactic role here in the honorificity process. Magahi therefore provides evidence that both subject and addressee honorificity are encoded in the syntax, and that the [HON] feature is independent of number. This supports the view that honorificity is a syntactic feature distinct from the canonical Phi (φ)-features (person, number, gender). Although Assamese lacks allocutive agreement like Magahi, Alok's HonP analysis can be extended to the subject honorificity in the case of 2P (as seen in 1-6) as well as the pronominal domain as: the [HON] feature of the pronoun is hosted on a Hon head and is checked via an Agree relation, in a parallel manner to clausal domain in Magahi.

Hindi-Urdu shows a point of contrast. Here, the suffix *-jī* is attached to nouns as a marker of respect. Once it is present, the verb no longer agrees as singular but instead takes plural marking, even when the noun clearly refers to just one person (Bhatt & Davis 2023) (see 14-16).

(14) *mina lambi hai*
Mi:na lambi hɛ
Mi:na.F tall.F be.PRS.3.SG

‘Mina is tall.’ (Bhatt & Davis, 2023: 22)

(15) *minaji lambi hain*

Mi:na-ji	lambi	hɛ̃
Mi:na.F-**HON**	tall.F	be.PRS.3.**PL**

‘Mina is tall.’ (Bhatt & Davis, 2023: 23)

(16) ***minaji lambi hai***

Mi:na-ji	lambi	hɛ
Mi:na.F-**HON**	tall.F	be.PRS.3.**SG**

‘Mina is tall.’ (Bhatt & Davis, 2023: 23)

The presence of *-jī* changes how the verb agrees. When the plain name *Mīnā* is used, the verb shows singular agreement. Adding the respectful suffix *-jī*, however, alters the agreement pattern, and the verb surfaces in the plural even though the referent is still one individual. To account for this, Bhatt & Davis (2023) interpret *-jī* as a Hon head inside the nominal phrase that carries a formal [+PL] feature. This feature is what causes the verb to appear in the plural, but its meaning contribution is one of respect rather than actual plurality. Bhatt & Davis interpret the same formal plural feature in Hindi-Urdu in two ways: a genuine plurality, or as a politeness marker. They further propose that due to the presence of Hon, the interpretation of genuine plural is blocked and vice versa. This keeps the nominal component of Hon morpho-syntactically singular (which is often the case in oblique form.)

Bhatt & Davis describe a system where honorificity and number are tightly linked, so that the two features cannot easily be teased apart. The Assamese pattern does not work this way. When a plural marker is added, it simply contributes number and leaves the honorific value of the root unchanged. A non-honorific stem, for instance, only combines with a non-honorific plural suffix, and the result is still interpreted as non-honorific. For example, singular to plural in (17) does not change the honorificity status.

(17)

toi	→	*tohot*
tɔi		tɔ.hɔt
2P:**NH**.SG		2P:**NH**.PL
‘You’		‘You’

This suggests that in Assamese, [HON] and [NUM] are independent features, as suggested in Alok (2020, 2021) for Magahi, contra [HON] tied to [NUM] in Hindi-Urdu (Bhatt & Davis 2023). Assamese clearly has distinct plural suffixes depending on honorificity level. Nevertheless, the insight that a Hon element can bear a formal plural feature may be useful, but not in the analysis I present here. I view Assamese plural suffixes as carrying number features, and also an honor feature that must match the root/stem’s.

In a large typological study covering 120 languages, Wang (2023) notes that honorific systems frequently draw on categories that are semantically unmarked, *e.g.,* plural number, third person, or indefiniteness. One familiar illustration is the use of a plural pronoun for a single polite addressee, as seen in French *vous* or Spanish *ustedes*. Wang takes this to be a pragmatic strategy based on feature markedness, rather than evidence for a distinct [HON] feature in the grammar.

Assamese seems to partly fit this pattern. *i.e.*, second-person high-honorific subjects trigger the use of the third-person agreement marker *-e* on the verb, a clear example of an unmarked category being pressed into the service of politeness. At the same time, Assamese does not reduce honorificity to such strategies. The pronominal system demonstrates that honorificity and number are kept apart: forms like *təhɔt* and *tɷmalʊk* are used in polite reference, and while the suffixes they carry unambiguously express plurality, the honorific value of the root remains intact.

Assamese, therefore, illustrates that plural morphology is recruited for plural functions only. From this perspective, the use of *-e* with 2P high-honorific subjects is best analyzed as syncretism with 3P agreement morpheme, rather than as true mismatch. At the syntactic level, the distinction among honorific tiers in 2P subject-verb agreement is derived through the Agree mechanism, a point to which I return in the next section.

4 [HON] is independent of [NUM]

I now present my approach that honorificity and number features are distinct from one another in Assamese. Honorification, thereby honorificity, is a syntactic process, while number features (singular and plural) are not. The evidence comes from the mismatch patterns observed in pronouns and from subject-verb agreement in second person and honorificity.

4.1 Honorificity in pronominal plural marking

As noted above in §1.2, pluralizing a non-honorific (NH) pronoun preserves its NH status, and pluralizing an honorific pronoun preserves its honor. I restate the crucial cases from Table 2 below:

(18)	*təi* '2P:NH.SG' +*-hɔt* 'NH PL suffix'	→ *tɔ.hɔt* '2P:NH.PL'	(grammatical)
(19)	*təi* '2P:NH.SG' + *-lʊk* 'H PL suffix'	→ **tɔ.lʊk*	(ungrammatical)
(20)	*tumi* '2P:H.SG' + *-lʊk* 'H PL suffix'	→ *tuma.lʊk* '2P:H.PL'	(grammatical)
(21)	*tumi* '2P:H.SG' + *-hɔt* 'NH PL suffix'	→ **tuma.hɔt*	(ungrammatical)

In (18-21), adding *-hɔt* or *-lʊk* to the root never changes the root's honorificity status, nor does it change the honorificity status of the output. The ungrammatical examples (19 & 21) show that attempting to override the root's honorific level yields a bad form. They serve as the empirical foundation. Any analysis of these instances must ensure that plural morphology is licensed only when honor features match. Importantly, the form *təhɔt* is

grammatical despite *tɔi* being NH; if Assamese had required plural suffixes to add honorificity status, *tɔhɔt* would incorrectly receive H status. Instead, *-hɔt* here simply pluralizes without changing honorificity, given that the honorificity status of the stem matches with that of the plural morpheme. To explain this mechanism formally, I propose the following feature specifications for Assamese-

(i) The pronoun root (DPs) carries an honorificity feature projected as HonP above DP with an interpretable [HON:][2] feature, valued as NH, H, or HH *wrt* the speaker's social status to the referent of the DP.

(ii) The plural suffix, projected as NumP above HonP, carries two features: an unvalued uninterpretable [*u*HON] feature and a valued [NUM:] feature. The [NUM:] is inherently valued as SG or PL depending on the pronominal, and it is not involved in syntactic processes. The [*u*HON] feature on Num head must be valued and checked for a successful derivation.

Under this system, the projections of the syntactic elements in question are in c-command relationship. This is schematically represented in (22) and hierarchically in (23), where angular parentheses (>>) represent structural hierarchy.

(22) [NumP [Num] [HonP [Hon] [DP]]]

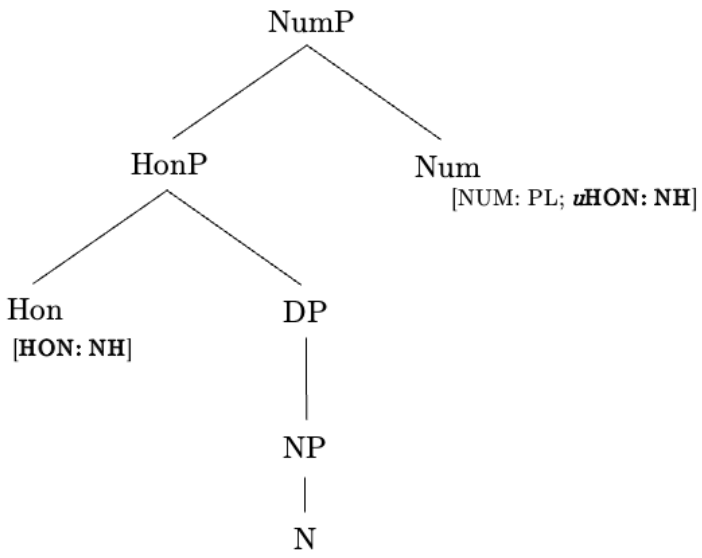
(23) NumP >> HonP >> DP

In Assamese pronominal forms, the plural morpheme comes last in the structure. What is interesting is that the exact shape of this plural marker depends on the honorific status of the base it attaches to. This ordering points to a hierarchy in which honorific information is settled closer to the noun root, and the projection of number takes place higher up. Following Ritter (1992), I assume that Number is represented as a functional head NumP dominating NP/DP. Since I take Abney's (1987) approach, on this view, HonP is structurally between NumP and DP, where HonP encodes the honorific distinctions (yielding *tɔi* vs. *tumi*), and NumP above introduces plural morphology (*e.g.*, *tɔi* vs *tɔhɔt*). Cross-linguistically, similar functional splits are attested. For instance, the honorific suffix *-jī* in Hindi-Urdu triggers plural agreement even on singular nominals, and Bhatt & Davis (2023) treat honorificity as occupying the Num slot. But Magahi explicitly encodes honorificity on an Hon head within the nominal spine (Alok 2021). These patterns accord with the general idea that morphemes closer to the root correspond to lower functional projections, which, in Assamese, is the fact that number suffixes are always outermost,

[2] In §2, I use "Hon" as the daughter branch of HonP instead of "iHon" as in (12) (where "i" is a mnemonic for *interpretable*). For uniformity, I make a distinction between features that are interpretable as X and uninterpretable as *u*X, where the italicized "*u*" in *u*X stands for *uninterpretable*, and where X is a variable for any syntactic element such as [HON].

requiring NumP to dominate HonP, and HonP in turn dominates the DP, validating the hierarchy presented in (23). This is structurally presented in (24) with feature specifications of Hon and Num.

(24)



The Num head with an uninterpretable unvalued [*u*HON:] feature probes for the goal Hon head with the interpretable valued [HON: NH/H/HH], thereby triggering Agree relationship. Since the [HON: NH/H/HH] feature is interpretable and valued, it can value the uninterpretable unvalued [*u*HON] feature on the Num head. For instance, if the DP pronoun root is NH, Hon carries [HON: NH] and it will probe and check [*u*HON:] on Num, resulting [*u*HON: NH]. The [NUM:] feature on Num is valued inherently by the plural meaning of the pronominal form, not via syntactic Agree. The outcome is that the plural form [NUM: PL; *u*HON: NH] is licensed, *i.e.*, *-hɔt* attaches to an NH root. Similarly, an H root leads to [*u*HON:H] on Num, licensing *-lʊ̀k*. The key point is that the [HON] feature is *checked* via Agree, unlike [NUM], so plurality is added independently of honor.

In minimalist syntax, uninterpretable features must be valued and checked against their interpretable counterparts in order for a derivation to converge (cf. Chomsky 1995, 2000, 2001). In the present case, the uninterpretable unvalued honorificity feature [*u*HON] on Num must be licensed before further derivation can proceed. This requirement is particularly visible when a pronominal root combines with a plural suffix that is sensitive to the honorific status of the pronominal root. Put differently, the Agree relation between Num and Hon is crucial only in plural forms; in singular forms, honorificity agreement appears redundant. To account for this, I consider two approaches.

The *first* approach treats the Num head as carrying two features: an obligatory [NUM] (valued as SG or PL), and an optional [*u*HON] (to be valued by its corresponding interpretable feature on Hon head). The [*u*HON] feature is triggered only when [NUM] is

plural (or [NUM: PL]); in singular contexts (when [NUM: SG]), it is absent, which explains why no separate honorific agreement appears on the singular pronoun. Earlier discussions of feature checking allow for a certain degree of optionality (Pesetsky & Torrego 2007). In contrast, I adopt the *second* approach, which assumes that Num obligatorily carries both [NUM] and [*u*HON], regardless of the particular number specification. Here, [*u*HON] is always valued by Agree with Hon, even when [NUM] is singular. On this view, singular pronouns lack overt morphology not because [*u*HON] is absent, but because the singular number feature has a null realization (cf. Harley & Ritter 2002 on null morphology in feature systems). This account is theoretically more economical: it does not require the grammar to condition the presence of [*u*HON] on the value of [NUM], thereby avoiding unnecessary stipulation. I propose a formal analysis in (25), following the steps of derivation below:

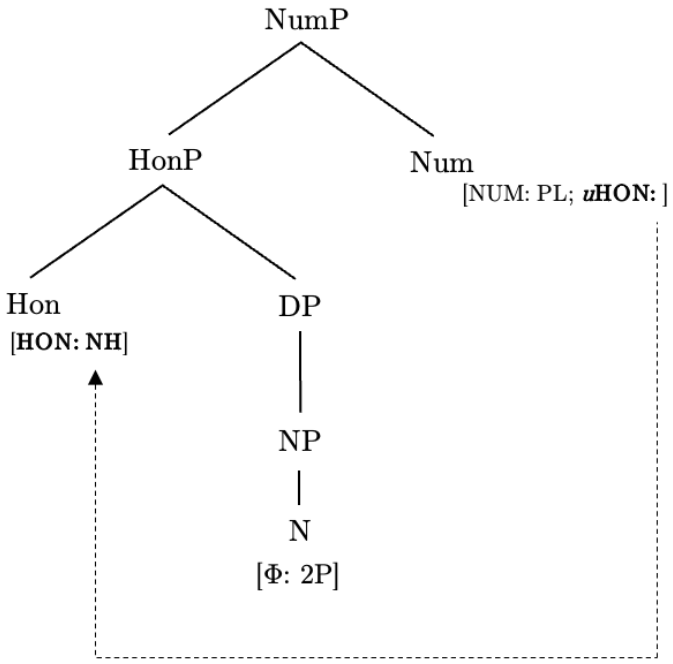
STEP 1: The derivation begins with a DP whose honorific specification is projected in HonP. For instance, a non-honorific (NH) root specifies its feature as [HON: NH] on Hon.

STEP 2: Above this, NumP is merged, introducing number features and an unvalued honorific feature: [NumP [Num $_{[NUM: ; uHON:]}$] [HonP [Hon $_{[HON: NH]}$] [DP]]]

STEP 3: Through Agree, the [HON: NH] feature on Hon values and deletes the uninterpretable [uHON] feature on Num.

STEP 4: The number feature on Num is valued independently. For instance, when Num is specified as [NUM: PL] by the semantics of plurality, the resulting feature bundle on Num is [NUM: PL; HON: NH]

(25)



This configuration yields the grammatical form *tɔ-hɔt*, *i.e.*, the non-honorific root *tɔi* combined with the non-honorific plural suffix *-hɔt*. By contrast, if the Num head hosting the plural suffix *-lʊk* (which carries [*u*HON: H]) were merged with an NH root (*tɔi*), the Agree relation would fail, since Hon [HON: NH] cannot value Num as [*u*HON: H]. This predicts the ungrammaticality of *tɔ-lʊk*. Assamese handles plural and honorific marking in a way that sets it apart from Hindi-Urdu. The two categories are realized independently.

Even though (25) abstracts away from some details, the core idea is that Num, bearing an unvalued [*u*HON] feature, probes downward to find a matching interpretable feature on Hon. When Hon is specified as [HON: NH], it values [*u*HON:] on Num accordingly. Number, by contrast, is introduced independently on Num and is valued by the semantics of plurality. In this view, honorificity is a syntactic feature that participates in Agree, whereas number is a feature whose valuation is determined compositionally by the semantics of pluralization. This analysis is consistent with Adger's (2003) characterization of Agree as feature valuation under c-command, and with Alok's (2020) proposal for Magahi, where honorificity is checked by a functional head in the clausal domain.

4.2 Honorificity in subject-verb agreement

As noted in §1.2. Assamese finite verbs show a three-way person/honorific agreement paradigm only for second-person subjects. In particular, 2P non-honorific (NH) subjects (*e.g.*, *toi/tohot*) trigger the suffix *-i* on the verb, 2P honorific (H) subjects (*tumi/tumaluk*) trigger *-a*, and 2P high-honorific (HH) subjects (*apuni/apunaluk*) trigger *-e*. Notably, these

markers encode person plus honorificity, not number: the same verbal morphology is used for singular vs. plural (no number inflection). Thus, verbs agree with the subject in [Person: 2] and [HON: NH/H/HH] with the subject, independent of any number feature. I repeat (4-6) below as (26-28).

(26) tɔi/tɔ.hot b^ha:t k^ha-l-i
2P:NH.SG/2P:NH.PL rice eat-PST-2P:NH.AGR
'You (hon-honorific) ate rice.'

(27) tumi/tuma.lʊk b^ha:t k^ha:-l-a
2P:H.SG/2P:H.PL rice eat-PST-2P:H.AGR
'You (honorific) ate rice.'

(28) apuni/apuna.lʊk b^ha:t k^ha-l-e
2P:HH.SG/2P:HH.PL rice eat-PST-2P:HH.AGR
'You (high honorific) ate rice.'

In (26), the subject is the 2P non-honorific pronoun *tɔi*. Here, a morphological agreement is inflected on the verb stem: the *-i* morpheme encodes both 2P and non-honorific status of the subject. In the same way, in (27), the subject is the 2P honorific pronoun, and the verb agrees in person and honorificity using the morpheme *-a*, used for 2P honorific subjects. Similarly, for 2P high honorific subject in (28), the verbal agreement is *-e*. This distinction in parallely maintained in imperative forms, as shown in (1-3). I summarize these morphemes in Table 3.

Subject		Verbal agreement	
2P form	Honorificity		
tɔi	NH	∅ (null)	-i
tumi	H	-a	-a
apuni	HH	-ɔk	-e

Table 3: Agreement morphemes on the verb based on the honorificity of the subject

Under the Minimalist framework, we assume the finite T head carries unvalued φ-features [*u*Person:] (cf. also Chomsky 2000, 2001; Adger 2003) and [*u*HON:] (Alok 2021) that must be valued by the subject. Each second-person pronoun (DP) is taken to project an HonP whose head Hon bears an interpretable [HON: NH/H/HH]. In syntax, T looks for the interpretable [HON: NH/H/HH] of the subject to value its [*u*HON:] (and [*u*Person]) via Agree. This mechanism is parallel to what Alok (2021) proposes for Magahi, where subject honorification and addressee agreement are fused and valued by Agree between functional heads. Because [HON: NH/H/HH] on Hon is interpretable and valued, it licenses [*u*HON] feature on T. The result is that T acquires [*u*Person: 2, HON: NH/H/HH] matching the person and honorific tier of the subject. Here, no [NUM] feature is involved in the verb, consistent with the empirical fact that number is not marked on the verb in Assamese.

5 Implications

The discussion I carry in this paper has implications for both typology and theory. It shows that honorificity in Assamese is not derived from number but instead has a dedicated functional head. In Assamese, the treatment of plural and honorific categories does not follow the pattern familiar from Hindi-Urdu. The two are expressed through distinct means rather than being fused into a single form. This separation adds to the understanding of the range of agreement strategies in the Indo-Aryan family. Once again, this distribution indicates that the language treats NumP and HonP as separate projections. Looking beyond Assamese, this distinction is typologically significant because it avoids a politeness strategy that Wang (2023) shows to be common cross-linguistically, namely the reanalysis of plural morphology as a marker of respect.

Seen in this view, Assamese provides evidence for a system in which [HON] operates alongside [NUM] rather than being dependent on it. Honorific distinctions are added to, rather than substituted for, number. From a theoretical standpoint, this supports the idea that the Agree mechanism can apply to features that encode social meaning. In turn, this strengthens the claim, developed by Alok (2020) and others, that honorificity belongs to the syntactic feature set itself rather than being reducible to pragmatic inference.

Importantly, we see that checking [HON] requires no special mechanism beyond Agree: it is a syntactic operation that requires an Agree mechanism; the number feature does not. This asymmetry- that [HON] participates in syntactic feature checking while [NUM] is not- is reminiscent of how person/gender and number sometimes behave differently in agreement (*e.g.*, Hindi-Urdu lacks explicit plural morphology on verbs for politeness; here Assamese does not use number but still checks honorificity). Finally, the Assamese pattern may generalize to other languages. The findings in this paper raise a possibility that [HON]-[NUM] independence holds more widely than hitherto assumed.

6 Conclusion

In this paper, I examine the morphosyntactic encoding of honorificity and number in Assamese, with particular attention to the interaction between pronominal forms and verbal agreement. I show that Assamese organizes honorificity in a three-tier system (non-honorific, honorific, and high honorific), which is expressed in the pronominal paradigm. Number distinctions, by contrast, are encoded in the pronoun system but are not consistently transmitted to the verbal domain.

Assamese finite verbs show honorific agreement only with second-person subjects, where the suffixes *-i*, *-a*, and *-e* directly encode the relative honorific status of the subject. These markers are insensitive to number, which supports the claim that honorificity and number are encoded in distinct functional projections. In analyzing the Assamese facts, I

adopt a Minimalist approach where the head of HonP carries interpretable [HON] features. This feature values the corresponding [*u*HON] on T through Agree (cf. also Chomsky 2000, 2001; Adger 2003; Pesetsky & Torrego 2007). A similar mechanism is argued for in Magahi by Alok (2021), but the Assamese pattern is narrower: only second-person subjects trigger honorific agreement on the verb. At the same time, number does not enter into the agreement relation. Although number morphology is available in the pronominal system, it does not extend to the verbal domain, which suggests that [NUM] projects independently of verbal agreement.

When seen in a broader Indo-Aryan context, Assamese therefore occupies an intermediate position. In Hindi, the highest honorific forms are realized through verbal patterns that overlap with plural agreement, so that respect is formally expressed as number. Magahi shows a different alignment, where the honorific status of the subject is tied together with allocutive agreement. Assamese, however, keeps the two categories apart. Verbal suffixes in the second person reflect only honorific distinctions, while number is marked solely in the pronominal system and never carried over to the verb. Assamese instead encodes a more restricted but structurally transparent system- honorificity is grammatically active only in the second person, and its effect is morphologically confined to verbal suffixes, while number remains pronominal.

By setting apart the role of [HON] and [NUM] features in Assamese, I argue for treating honorificity as an independent grammatical dimension, rather than a derivative of person or number.

Acknowledgements

I would like to thank my supervisor, Deepak Alok, for many helpful discussions.

References

Abney, Steven Paul. 1987. *The English noun phrase in its sentential aspect*. Doctoral dissertation, Massachusetts Institute of Technology.

Adger, David. 2003. *Core syntax: A minimalist approach*. Oxford: Oxford University Press.

Alok, Deepak. 2020. *Speaker and addressee in natural language: Honorificity, indexicality and their interaction in Magahi*. Doctoral dissertation, Rutgers University.

Alok, Deepak. 2021. The Morphosyntax of Magahi Addressee Agreement. *Syntax* 24(3). 263–296.

Bhatt, Rajesh and Christopher Davis. 2023. Number, honor, and agreement in Hindi-Urdu. In *Proceedings of Formal Approaches to South Asian Languages 12.*

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels, and Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language* 1-52. Cambridge, MA: MIT Press.

Harley, Heidi, and Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78(3). 482–526.

Kaur, Gurmeet, and Akitaka Yamada. 2022. Honorific (mis)matches in allocutive languages with a focus on Japanese. *Glossa: A Journal of General Linguistics* 7(1). 1–29.

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

Miyagawa, Shigeru. 2012. Agreements that occur mainly in the main clause. In Liliane Aelbrecht, Liliane Haegeman, and Rachel Nye (eds.), *Main clause phenomena: New horizons*, 79–112. Amsterdam: John Benjamins.

Miyagawa, Shigeru. 2017. *Agreement beyond phi*. Cambridge, MA: MIT Press.

Miyagawa, Shigeru & Mamoru Saito (eds.). 2008. *The Oxford handbook of Japanese linguistics*. Oxford: Oxford University Press.

Niinuma, Fumikazu. 2003. The syntax of honorification. Doctoral dissertation, University of Connecticut.

Oyharçabal, Beñat. 1993. Verb agreement with non-arguments: On allocutive agreement. In José Ignacio Hualde and Jon Ortiz de Urbina (eds.), *Generative studies in Basque linguistics*, 89–114. Amsterdam: John Benjamins.

Palash, Nilotpal, Sharma, Dipankar & Bipul Dutta. 2019. *Assamese grammar*. Guwahati: Assam Publication Board.

Pesetsky, David, and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In Simin Karimi, Vida Samiian, and Wendy Wilkins (eds.), *Phrasal and clausal architecture: Syntactic derivation and interpretation*, 262–294. Amsterdam: John Benjamins.

Portner, Paul, Miok Pak & Raffaella Zanuttini. 2019. The speaker-addressee relation at the syntax-semantics interface. *Language* 95(1). 1–30.

Ritter, Elizabeth. 1991. Two functional categories in noun phrases: Evidence from Modern Hebrew. In Susan Rothstein (ed.), *Syntax and semantics: Vol. 25. Perspectives on phrase structure*, 37–62. San Diego, CA: Academic Press.

Ritter, Elizabeth. 1992. Cross-linguistic evidence for number phrase. *Canadian Journal of Linguistics* 37(2). 197–218.

Verma, Mandira K. 1991. Exploring the parameters of agreement: The case of Magahi. *Language Sciences* 13(2).125–143.

Wang, Ruoan. 2023. Honorifics without [hon]. *Natural Language & Linguistic Theory* 41 (3). 1287–1347.

Towards a typology of information structuring in Munda languages

Gregory D. S. Anderson, *Living Tongues Institute for Endangered Languages*
Bikram Jora, *Living Tongues Institute for Endangered Languages*
Moses Oyeleye, *Living Tongues Institute for Endangered Languages*
Izza Armand, *Living Tongues Institute for Endangered Languages*
Tara Shabazaz, *Living Tongues Institute for Endangered Languages*
Luke Horo, *Living Tongues Institute for Endangered Languages*

Abstract

Munda languages use morpholexical, syntactic and prosodic means to encode information structure of various types. These parameters are independent but may be interdependent. Based on two distantly related languages of the family, Sora of the Sora-Gorum branch and Mundari of the Kherwarian branch we provide details on each of these three means of encoding information structure. There is some evidence that two of the morpholexical means that are productive in Mundari and restricted in Sora may be old formants encoding information structure in the family. Prosodic means appear to show no correlation across the languages and syntactic means of encoding information structure show some commonalities that could be retentions of an earlier proto-language system, while others are clearly language-specific.

1 Introduction

In this paper we discuss formal means of encoding information structure in two Munda languages, Sora of the Sora-Gorum subgroup and Mundari of the Kherwarian subgroup. We discuss three logically independent but often interconnected means of highlighting discursively and pragmatically important arguments, predicates and adjuncts in these languages, viz., morpholexical indices §2, syntactic configurations §3 and prosodic means §4 to track or highlight elements that are topical(ized) or in focus.

Under morpholexical means of encoding focus or topic, we include clitics, affixes or prosodically full forms of functional words. Under syntactic configurational means of encoding information structure, we include scrambled syntax or the pragmatically and information structurally activated orders of elements. Cleft-structures seem to be an alien strategy to Munda, as no such structures occur in our collection of spontaneous discourse texts. Under prosodic means we cover the differential use of the acoustic cues of prominence of duration, intensity or pitch/f0 to encode not lexical or grammatical or word-level or phrasal or utterance level contrasts but rather discourse/pragmatic and information structural ones. This is the first ever analysis of the different formal means of encoding information structure in the Munda language family.

In descriptions of Munda languages, the concepts of topic and focus get confused or conflated and the morphological elements involved may indeed straddle both functions.

This functional overlap is particularly true in contrastive contexts where contrastive topic and contrastive focus may converge. To be sure, even in non-contrastive contexts, both topic and focus reference arguments and events in a discourse space larger than a sentence, shifting attention and noteworthiness or referentiality of such referents and events (and indeed adverbial adjuncts) as suits the narrative or conversational structure. Of course, topic is an accepted node or functional projection in Chomskyan syntax and has a formal role in Lexical Functional Grammar in the functional domain. Nevertheless, some scholars dispute topic as a singular concept but rather view it as a cluster of prototypes (Asher, 2004; Jacobs, 2001).

In traditional approaches, topic should represent given information and focus conversely should be new information on some meta-discursive level. But topicalization can reintroduce a referent for example that may have dropped from the immediate attention of the hearer. In many formal syntactic approaches, and indeed supported by cross-linguistic evidence, topics may be external to the rest of the sentence and occupy a separate field in the syntax often on the left-periphery. But focalized elements can also appear in such a position, which further contributes to the potential confusion between these two information structuring concepts. Munda languages support this generalization as exemplified in relevant places below.

Topic is widely used in linguistic analysis even while defining it is elusive (Asher, 2004), and indeed while the concepts of contrastive topic, discourse topic and sentential topic are interconnected, they formally have little in common other than a vague sense of 'aboutness', and being 'known' information.

With respect to focus some standardly followed or accepted definitions in linguistics include those of Lambrecht (1994) and Dik (1997). Lambrecht defines focus as "the semantic component of a pragmatically structured proposition whereby the assertion differs from the presupposition. The focus component is by definition an unpredictable part of the proposition." Dik (1997) for his part uses the following definition "[t]he focal information in a linguistic expression is that information which is relatively the most important or salient in the given communicative setting and considered by S[peaker] to be the most essential for A[ddressee] to integrate into his pragmatic information."

Many of the other nuanced takes on this complex and somewhat elusive phenomenon of 'focus' include those found in such a diverse array of studies as works by Bearth (1999), Büring (1997, 2010), Caron (2000), Chafe (1976), Dik (1980, 1981), Dik et al. (1981), Drubig & Schaffar (2001), Fiedler et al. (2010), Gundel & Fretheim (2004), Krifka (2004, 2007), Sasse (1987), Uhmann (1991), van der Wal (2016), Zimmermann (2008, 2016), Zimmermann & Onea (2011) and Zubizarreta (1998), all of which we consulted in an attempt to determine the functional domains of the morpholexical elements used in Sora and Mundari that structure information in spontaneous narrative and conversational discourse.

Like topic, there are several different meanings of focus as used in the literature, whether the communicative point in is informational, presentational, contrastive, corrective or counter-presuppositional. The scope of the focalization phenomena also varies, including whether the focus is on an argument or is predicate/event-centered, or whether the focalizing is ex-

haustive, restrictive, additive, emphatic, verum-based, etc., semantically.

2 Morpholexical means of encoding information structure in Munda

Use of morpholexical means in flagging focus and topic is the most common strategy found in Munda languages. This includes uninflected particles, affixes and/or clitics used to encode information structuring semantics.

2.1 Morpholexical (-cum-syntactic) Means of Encoding Focus in Sora

Contrastive topic in Sora is marked by the postposing of the particle *ati*/*ate* or the clitic/affix *-ti*/*te* after the topicalized element. It sometimes occurs left-dislocated in a quasi-cleft-like field on the left periphery of the clause.

(1) a. ao kən a-kəɽibab-ən ate titti-n sed-ir-ai-dʒi
DISC this DEP-skull-N.SFX TOP over.there-N.SFX throw-go-PFV-3PL
'so as for that skull of his, they took and threw it away over there'

b. bəru-n a-dʒaʔmol ati daku biɲdo
mountain-N.SFX DEP-seed TOP COP but
saroba-n a-dʒaʔmol ati ɲen arika
field-N.SFX DEP-seed TOP I NEG.COP
'I have some mountain seeds, but as for field seeds, I don't have any'

The most common focalizing element in Sora is *=na*. It is quite promiscuous in the host it selects for (2a-2f). It commonly is translated as 'really, surely, indeed'.

(2) a. Verb
a-pəsiʔ ati anin kəjiʔ-te=na kəjiʔ-ti
DEP-child TOP 3PRON die-NPST=FOC die-NPST
'as for the child, he will die, he will definitely die'
(note also the left dislocated TOP–marked NP)

b. Object NP
kun a-kansim-ən=na kudub-ən tubob-tə-dʒi
that DEP-rooster-N.SFX=FOC all-N.SFX kill(head)-NPST-PL
'they kill all the roosters (that were brought)'

c. Manner Adverb
ɲen enegoi=na doʔoŋ-bin sum-le ambin dʒanaŋ ɲen
1SG like.this=FOC body-2PL catch-CV 2PL also 1SG
kabɲid-ta-bin
kill-NPST-2PL.UND
'like this I will catch you, and I will kill you also'

d. Postposition

kon baṯ̪i=na anin-dʒi a-rəb-rəb-ḏ̪om-li-n-dʒi anin-dʒi
that with=FOC 3PRON-PL DEP- protect-RFLXV-ITR-N.SFX-PL 3PRON-PL

sənu-ba-n dʒanaŋ jir-rə-dʒi
fight-DIR-N.SFX also go-PST-PL

'they would go also to fight (battles/wars) for their protection with these'

e. Temporal adverb/adjunct

adnəŋ=na kəm pasidʒ-an kəna-n paŋ-ṯ̪e
at.that.time=FOC this child-N.SFX tiger-N.SFX take-NPST

a-monsil-ɲen ɠam-le
DEP-niece-1SG say-PST

'they would go also to fight (battles/wars) for their protection with these'

f. Subject/Agent

o-ɲen=na ajer-təm ɠam-li
child-1SG.POSS=FOC return-PRF say-PST

'my child really has returned, she said'

With locative adverbs/adjuncts, the focalizer =*na* has an exhaustive/restrictive focal interpretation in Sora.

(3) a. tidtin=na
there:N.SFX=FOC
'only there'

b. iʔdʒa ṯ̪iʔnen=na
no here:N.SFX=FOC
'no, only here (her uncle said)'

2.2 Morpholexical Means of Encoding Focus in Mundari

There are at least two affixes/clitics - that function in the information structure system in Kherwarian languages: -*do* and –ɠ*e*. In Mundari -*do* appears to be a topicalizer (4) mainly while -ɠ*e*/-ɠ*i* functions in a manner similar to Sora =*na*, with similar promiscuity in host selection (5a-5d).

(4) Mundari *do*

enlo=do miad hoɽo hente-te lolo lidʒa
that.moment=TOP one man there-ABL hot cloth

tusiŋ-kete oḏ̪oŋ-dʒa-n-a
wear-CV come.out-TAM-ITR/MDL-IND

'at that very moment one man came out from there wearing a hot cloth'

(5) a. Subject

aɲ-ge=ɲ senog-a
1SG-FOC=1SG.SUBJ go:IPFV-IND

'I will go' (Osada, 1992, p. 136)

b. Object

manɖi-ge=ɲ dʒom-ke-n-a
food-FOC=1SG.SUBJ eat-AOR-ITR/MDL-IND

'I ate the food' (Osada, 1992, p. 136)

c. Locational/Directional NP

Ranci-ate-ge=m hidʒuʔ-aka-n-a
Ranchi-ABL-FOC=2SG.SUBJ come-ANT-ITR/MDL-IND

'you have come from Ranchi' (Osada, 1992, p. 137)

d. Verb

manɖi=ko dʒom-tan-ge-a
food=3PL.SUBJ eat-IPFV-FOC-IND

'they are eating food' (Osada, 1992, p. 137)

3 Syntactic Means of Encoding Focus in Munda languages

Text data suggests SOV is overwhelmingly dominant in all Munda languages. SVO order on the other hand focuses on the subject.

(6) Sora

mənra-n-dʒi gijʔ-le-dʒi ənsəlo-n-dʒi=adoʔoŋ
person-N.SFX-PL see-PST-PL woman-N.SFX-PL=OBJ

'*the men* saw the women'

(7) Mundari

sumit dal-dʒa-d-ko-a=e hon-ko
Sumit beat-TAM-TR/ACT-3PL.OBJ-IND=3SG.SUBJ child-PL

'*Sumit* beat the children'

Flipping objects and subjects focuses on the object and does so exhaustively in Sora or adds an emphatic connotation 'indeed'.

(8) a. Sora

ənsəlo-n-dʒi=adoʔoŋ mənra-n-dʒi gijʔ-le-dʒi
woman-N.SFX-PL=OBJ person-N.SFX-PL see-PST-PL

'the men saw *only the women*'

b. Sora

bogad sonum-ən amaŋ anlen pur-re-bi-n=den
if spirit-N.SFX DEP:OBJ.II we worship-PST-1PL-N.SFX=COND

'if indeed it is the spirit we have worshipped'

In Mundari, this scrambled word order rather also focalizes subjects.

(9) Mundari

hon-ko sumit dal-dʒa-d-ko-a=e
child-PL Sumit beat-TAM-TR/ACT-3PL.OBJ-IND=3SG.SUBJ

'*Sumit* beat the children'

OVS order in Sora marks exhaustive focus on the object but also carries a passive connotation.

(10) a. Sora

ənsəlo-n-dʒi=adoʔoŋ gijʔ-le-dʒi mənra-n-dʒi
woman-N.SFX-PL=OBJ see-PST-PL person-N.SFX-PL

'*only the women* were seen by Opino'

b. Sora

anin-dʒi sanil-ta-dʒi gordʒaŋ-mər-ən-dʒi
3PRON-PL follow/escort-NPST-PL village-person-N.SFX-PL

'they were followed by the villagers'

OVS order in Mundari on the other hand conveys focus on the object. Basically any element that appears post-verbally in Mundari typically assigns a focalized reading on that element.

(11) Mundari

hon-ko=e dal-dʒa-d-ko-a sumit
child-PL=3SG.SUBJ beat-TAM-TR/ACT-3PL.OBJ-IND Sumit

'Sumit beat the *children*'

VSO order in Sora focalizes the subject. This formation is somewhat common in our large text corpus.

(12) Sora

aŋanʈi dʒʰundʒum-boʲdʒ-ən a-ɖinna jər-ai-ʈen
when see.off-bride-N.SFX DEP-day come-CLOC-PRF

karra-n sinʈa-n-eʈin kən-i məndra
excess-N.SFX worry-ITR/MDL-PRF this-FOC man

‘When the day came for the sending off of the bride this very man was extremely worried.’

The word that immediately follows an initial verb in Mundari is in contrastive focus (13), which of course can include a subject NP (14), with the result being VSO order.

(13) Mundari

dal-dʒa-d-ko-a=e sumit hon-ko
beat-TAM-TR/ACT-3PL.OBJ-IND=3SG.SUBJ Sumit child-PL

‘*Sumit* beat the children.’

VOS in Sora is reported to focus the whole VP exhaustively according to speaker interviews, but it is textually unattested in our collection of sixty texts in six varieties of Sora.

(14) Sora

dʒum-le kinte-n opino-n
eat-PST banana-N.SFX Opino-N.SFX

‘Opino *already surely ate the banana only*’

VOS order entails focus on the object in Mundari on the other hand, as the element immediately following an initial verb focalizes it, including the object NP.

(15) Mundari

dal-dʒa-d-ko-a=e hon-ko sumit
beat-TAM-TR/ACT-3PL.OBJ-IND=3SG.SUBJ child-PL Sumit

‘Sumit beat *the children*.’

Cleft constructions, whether of the Wh- or ‘it’-type do not seem to form part of the natural repertoire of devices used for focalization in either Sora or Mundari, or at least no such formations spontaneously occur in our text corpus and are rejected by speakers when generated.

4 Prosodic Means of Encoding Illocutionary Force (and Focus?) in Munda languages

The acoustic cues associated with phonetic prominence in the Munda languages include vowel intensity, vowel duration, and fundamental frequency, e.g., in Sora (Horo et al., 2020, 2022, 2023a; Horo & Anderson, 2021a) or Assam Santali (Horo & Anderson, 2021b, 2023a). However, there is no commonality across all the languages as to which cue or sets of cues is favored for word- or phrase-level prominence and which may be used to indicate other categories, e.g., word boundary, phrasal edge or information structure.

For example, in Sora one finds vowel intensity peaks on the second syllable followed by a lengthening of word-final syllables to mark word-edge. Phrasally, pitch declines at the end except in Y/N questions where it rises.

4.1 Sora prosody, illocutionary force and focus

In Sora declaratives, with no special information structural properties, there is a decline in pitch in each word and the lowest pitch is on the final syllable of the utterance; see Figure 1 'little by little I will get cold' [∼little cold feel-NPST-1SG.UND].

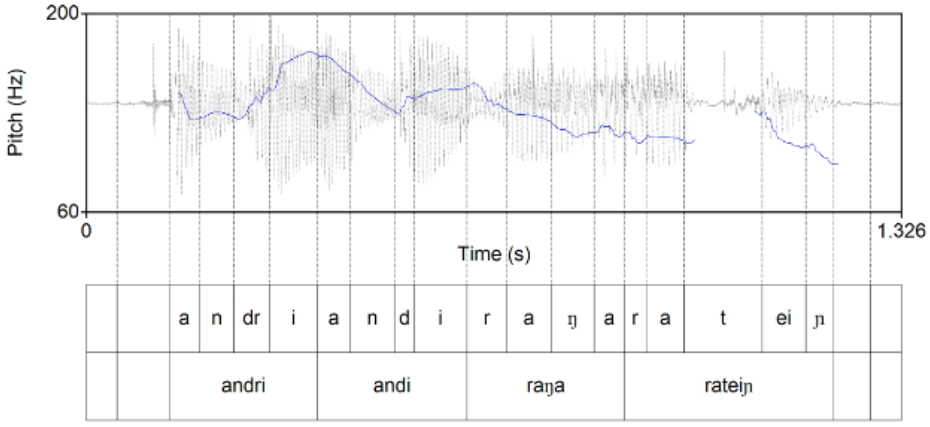


Figure 1: Pitch contour in a declarative sentence in Sora

In imperative formations as in Figure 2 'sleep carefully!' [one self-CV-I/M-sleep-I/M-IMP], all but the final word show pitch rising throughout the word and the imperative verb ending taking low pitch.

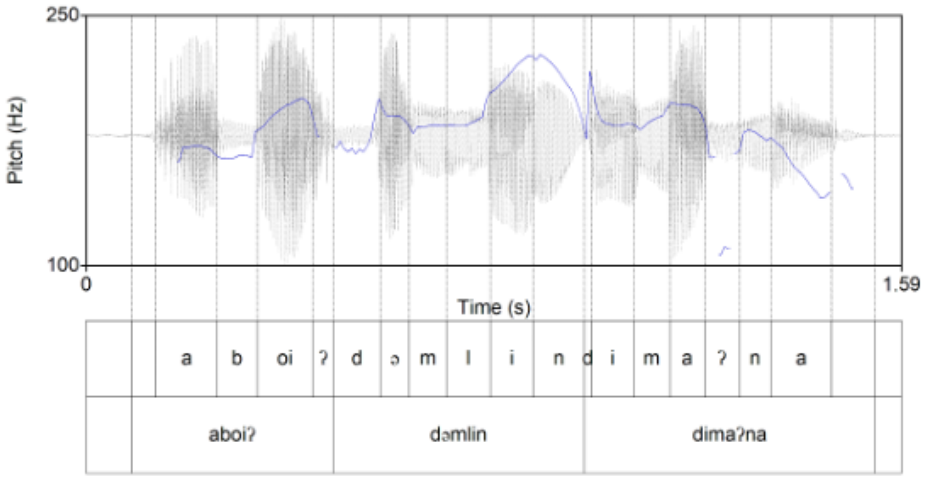


Figure 2: Pitch contour in an imperative sentence in Sora

Interrogative sentences with Wh-words show a similar pattern; see Figure 3 for 'what did you say?' [2SG what say-PST]. The Wh-element is by definition under focus and correspondingly there is high pitch on this element.

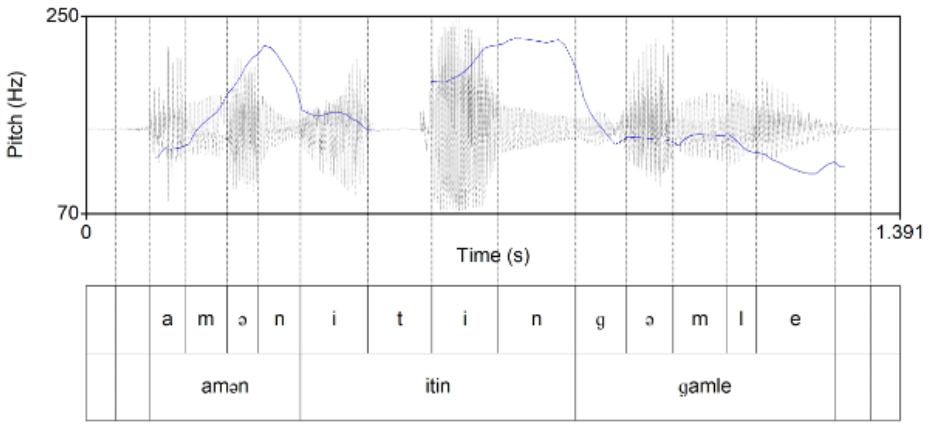


Figure 3: Pitch contour in a wh-word interrogative sentence in Sora

Yes/No questions, however, have rising pitch on the utterance final word in Sora which here is the interrogative particle *po*, as in Figure 4 for 'are you voiceless?' [2SG NEG.NMLZR-VOICE Q]. The entire predicate is in focus in such constructions and consequently there is also a rise in pitch on the Q-particle as is quite common cross linguistically, but noteworthy in Sora as there is typically a declination in pitch at the end of utterances including Wh-questions (see Figure 3) and declaratives (Figure 2).

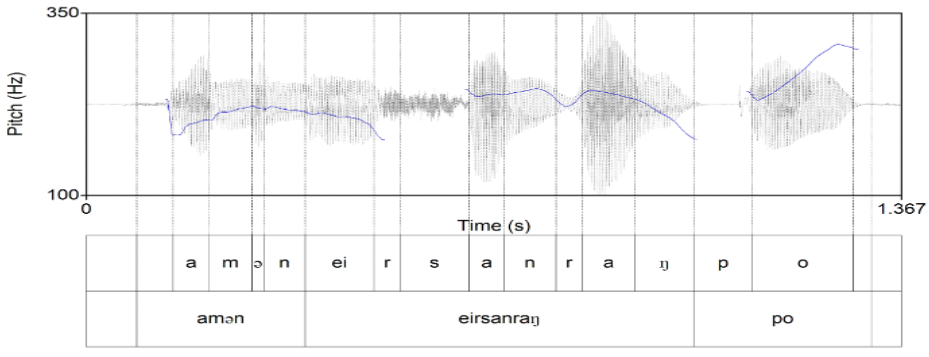


Figure 4: Pitch contour in a yes/no interrogative sentence in Sora

However, forced interpretation of elements as focalized in Sora (not using the focalizing element *=na* nor syntactic scrambling), using Odia and English translation to reinforce the

focalization shows no consistent realization acoustically in Sora, e.g., whether contrastive focus is on the subject or the predicate, these show the same pitch contours; see Figures 5 and 6.

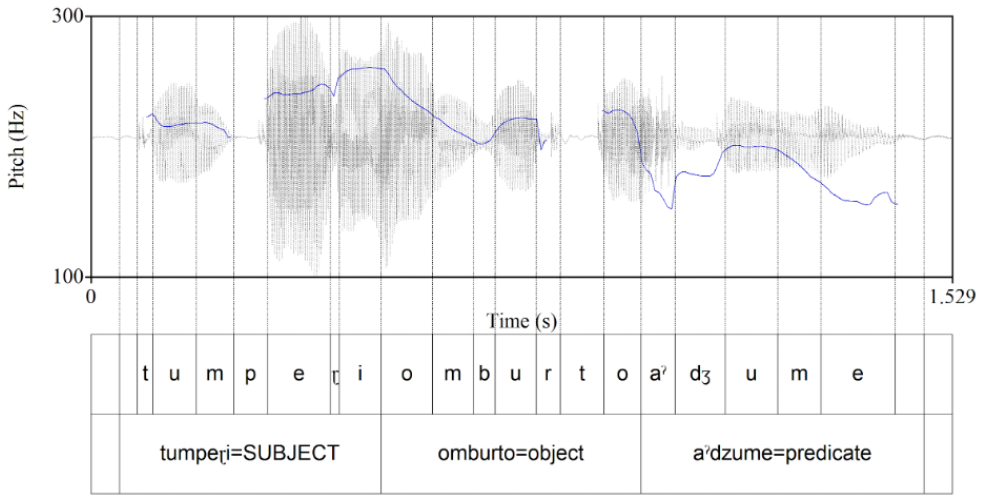


Figure 5: Pitch contour in a declarative form with contrastive focus on subject

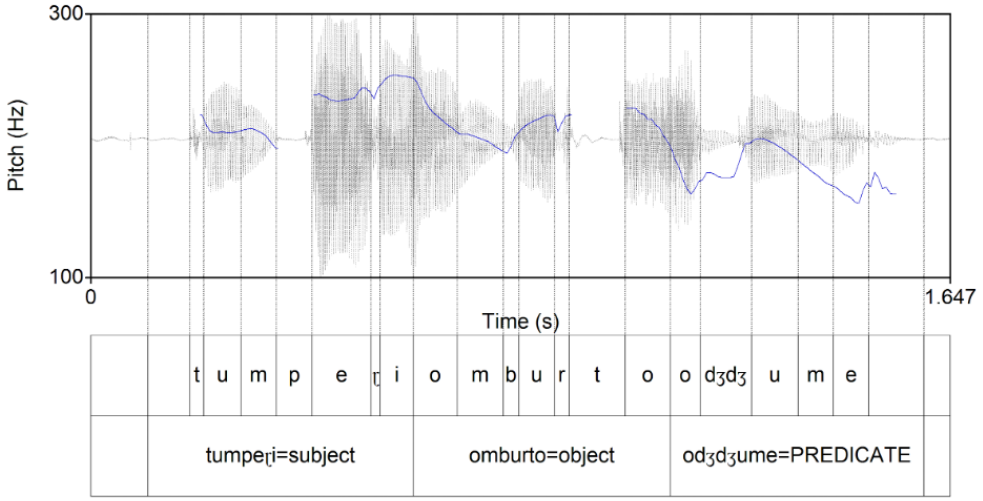


Figure 6: Pitch contour in a declarative form with contrastive focus on predicate

Therefore, prosodic means alone do not seem to be able to be used to encode focus in Sora, rather this should also accompany morpholexical means such as the interrogative particle *=po* or *=na*.

4.2 Mundari prosody and focus

Mundari disyllables show no consistent acoustic cue of lexical prominence. Longer duration on the second syllable is found in isolation (Figure 7), while intensity (Figure 8) and pitch (Figure 9) rise across the word. Note that all Mundari recordings were made in four contexts: in isolation (i), in a sentential/phrasal frame (p), a non-focal frame (n), and in a focal (f) frame.

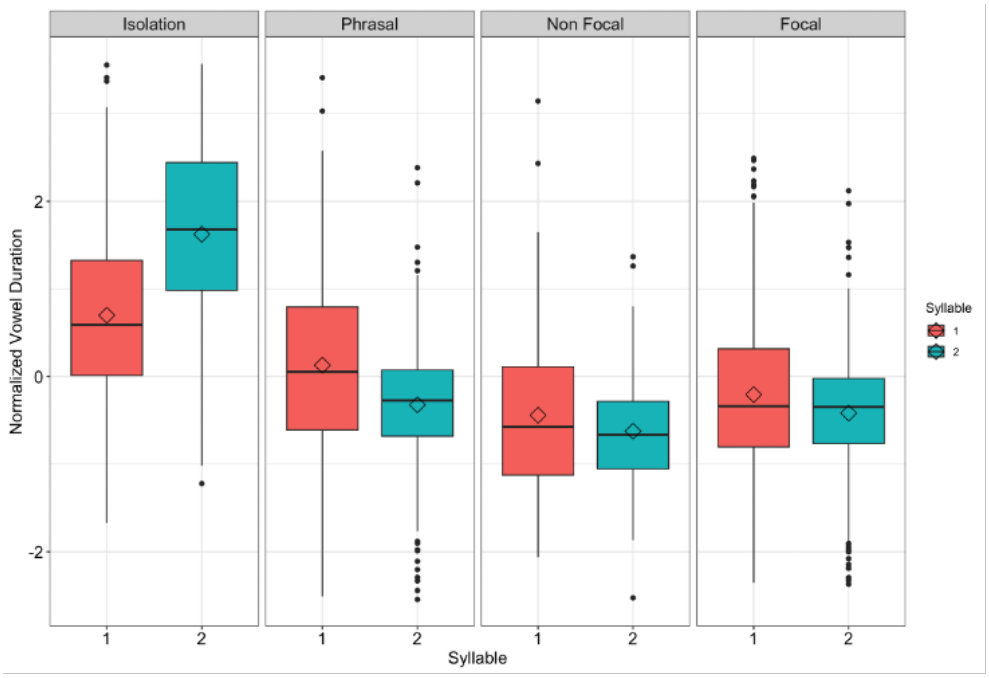


Figure 7: Distribution of vowel duration in Mundari disyllables produced in isolation and non-isolation contexts

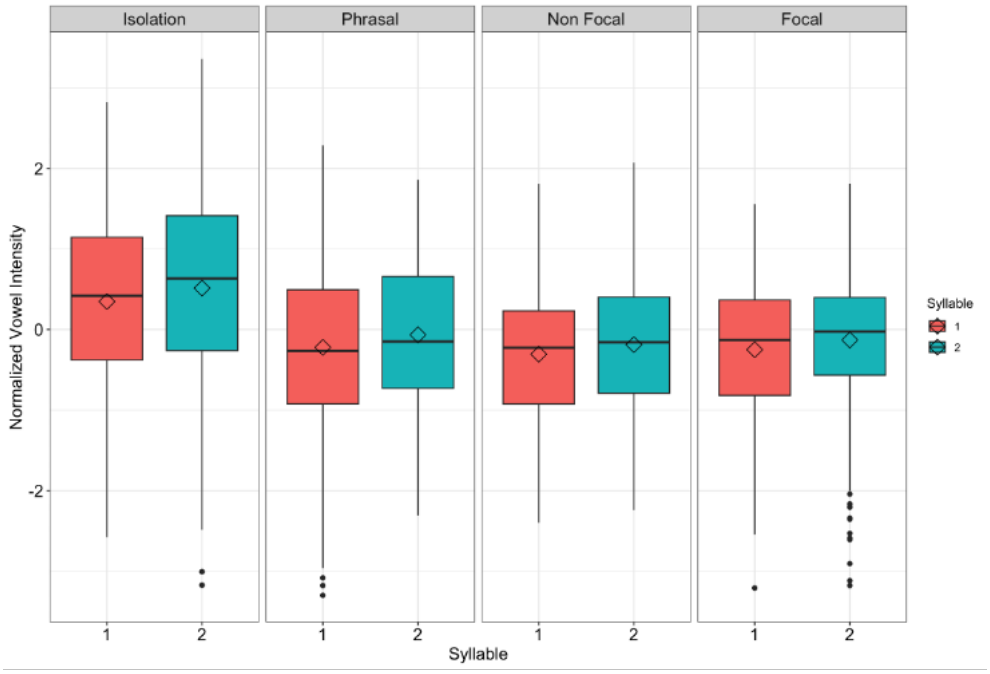


Figure 8: Distribution of vowel intensity in Mundari disyllables produced in isolation and non-isolation contexts

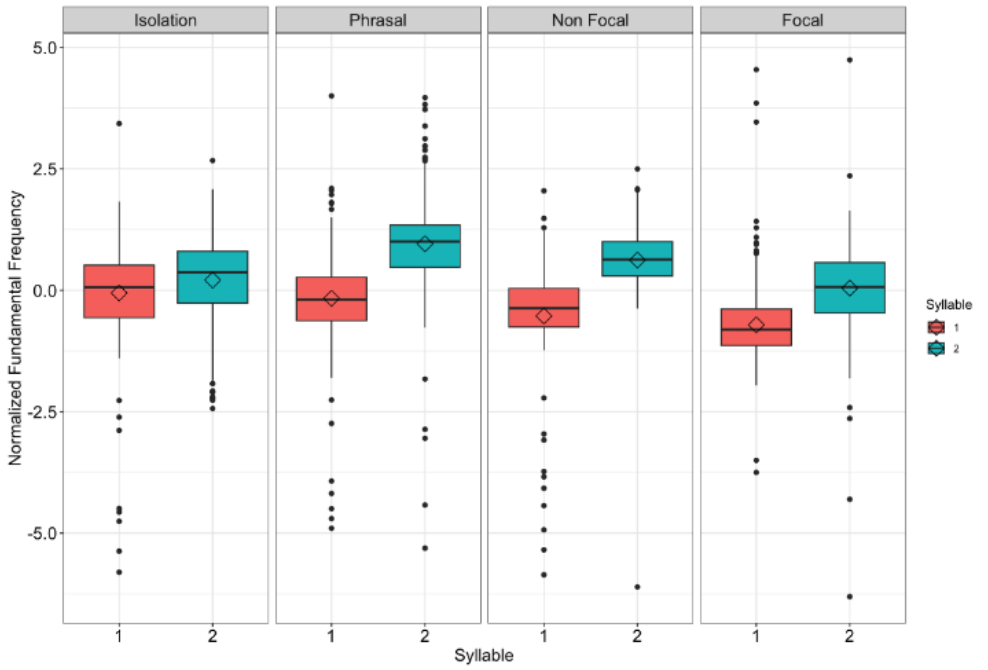


Figure 9: Distribution of mean *f*0 in Mundari disyllables produced in isolation and non-isolation contexts

Trisyllabic and tetrasyllabic words also show final lengthening in words said in isolation but there is no consistent pattern for vowel intensity or duration in non-isolation contexts (Figure 10-13).

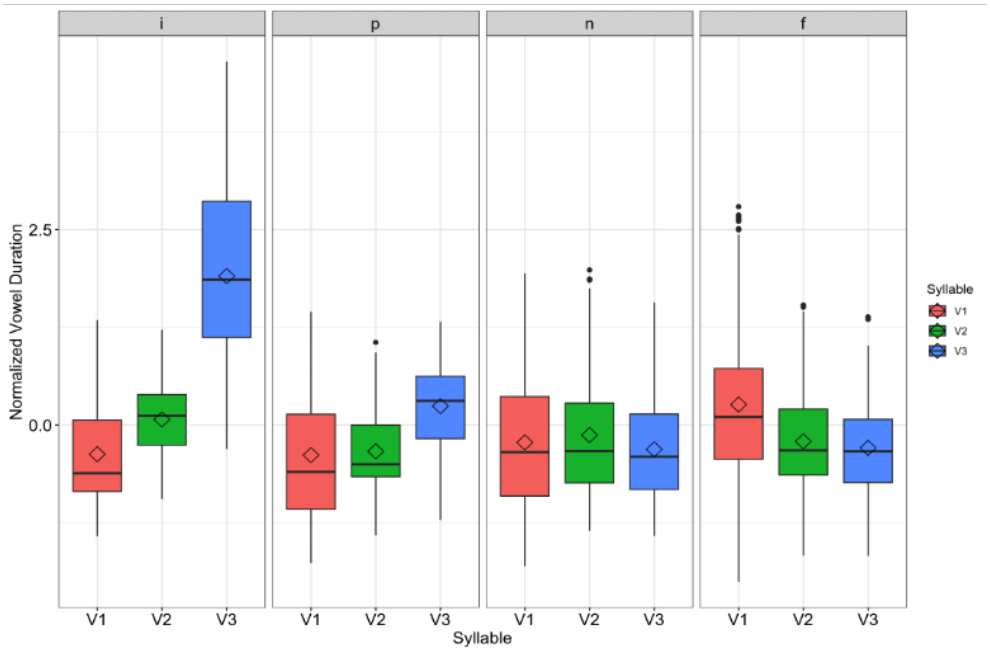


Figure 10: Distribution of vowel duration in Mundari trisyllables produced in isolation, phrasal, non-focal and focal contexts

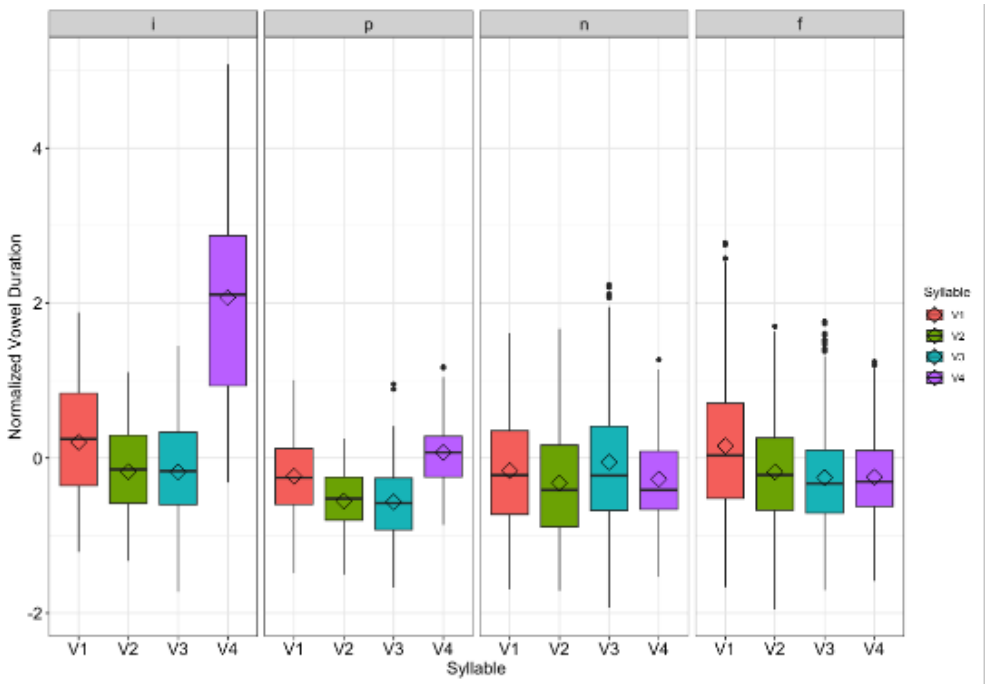


Figure 11: Distribution of vowel duration in Mundari tetrasyllables produced in isolation, phrasal, non-focal and focal contexts

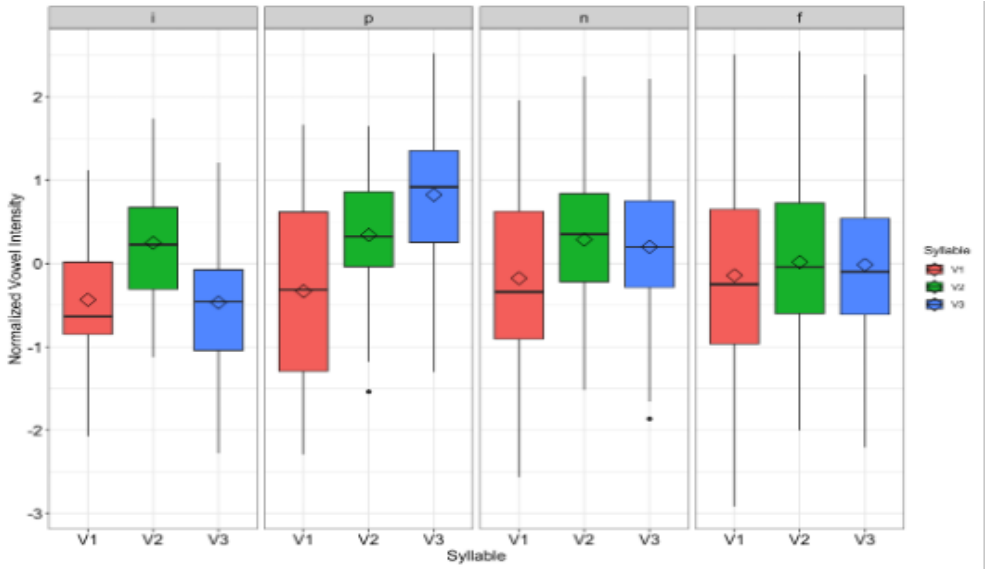


Figure 12: Distribution of vowel intensity in Mundari trisyllables produced in isolation, phrasal, non-focal and focal contexts

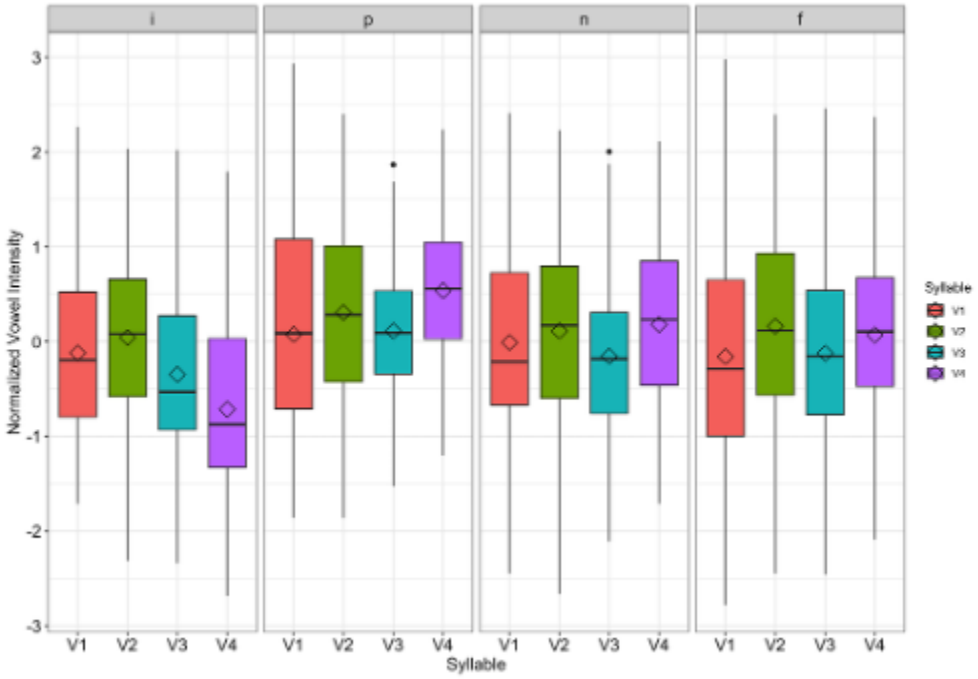


Figure 13: Distribution of vowel intensity in Mundari tetrasyllables produced in isolation, phrasal, non-focal and focal contexts

Moreover, as Figures 14-17 show, there is a confounding lack of patterning for five and six syllable words in Mundari with respect to either acoustic cue.

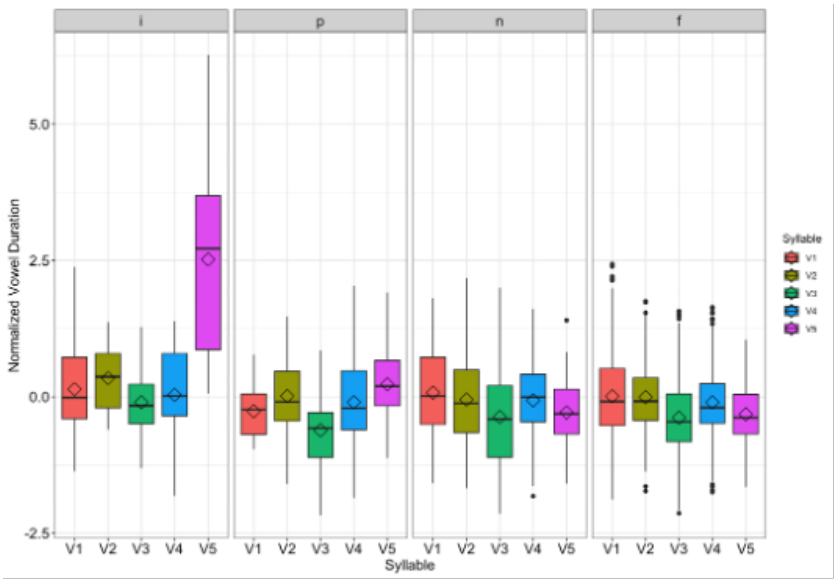


Figure 14: Distribution of vowel duration in Mundari pentasyllables produced in isolation, phrasal, non-focal and focal contexts

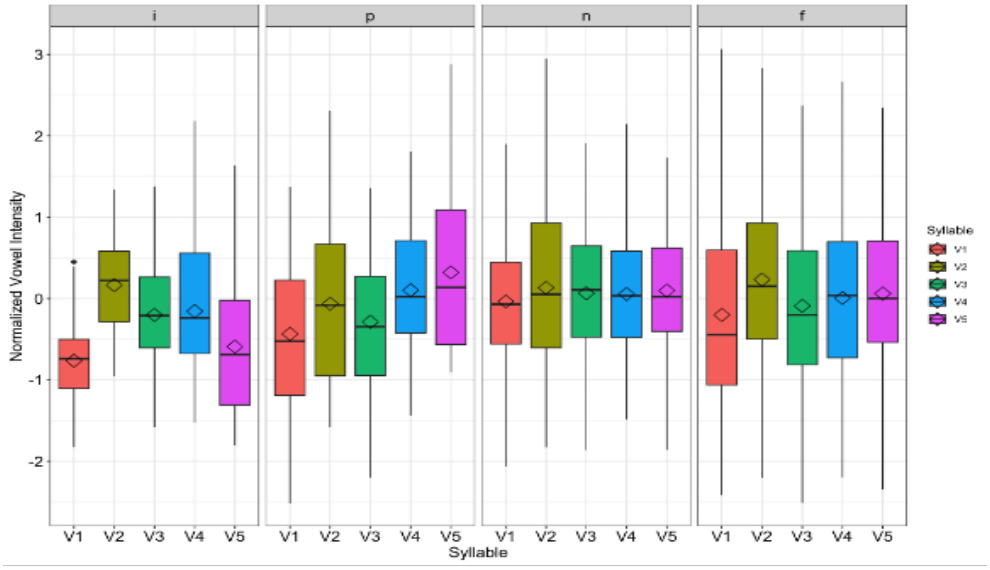


Figure 15: Distribution of vowel intensity in Mundari pentasyllables produced in isolation, phrasal, non-focal and focal contexts

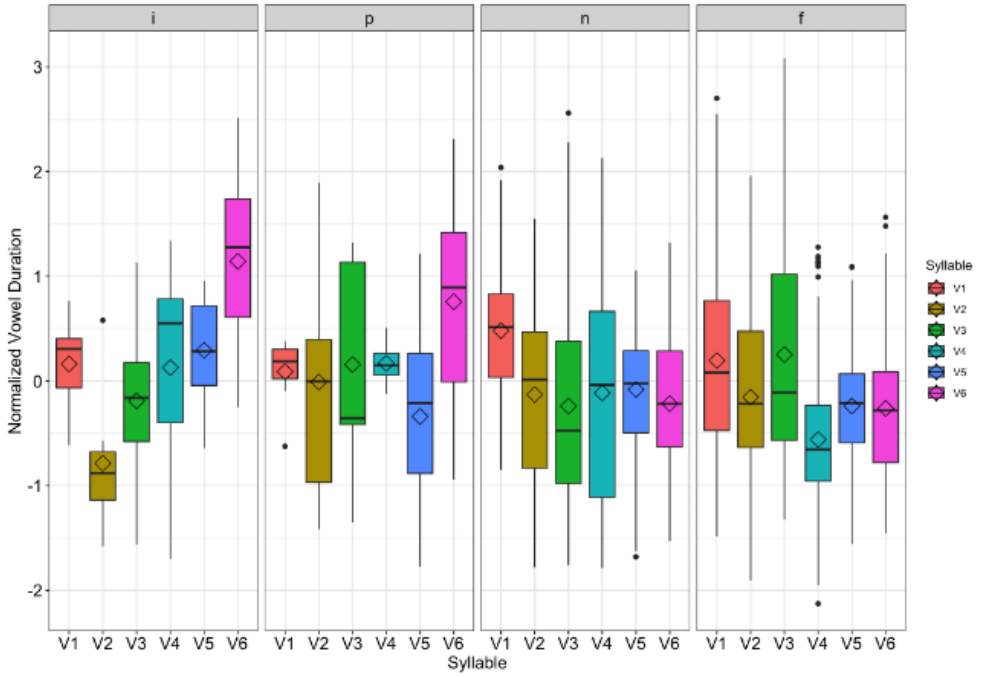


Figure 16: Distribution of vowel duration in Mundari hexasyllables produced in isolation, phrasal, non-focal and focal contexts

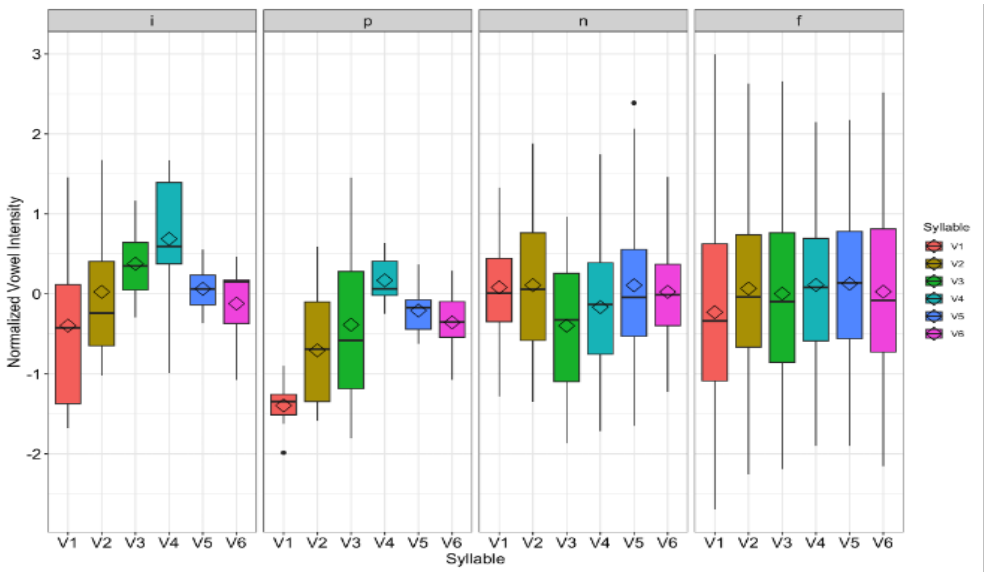


Figure 17: Distribution of vowel intensity in Mundari hexasyllables produced in isolation, phrasal, non-focal and focal contexts

Pitch, cued by fundamental frequency, shows nothing better either, whether trisyllables, tetrasyllables, pentasyllables or hexasyllables (Figures 18-21). These facts and others led us to propose that there is no level of phonological word in the phonoprosodic architecture of Mundari (Gogoi et al., 2024; Horo et al., 2024/2025). Note that while phonological word is not a category per se in Mundari, foot is the domain for a robust harmonic restriction against high and mid vowels. This domain is not, however, necessarily coterminous with the word or stem in Mundari.

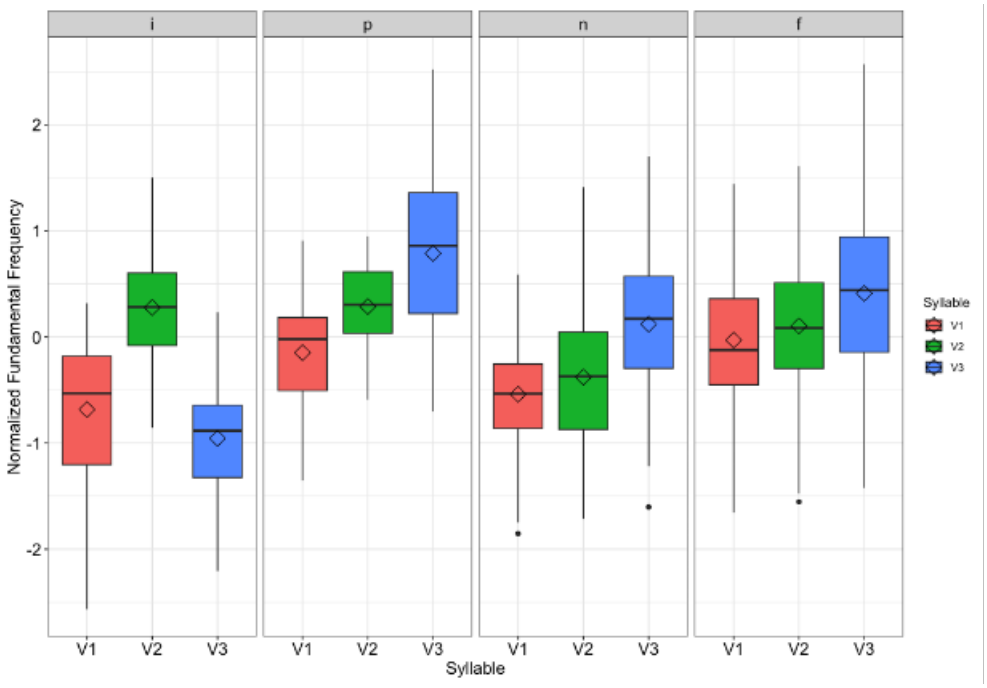


Figure 18: Distribution of mean F0 in Mundari trisyllables produced in isolation, phrasal, non-focal and focal contexts

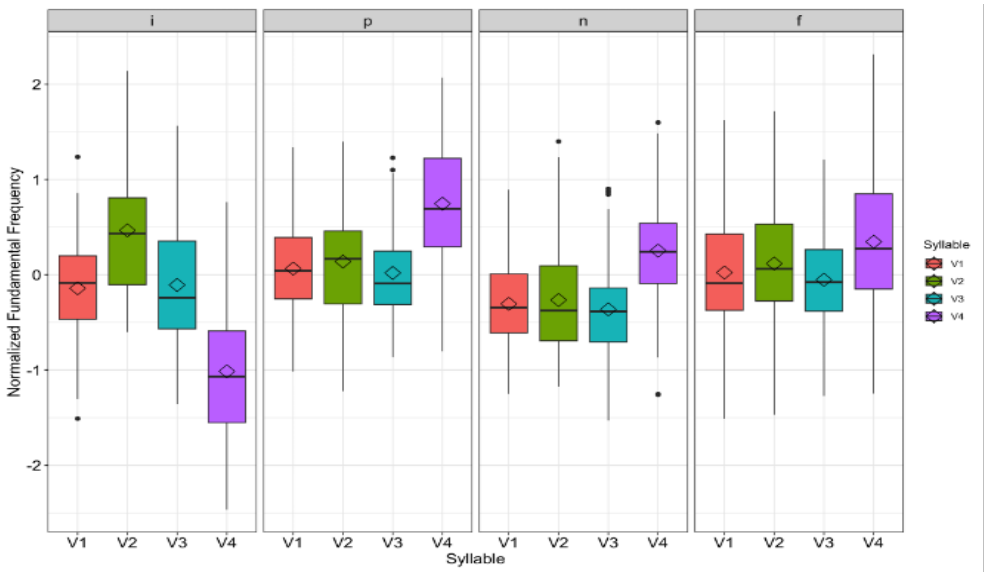


Figure 19: Distribution of mean F0 in Mundari tetrasyllables produced in isolation, phrasal, non-focal and focal contexts

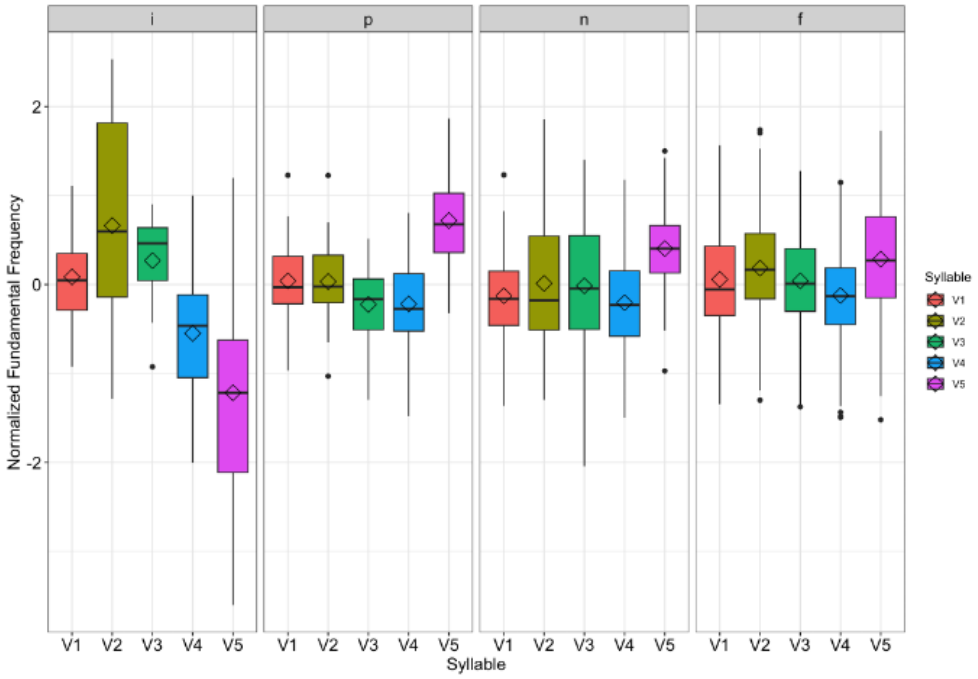


Figure 20: Distribution of mean F0 in Mundari pentasyllables produced in isolation, phrasal, non-focal and focal contexts

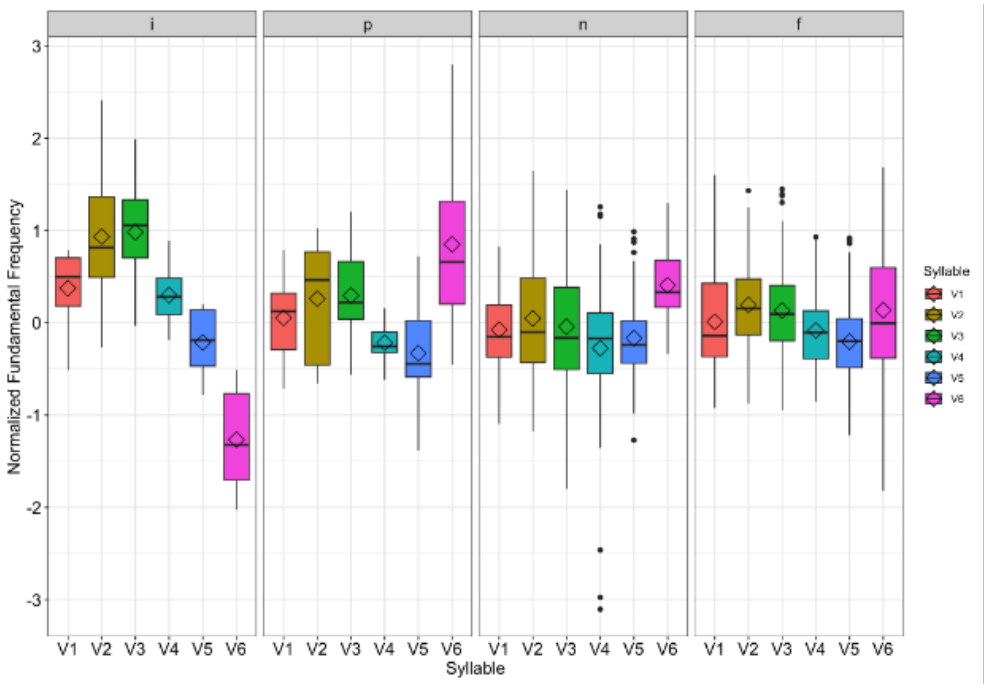


Figure 21: Distribution of mean F0 in Mundari hexasyllables produced in isolation, phrasal, non-focal and focal contexts

Put differently, there are no phonoprosodic processes or patterns in acoustic cuing that can justify any level larger than foot and smaller than utterance in Mundari, and none of the usual cues seem to have any role in information structuring at all. Perhaps this is predictable due to the presence and use of both morpholexical and syntactic means to encode information structure categories like topic or focus that are robustly used in Mundari. To be sure, there are general tendencies in utterances to see an increase in duration in final syllables, and a slight decrease in intensity and a decrease in pitch as a rule.

5 Summary

Munda languages may use morphological and syntactic means of encoding focus, but prosodic means are weakly developed and typically must be accompanied by morpholexical indices of information structure. They are each logically independent but can be interconnected. Individual languages show some commonalities, e.g., syntactically SOV marks subject focus and OSV object focus in both Sora and Mundari. But there are also differences found in the information structuring properties of scrambled syntactic configurations too, e.g., verb-initial order signals focusing on the verb or verb phrase in Sora, but VSO rather focuses on the subject and VOS on the object in Mundari. Some morphological markers might be old in the family in IS functions too, e.g., *do*, *ge*, *-de/i(ʔ)*, possibly others. Prosodic means of information focus in Sora interrogatives may show pitch peaks on the requested new information. Intensity is the cue to lexical prominence and generally peaks on the second syllable. Mundari has no prosodic cuing of prominence at the lexical level at all. Thus, Sora and Mundari prefer to utilize scrambled syntax and/or morphological indices encoding information structure together or independently rather than combining such with prosodic cuing of focalization for example, as commonly occurs with cleft constructions or scrambled syntax in English focalization.

Moving forward we will continue to map prosodic structures onto words (i.e., p- and g-words), phrases and sentences, as well as investigate the role of prosody in encoding information structure relations like focus in texts collected in four more Munda languages Kharia, Remo, Juang and Gtaʔ. This spans the full genetic spectrum within Munda. Once completed we will have a solid understanding of the formal and functional properties of information structuring in this important family of Austroasiatic languages of South Asia. The present study is just a first step into understanding the ways that Munda languages do and do not structure information. Once complete, we will have a clear idea the varied formal systems of prominence and focus attested in the family are and begin to be able to piece together how the individual attested systems may have arisen and changed over time.

Acknowledgment

Thanks to the US National Science Foundation grants #2041248 (Santali, Mundari, Kharia, Juang, Remo, Gorum, Gtaʔ) #1844532 (Sora) & the US National Endowment for the Humanities #PD-281083-21 (Santali, Mundari, Kharia, Juang, Remo, Gorum, Gtaʔ)

References

Asher, Nicholas. 2004. Discourse topic. In *Theoretical Linguistics 30*, 163–201.

Bearth, Thomas. 1999. The contribution of African linguistics towards a general theory of focus: Update and critical review. *Journal of African Languages and Linguistics* 20. 121—156. https://doi.org/10.1515/jall.1999.20.2.121.

Büring, Daniel. 1997. *The Meaning of Topic and Focus: The 59th Street Bridge Accent*. London: Routledge.

Büring, Daniel. 2010. Toward a typology of focus realization. In Malte Zimmermann & Caroline Féry (eds.), *Information Structure*, 177—205. Oxford: Oxford University Press.

Caron, Bernard. 2000. *Topicalisation et focalisation dans les langues africaines*. Louvain: Peeters [Afrique et Langage 1].

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles Li (ed.), *Subject and Topic*, 25—55. New York: Academic Press.

Dik, Simon C. 1980. On the typology of focus phenomena. *GLOT: Leids Taalkundig Bulletin* 3. 21—40.

Dik, Simon C. 1981. *Functional Grammar*. Amsterdam/New York: Oxford: North-Holland Publishing Company [North-Holland Linguistic Series 37].

Dik, Simon C. 1997. *The Theory of Functional Grammar I: The Structure of the Clause*. Berlin: Mouton de Gruyter.

Dik, Simon C, Maria Hoffmann, Jan R. de Jong, Djiang Sie Ing, Harry Stroomer & Lourens de Vries. 1981. On the typology of focus phenomena. In Harry van der Hulst Teun Hoekstra & Michael Moortgat (eds.), *Perspectives on Functional Grammar*, 41—74. Dordrecht: Foris Publications.

Drubig, Hans Bernhard & Wolfram Schaffar. 2001. Focus Constructions. In Martin Haspelmath et al (ed.), *Language Typology and Language Universals*, 1079—1104. Berlin: Mouton de Gruyter [Handbücher zur Sprach- und Kommunikationswissenschaft 20.2].

Fiedler, Ines, Katharina Hartmann, Brigitte Reineke, Anne Schwarz & Malte Zimmermann. 2010. Subject focus in West African languages. In Malte Zimmermann & Caroline Féry (eds.), *Information Structure: Theoretical, Typological, and Experimental perspectives*, 234—257. Oxford, New York: Oxford University Press. https://doi.org/10.18452/13433.

Gogoi, Pamir, Luke Horo & Gregory D. S. Anderson. 2024. Phonetic and Phonological Analysis of the Mundari Vowel System. In P. Sidwell (ed.), *Austroasiatic linguistics. in honour of gérard diffloth (1939-2023)*, 75–90. Chiang Mai: Myanmar Center, Chiang Mai University.

Gundel, Jeanette & Thorstein Fretheim. 2004. Topic and focus. In Larry Horn & Gregory Ward (eds.), *The Handbook of Pragmatics*, 175—196. Malden, MA: Blackwell Publishers.

Horo, Luke & Gregory D. S. Anderson. 2021a. Prosody and Morphosyntax in Sora: A preliminary study. In *Proceedings of 1st International Conference on Tone and Intonation (tai)*, 51–55. doi:10.21437/TAI.2021-11.

Horo, Luke & Gregory D. S. Anderson. 2021b. Towards a prosodic typology of Kherwarian Munda languages: Santali of Assam. In *Papers from the 30th Annual Meeting of the Southeast Asian Linguistics Society*, vol. 8, 298–317. JSEALS Special Publications.

Horo, Luke & Gregory D. S. Anderson. 2023a. A preliminary study on prominence in Assam Santali disyllables. In Gregory D. S. Anderson (ed.), *Munda Linguistics. Typological, Descriptive and Diachronic Perspectives*, 127–150. Pune: Deccan College.

Horo, Luke & Gregory D. S. Anderson. 2023b. Towards a prosodic typology of Kherwarian Munda languages: Hasadaʔ and Naguri Mundari. Presented at Northeast Indian Linguistics Society. (NEILS).

Horo, Luke, Gregory D. S. Anderson & Pamir Gogoi. 2024/2025. Can Vowel Height and Onset Duration Cue Prominence in Mundari? In *Papers from the 33rd Annual Meeting of the Southeast Asian Linguistics Society*, vol. 13, 137—151. JSEALS Special Publications.

Horo, Luke, Gregory D. S. Anderson, Aman Singha, Ria Borah Sonowal & Opino Gomango. 2023a. Acoustic phonetic properties of p-words and g-words in Sora. In *Proceedings of (formal) Approaches to South Asian Linguistics*, vol. 12, https://ojs.ub.uni-konstanz.de/jsal/index.php/fasal/article/view/290.

Horo, Luke, Pamir Gogoi & Gregory D. S. Anderson. 2023b. Prominence in Mundari in disyllables and inflected polysyllables. Presented at HisPhonCog.

Horo, Luke, Pamir Gogoi, Aman Kumar Singha, Ria Borah Sonowal, Kelsey Bialo, Bikram Jora & Gregory D. S. Anderson. 2023c. Prominence in Mundari in disyllables and inflected polysyllabic nouns. Presented at SEALS.

Horo, Luke, Priyankoo Sarmah & Gregory D. S. Anderson. 2020. Acoustic Phonetic Study of the Sora Vowel System. *Journal of the Acoustical Society of America* 147. 3000–3011. https://doi.org/10.1121/10.0001011.

Horo, Luke, Aman Kumar Singha, Ria Borah Sonowal & Gregory D. S. Anderson. 2022. Acoustic phonetic properties of polysyllabic Sora words. Presented at ICoLSI.

Jacobs, Joachim. 2001. The dimensions of topic-comment. *Linguistics* 39. 641–681.

Krifka, Manfred. 2004. The semantics of Questions and the Focusation of Answers. In Gordon C. Lee & Daniel Büring (eds.), *Topic and Focus: A Cross-Linguistic Perspective*, 139–151. Dordrecht: Kluwer Academic Publishers.

Krifka, Manfred. 2007. Basic Notions of Information Structure. In Caroline Féry et. al (ed.), *Interdisciplinary Studies on Information Structure (isis) 6*, 13–56. Potsdam: Universitäts-Verlag [Working papers of the SFB 632].

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: A theory of topic, focus, and the mental representations of discourse referents*. New York, NY, USA: Cambridge University Press.

Osada, T. 1992. *A reference grammar of Mundari*. Tokyo University of Foreign Studies: Institute for the Study of Languages and Cultures of Asia and Africa.

Sasse, H.-J. 1987. The thetic-categorical distinction revisited. *Linguistics* 25. 511—580.

Uhmann, Susanne. 1991. *Fokusphonologie*. Tübingen: Niemeyer.

van der Wal, Jenneke. 2016. Diagnosing focus. *Studies in Language* 40. 259—301.

Zimmermann, Malte. 2008. Contrastive focus and emphasis. *Acta Linguistica Hungarica* 55. 347—360.
Zimmermann, Malte. 2016. Predicate focus. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford Handbook of Information Structure*, Oxford: Oxford University Press.
Zimmermann, Malte & Edgar Onea. 2011. Focus marking and focus interpretation. *Lingua* 121. 1651—1670.
Zubizarreta, Maria Luisa. 1998. *Prosody, focus, and word order*. The MIT Press, MA: Cambridge dissertation.

Moving subjects out of finite *je*-clauses in Bangla: A case of Hyper-raising?

MOUMA BANIK, *University of Delhi*

ABSTRACT

In Bangla (Bengali; Eastern Indo-Aryan), an embedded subject from a finite complement clause can be moved out to the matrix clause which violates syntactic constraints like PIC, Activity Condition and Case theory. This phenomenon has been observed cross-linguistically and consequently named as hyper-raising (Ura 1994). Bangla exhibits an interesting case of this hyper-raising constructions where the embedded subject can only move out when the finite post-verbal complement clause lacks an overt C head *je* (initial complementizer, Singh 1980; Bayer 1996, 1999, 2001; Bhattacharya 2001, 2002, 2015, *among others*). This paper examines whether or not this movement is actually an instance of true A-movement. Along with this, a possible solution is also presented for the puzzle of the Bangla hyper-raising which lacks an overt complementizer head. This is different from other instances of hyper-raising seen cross-linguistically. The proposed solution shows why all the previous analyses may not be a right fit to the type of hyper-raising seen in Bangla, but can be simply answered through drawing a parallel to the *that*-trace effect seen in English *wh*-movement. Following Obata (2018) to explain the *that*-trace effect seen in the A-movement of the (hyper)-raised NP, it is shown that not C deletion but it is rather the external merge of the C to T that captures the absence of the IC head *je*- in Bangla hyper-raising.

1 Introduction

Movements of syntactic elements (such as nominals or pronominals, phrases, clauses, and so on) have taken center stage in syntactic work since the beginning of the generative grammar framework. One of the most ubiquitous and prominent movements of such is *raising* where a nominal is raised out of a non-finite complement clause to the [Spec, TP] of the matrix clause for reasons like Case and EPP in languages like English. Chomsky's idea of Phase (2000, 2001) is centered around these concepts of Raising, ECM, and Control. A non-finite embedded clause, which lacks a CP layer (as seen in English-type Raising), fails to be a phase – thus allowing movements out of it (PIC, Chomsky 2000, 2001). This movement is an instance of A-movement as the raised element goes to a targeted position ([Spec, TP]) in the matrix clause to check the EPP on the matrix T head and gets nominative Case on it.

(1) John$_i$ seems [$_{TP}$ t$_i$ to like the cake.]

But investigation into other languages shows us something different. A-movement of subjects or objects is allowed out of a finite embedded clause even when it is a CP, violating constraints like Phase Impenetrability Condition, Activity Condition (PIC, Chomsky 2000, 2001), and Case theory. This phenomenon has been termed as *hyper-raising* (Ura 1994). Languages such as Brazilian Portuguese (Ferreira 2000, 2004; Martins & Nunes 2005, 2009; *among others*), Bantu languages such as Zulu (Halpert 2012, 2015, 2019), Logoori and Tiriki (Diercks et al. 2022), Vietnamese and Cantonese (Lee & Yip 2024) and many more reveal instances where the embedded nominals can easily raise out of the finite embedded clause to an A-position in the matrix clause, violating all the constraints that have been discussed.

Although there has been much discussion on hyper-raising constructions in a large number of languages, there has not been any such discussion for Bangla. In this paper, I claim that these instances of A-movements across a finite clause boundary are also available in Bangla (Bengali; Indo-Aryan). In the language, subject-to-subject hyper-raising is seen across post-verbal complement clauses in the presence of a perception predicate like *mone hOy* ('seem') in the matrix clause.

(2) **Oroni**$_i$ **mon-e** **hO-y** [(*je) **t**$_i$ gotokal skul-e eS-ech-il-o]
oroni.NOM mind-LOC be-HAB.3 IC yesterday school-LOC come-PFV-PST-3
'Oroni seems to have come to school yesterday.'

As we see in example (2), the matrix predicate has default agreement (3P) on it which generally occurs when the subject is non-nominative (e.g., *amar* ('1SG.GEN'), *tomar* ('2SG.GEN'), *Rima-r* ('Rima-GEN') and so on). Along with this, I argue that in Bangla hyper-raising construction (see the example in (2)), there is a phonologically null or non-overt C head present and this IC *je* actually behaves similarly to the *that* in English *wh*-constructions. In English, movement of the embedded *wh*-phrase to the matrix [Spec, CP] position is disallowed but is available once the complementizer is deleted (Chomsky 2013, 2015). This has been known as the *that*-trace effect in literature. Chomsky (2015) has shown through labeling that this complementizer head is actually deleted syntactically, making T the new phase-head in the derivation. The deletion of this C makes the embedded CP phase weaker – which in turn, allows the movement of the embedded *wh*-phrases to a higher position (Chomsky 2013, 2015) in the matrix clause. But this stipulation becomes problematic for reasons like 'selection', 'legibility conditions', and 'recoverability conditions' (Obata 2018). I follow Obata (2018) to show that instead of C-deletion, external pair-merge of C to T will give us a neater solution for the ungrammaticality of the IC head *je* in example (2).

The sectioning of the paper will be as follows. In the next section, I give some data of the subject movement in Bangla across a finite clause boundary and show some of the key factors related to such movements in the language. In section 3, whether or not the movement that we see is a true A-movement will be examined through a few diagnostics. In section 4, I will propose a solution for the type of A-movement across finite clause boundary that we encounter in Bangla. Section 5 concludes the paper.

2 Empirical background: some key factors from Bangla

This section discusses some data from Bangla[1] where subject-to-subject movement is allowed across a finite clause boundary along with some other interesting factors seen in such constructions like the complementizer inside the complement clause, the raising predicate, and the agreement on such raising predicate.

Bangla, like other Eastern Indo-Aryan languages of India (e.g., Assamese, Odia) can have two ways of sub-ordination with respect to the position of the verb in the matrix clause (Singh 1980; Bayer 1999, 2001; Bhattacharya 2001, 2002, 2015; *among others*). An embedded subject from a finite post-verbal complement clause can raise out to the matrix subject position in presence of a specific predicate in Bangla. In example (3) below, the embedded subject *Oroni* has been moved to the subject position in the matrix clause. This subject-to-subject movement occurs across a finite clause boundary and in presence of a predicate *mone hOy* ('seem') which is a perception verb[2]. The embedded clause in the example below is also post-verbal in nature.

(3) **oroni**$_i$ **mon-e** **hO-y** [$_{CP}$ (*je) **t**$_i$ gotokal skul-e
oroni.NOM mind-LOC be-HAB.3 IC yesterday school-LOC
eS-ech-il-o]
come-PFV-PST-3
'Oroni seems to have come to school yesterday.'
Lit: 'Oroni seems has come to school yesterday.'

This subject-to-subject raising has been known as hyper-raising cross-linguistically (Ura 1994). But before we turn to see whether the subject-to-subject raising in Bangla (as seen in example (3)) is indeed a case of hyper-raising in the language or not, there needs to be a discussion about some of the factors related to such a movement in the language. There are similarities between the constructions seen in (3) and other languages that show hyper-raising. But interestingly, there are also a few differences in what happens in the embedded clause in Bangla as opposed to other languages exhibiting the same movement. I discuss both the similarities and the dissimilarities in details below.

2.1 Optionality of the movement

Although the movement in example (3) is readily available in Bangla, it is not obligatory; i.e. the subject may not move out of the embedded clause at all.

[1]Unless stated otherwise, judgement for all the Bangla data has been taken from the author herself who is a native speaker of the language.

[2]Perception verb is a type of attitude verb which talks about the speakers' sensory perceptual (visual or non-visual) experiences. Verbs like *seem*, *appear*, etc., are examples of visual perception verbs whereas verbs like *taste*, *feel*, *sound*, etc., are examples of non-visual perception verbs.

(4) (amar) mon-e hO-y [(je) **Oroni** gotokal skul-e
1SG.GEN mind-LOC be-HAB.3 IC oroni.NOM yesterday school-LOC
eS-ech-il-o]
come-PFV-PST-3
'I think/ feel that Oroni has come to school yesterday.' Or 'It seems to me, that Oroni has come to school yesterday.'

In the example above, the subject *Oroni* does not move out of the embedded clause. It remains in-situ and a genetive marked experiencer subject *amar* (1SG) can sit in the subject position of the matrix clause or be dropped.

This movement of the subject is completely unavailable out of a non-finite embedded clause as seen in the example below.

(5) *Oroni mon-e hO-y [t_i gotokal skul-e as-te]
Oroni.NOM mind-LOC be-HAB.3 yesterday school-LOC come-INF
'Oroni seems to have come to school yesterday.'

Such optionality of movement has also been observed across Bantu languages like Zulu, Logoori, Tiriki (Halpert 2012, 2015, 2019; Diercks et al. 2022), etc. Bangla also patterns similarly to these Bantu languages in how the movement of the subject is optional and yet banned out of a non-finite embedded clause. This gives us some useful insight into what happens in Bangla.

2.2 Complementizer in the embedded clause

A-movement of the subject from embedded clause to the matrix clause, as seen in example (3) above, is only allowed across a post-verbal[3] complement clause in Bangla. And it is a known fact that the post-verbal clauses in Bangla are headed by an initial complementizer (or IC) *je* (Singh 1980; Bayer 1999, 2001; Bhattacharya 2001, 2002, 2015, *among others*), which can be dropped from the embedded clause it appears in.

[3]An anonymous reviewer pointed out whether or not such A-movements across pre-verbal *bole*-clauses are allowed in the language. To answer this question, we have to look at what happens both inside and outside such pre-verbal clauses headed by a final complementizer *bole* (derivative of SAY-verb). When the embedded clause is pre-verbal, out of which the subject moves out to a higher position, it becomes problematic to examine whether the subject is raised or base-generated in the matrix clause due to the linear order observed in the example.

(i) Oroni$_i$ [t_i gotokal skul-e eS-ech-il-o (bole)] mon-e hO-y
oroni.NOM yesterday school-LOC come-PFV-PST-3 mind-LOC be-HAB.3
'Oroni seems to have come to school yesterday.'

(6) robi bhab-l-o [(**je**) Oroni gotokal skul-e eS-ech-il-o]
robi.NOM think-PST-3 IC oroni.NOM yesterday school-LOC come-PFV-PST-3
'Robi thought that Oroni came to school yesterday.'

Interestingly, this is not what is seen in examples like (3) where the movement of the embedded nominal across finite clauses is allowed in the language. The initial complementizer *je* in these constructions is non-overt. In fact, presence of the IC *je* inside the post-verbal finite embedded clause makes the sentence ungrammatical[4].

(7) *Oroni$_i$ mon-e hO-y [**je** t$_i$ gotokal skul-e eS-ech-il-o]
oroni.NOM mind-LOC be-HAB.3 IC yesterday school-LOC come-PFV-PST-3
'Oroni seems to have come to school yesterday.'
Lit: *'Oroni seems that has come to school yesterday.'

This absence of an overt complementizer inside the embedded clause is not what we see in a canonical hyper-raising construction generally. Also, this ungrammaticality of *je* inside the embedded clauses brings forth a question: does the embedded clause even constitute a CP when the complementizer itself is absent? I will return to this question in the later sections of the paper.

2.3 Selected predicate that aids the movement

In Bangla, such an embedded subject movement is licensed by a specific perception predicate *mone hOy* ('seem') as seen in example (3)[5]. The interesting thing to notice here is that this verb is restricted in its form when it licenses such a movement in the language.

(8) a. Oroni$_i$ **mon-e hO-y/ ho-cch-e** [(*je) t$_i$ gotokal
oroni.NOM mind-LOC be-HAB.3/ be-PRES.PROG-3 IC yesterday
skul-e eS-ech-il-o]
school-LOC come-PFV-PST-3
'Oroni seems to have come to school yesterday.'

b. *Oroni **mon-e ho-l-o/ ho-yech-il-o/ ho-cch-il-o/ ho-t-o/**
oroni.NOM mind-LOC be-PST-3/ be-PFV-PST-3/ be-PROG-PST-3/ be-IMPF-3/
hO-b-e [(*je) t$_i$ gotokal skul-e eS-ech-il-o]
be-FUT-3 IC yesterday school-LOC come-PFV-PST-3
'Oroni seems to have come to school yesterday.'
Lit: 'Oroni is seeming to have come to school yesterday.'

[4]For some speakers of the language, presence of this *je* inside the embedded clauses does not sound as bad. One reason for the acceptability can be due to the possibility of *je* being a discourse particle (see Bayer & Dasgupta (2016) for more discussion.)

[5]*mone hOy* ('seem') in Bangla is an example of a complex predicate where the first verb is a nominal marked with a locative *-e* marker and the second verb *hO-y* ('be-HAB.3') is an inflected auxiliary.

In the set of examples shown above, the tense morphology on the second verb (which generally takes the tense and agreement markers in a complex predicate in Bangla) of the predicate is restricted to either *hO-y* ('be-HAB.3) or *ho-cch-e* ('be-PRES-PROG.3'), no other tense markers are allowed in an environment where the subject movement happens.

But this changes when the subject is non-nominative and the complementizer *je* inside the embedded clause is overt. In example (9), the matrix verb can take all the different tense inflections on it[6].

(9) amar mon-e **hO-y/ ho-cch-e/ ho-l-o/ ho-yech-il-o/**
1SG.GEN mind-LOC be-HAB.3/ be-PRES.PROG-3/ be-PST-3/ be-PFV-PST-3/
ho-cch-il-o/ ho-t-o [(je) Oroni gotokal skul-e
be-PROG-PST-3/ be-IMPF-3/ IC oroni.NOM yesterday school-LOC
eS-ech-il-o]
come-PFV-PST-3
'I think/ feel/ am thinking/ am feeling/ thought/ felt/ had thought/ had felt/ was thinking/ was feeling that Oroni has come to school yesterday.' Or, 'To me, it seems/ is seeming/ seemed/ had seemed/ was seeming/ used to seem that Oroni has come to school yesterday.'

In example (9), the verb does not allow A-movement of the embedded nominal *Oroni* to the matrix clause[7]. There is also a noticeable distinction in the meaning of the two verbs used in both examples (8) and (9). In the first set of examples, when the verb is restricted in its form and allows movement, the meaning is similar to a raising verb like 'seem' in English. But the verb in example (9) gives a meaning similar to the attitude verbs 'feel' or 'think' or non-raising 'seem' in English.

[6]I purposefully avoid the usage of future tense (*mone hObe*, ('will seem/ feel/ think')) of the perception verb in the current set of examples as it will bring forth the discussion of different properties of this verb which is beyond the scope of our current discussion. But the usage of this form does exist in the language as seen in the example below:

(ii) amar mon-e hO-b-e ar Oroni skul-e co-l-e ja-b-e?
1SG.GEN mind-LOC be-FUT-3 and oroni.NOM school-LOC go-PRT go-FUT-3?
'Just because I thought of it, it won't mean that Oroni will go to school.'
Lit: 'I will think and Oroni will go to school?'

[7]A'-movement (like topicalization) of the embedded nominal in such sentences is allowed.

(iii) Oroni$_i$, amar mon-e **hO-y/ hocche/ ho-l-o/ ho-yech-il-o/**
oroni.NOM 1SG.GEN mind-LOC be-HAB.3/ be-PRES.PROG.3/ be-PST-3/ be-PFV-PST-3/
ho-cch-il-o/ ho-t-o [(je) t$_i$ gotokal skul-e eS-ech-il-o]
be-PROG-PST-3/ be-IMPF-3 IC yesterday school-LOC come-PFV-PST-3
'Oroni, I think/ feel/ am thinking/ am feeling/ thought/ felt/ had thought /had felt/ was thinking/ was feeling that he has come to school yesterday.' Or 'Oroni, to me, it seems/ is seeming/ seemed/ had seemed/ was seeming/ used to seem that he has come to school yesterday.'

This duality of the perception verb *mone hOy* ('seem'/ 'feel'/ 'think') in how they behave differently with respect to their tense morphology, availability of an overt complementizer, and meaning will give us more insight into the movement of the embedded subject across the finite post-verbal clauses in Bangla.

2.4 Default agreement on the matrix verb

In Bangla, agreement between subject and verb is restricted to only person feature. Agreement failure happens when the subject is non-nominative and as a result the verb gets default agreement on it.

(10) **amar/ tomar/ o-r/ Oroni-r** khub khide pe-yech-**e**
1SG.GEN/ 2SG.GEN/ 3SG-GEN/ oroni-GEN very hunger get-PRES.PFV-3
'I/ you/ he/ she/ Oroni feels hungry.'

The default agreement in Bangla as shown in example (10), is realized as 3P at the end of the verb in the clause. This is exactly what we see in our examples where the subject movement is allowed across a finite clause boundary.

(11) **ami**$_i$/ **tumi**$_i$/ **Se**$_i$/ **Oroni**$_i$ mon-e hO-**y** [t$_i$ gotokal
1SG.NOM/ 2SG.NOM/ 3SG.NOM/ oroni.NOM mind-LOC be-HAB.3 yesterday
skul-e eS-ech-il-**am/e/o**]
school-LOC come-PFV-PST-1/2/3
'I/ You/ He/ She/ Oroni seem(s) to have come to school yesterday.'

As seen in example (11), even when the moved subject in the matrix clause is in first person, second person or third person, the agreement on the matrix clause remains 3P which is the default agreement in the language. Again we encounter a puzzling issue that even though the embedded subject moves to the matrix clause, it does not agree with the matrix verb at all. This non-agreeing pattern of the matrix verb is similar to what we see in languages like Zulu (Halpert 2012, 2015, 2019), Logoori and Tiriki (Diercks et al. 2022), and so on.

I will now turn to diagnose whether these movements are true instances of A-movement or not along with other factors that are important to understand what happens in hyper-raising constructions in Bangla.

3 Diagnosing hyper-raising in Bangla

Beginning from Ura (1994) and continuing to the most recent syntactic works across various languages (Ferreira 2000, 2004; Halpert 2012, 2015, 2019; Diercks et al. 2022; Martins & Nunes 2005, 2009; Greeson 2023; Lee & Yip 2024), A-movement of subjects (or objects) out of finite clauses or hyper-raising has started to gain enough attention in the literature. More and more investigation into the topic has revealed a lot of interesting factors regarding

hyper-raising cross-linguistically. Along with this, there have been multiple ways to diagnose such A-movement cross-linguistically (Martins & Nunes 2005, 2009; Halpert 2015, 2019; Diercks et al. 2022; Greeson 2023; Lee & Yip 2024, *among others*). In this section, I will adopt some of these tests to show what happens in case of Bangla[8].

3.1 Against base-generation

I argue that the movement that we see in examples such as (3), are due to displacement of the subject from the embedded clause to the matrix clause. The subject in the matrix clause is not base-generated in that position, but is moved from a finite embedded clause to that position. The evidences are given below.

A subject that has been moved to a higher position in the matrix clause can be sensitive to island constraints (Greeson 2023; Lee & Yip 2024). In a complex NP island in Bangla, the moved subject fails to establish any thematic relation with the embedded predicate inside that island.

(12) *Island sensitivity*

*__Oroni__$_i$ mon-e hO-y [[__t__$_i$ atokkhone skul-e col-e
oroni.NOM mind-LOC be-HAB.3 already school-LOC go-PRT
g-ech-e khobor-Ta] mithye chi-l-o]
go-PRES.PFV-3 news-CLF false be-PST-3

'The news that Oroni has already gone to school was false.'

The raised subject *Oroni* in example (12) is sensitive to subject-island constraint when the verb *mone hOy* ('seem') is present in the matrix clause. The ungrammaticality of the sentence in the example above gives us evidence that the NP *Oroni* is indeed a result of movement in examples like (3).

3.2 A-movement, not A'-movement

As shown in the section above, the subject in the matrix clause is indeed a result of movement. But it cannot prove whether this movement is A-movement or A'-movement. I will now give some evidence for the type of movement that happens in these constructions.

As many of the languages that exhibit hyper-raising are also instances of *pro*-drop languages, a doubt arises: whether the moved subject is a left dislocated topic with a *pro*-

[8]One important thing to note here is that Bangla is a typologically different language than a lot of the languages that have been diagnosed with hyper-raising. Not every test that has been used for the other group of languages, can be used for what we see in Bangla as the properties of the language like subject-hood, topicalization, prosody, nature of the complementizer, and even the raising predicate differ vastly. So, it is crucial to understand that such A-movements are truly language specific and the tests need to be adapted according to the specific language that is under discussion.

drop[9] in its matrix subject position. Bangla, a partial *pro*-drop language[10], also can be an instance of such mechanism. But I show through the tests given below, this is not what happens in Bangla. The subject-to-subject movement seen in example (3) is a true instance of A-movement.

In hyper-raising languages where a subject has been raised to a higher position, tests like quantifier raising have been used to show if this subject displacement is an A'-movement (topicalized NP) or true A-movement (cf. Martins & Nunes 2009; Lee & Yip 2024, *among others*). In Bangla, a quantified phrase *keu* ('someone') fails to be topicalized, but can be raised.

(13) *Quantifier raising*

a. ***keu**$_i$*, Oroni amake **bol-l-o** [(je) **t**$_i$ por-e g-ech-e]
someone Oroni.NOM 1SG.ACC say-PST-3 IC fall-PRT go-PRES.PFV-3
'Oroni said to me that someone has fallen down.'

b. **keu**$_i$ **mon-e** **hO-y** [**t**$_i$ por-e g-ech-e]
someone mind-LOC be-HAB.3 fall-PRT go-PRES.PFV-3
'It seems that someone fell.'

In example (13a), when the embedded quantifier phrase is moved to the matrix clause in presence of a verb like *bollo* ('said'), the sentence sounds ungrammatical. But in example (13b), the quantified NP has been moved across the finite clause boundary in presence of the matrix predicate *mone hOy* ('seem'). This gives evidence towards A-movement of the raised subject in examples like (13b).

Other diagnostics like the idiom test have been used to examine extractibility of an element. For hyper-raising constructions too, people have used idioms to see if the movement across the raising predicate is an A-movement or not (Halpert 2012, 2015, 2019; Diercks et al. 2022; Lee & Yip 2024, *among others*). In an A'-movement like topicalization, the idiomatic meaning would not be retained after the nominal has been moved. In Bangla, in presence of the matrix predicate *mone hOy* ('seem'), even if we raise the embedded NP to a matrix subject position, the idiomatic meaning of the sentence is retained, proving it is an A-movement.

(14) *Idioms*

a. koi maach-er pran boro-i Sokto
koi fish-GEN life very-EMPH tough
Int: 'A tough nut to crack.' or 'Someone who has a very tenacious soul.'
Lit: 'The koi fish has a strong will to live.'

[9]See Diercks et al. (2022) to understand how 'LD + *pro*-expletive' analysis works or does not work to diagnose hyper-raising in a language.

[10]See Holmberg et al. (2009) to know more about why Bangla is a partial null subject language (NSL).

b. **koi**$_i$ maach-er pran **mon-e** **hO-y** [**t**$_i$ boro-i Sokto]
koi fish-GEN life mind-LOC be-HAB.3 very-EMPH tough
Lit: 'The koi fish seems to have a tough life.'

c. *koi mach-er pran$_i$, ram bol-l-o [(je) t$_i$ bOro-i Sokto]
koi fish-GEN life ram.NOM say-PST-3 IC very-EMPH tough
'The koi fish's life, Ram said, is very tough.'

The example in (14c), when the NP *koi maacher pran* ('the koi fish's life') is topicalized to the front of the matrix clause, the sentence becomes ungrammatical. The idiomatic meaning in this sentence is lost; contrary to what can be seen in (14b) where it is retained fully.

Along with the tests that have shown that the movement in example (3) is indeed an instance of an A-movement, we also now turn to test the weak cross-over effects (or WCO) in Bangla. WCO is generally used to differentiate between A and A'-movement in a langauge. It also creates new binding possiblities (see Greeson (2023) for English, and Lee & Yip (2024) for Vietnamese and Cantonese WCO). Weak cross-over effect arises when a pronominal fails to co-refer to a displaced and quantified antecedent of it. A lack of this WCO effects attest to the fact that the movement is an A-movement.

(15) a. *Bound reading impossible*
*[**tader**$_i$ **malik-ra**] mon-e hO-y [**proti-Ta kukur-ke**$_i$] roj
2PL.GEN owner-CLF mind-LOC be-HAB.3 each-CLF dog-ACC everday
park-e niy-e ja-y
park-LOC take-LOC go-HAB.3
Int: 'The owners of each and every dog seem to take their dog to the park everyday.'

b. *Bound reading possible*
[**proti-Ta kukur-ke**$_i$] mon-e hO-y [**tader**$_i$ **malik-ra**] roj
each-CLF dog-ACC mind-LOC be-HAB.3 2PL.GEN owner-CLF everday
park-e niy-e ja-y
park-LOC take-LOC go-HAB.3
'The owners of each and every dog seems to take their dog to the park everyday.'

As seen in examples (15b), the pronominal *tader* ('their') can be bound by the antecedent *proti-Ta* ('every') which deletes the WCO effect from these sentences. This is completely different to what is seen in example (15a) where the binding between the pronominal and antecedent fails, resulting in WCO effect in the structure. This further proves that the movement shown in example (3) is A-movement and not topicalization.

3.3 Movement across a finite CP

The embedded clause, as shown in example (3), lacks an overt complementizer (IC *je*). This might lead us to believe that the embedded clause might not be a CP, but something smaller than that (a TP or *v*P/VP). But the evidence shows that the embedded clause is indeed a finite CP.

Only a finite TP in Bangla can host agreement, tense and aspectual information inside and this is what we see in example (16). The raised subject *Oroni* exhibits person agreement (3P) with the embedded verb *eSechilo* ('has come') in the example below. Along with this, the embedded verb has a past tense (*-il-*) and aspect (*-ech-*) markers on it.

(16) *Tense and agreement inside the embedded clause*

Oroni$_i$ mon-e hO-**y** [(*je) t$_i$ gotokal skul-e eS-**ech-il-o**]
oroni.NOM mind-LOC be-HAB.3 IC yesterday school-LOC come-PFV-PST-3

'Oroni seems to have come to school yesterday.'

Following *feature inheritance* (Chomsky 2008), the finite T head inside the embedded clause gets all of its ϕ-features from the C head (which is the non-overt IC *je* in the example (16)) above in virtue of C being a phase head. The transferred ϕ-features on the embedded T then establish an Agree relation with the embedded nominal *Oroni* before it raises to the matrix clause. This proves that the embedded clause is a CP, and not anything smaller.

As Cinque (1999) suggests, speaker-oriented adverbs like *unfortunately*, *luckily*, *mistakenly* etc., appear high on a clausal spine. The presence of such adverbs gives us evidence that the embedded clause must have a CP projection and is finite (Huang 2022; Lee & Yip 2024). In Bangla, an adverb like *bhul kore* ('mistakenly') can appear inside the embedded clause.

(17) *Availability of speaker-oriented adverb inside the embedded clause*

ami$_i$/ **tumi**$_i$/ **Se**$_i$/ **Oroni**$_i$ mon-e hO-**y** [bhul
1SG.NOM/ 2SG.NOM/ 3SG.NOM/ oroni.NOM mind-LOC be-HAB.3 mistake
kor-e t$_i$ gotokal skul-e col-e eS-**ech-il-am/e/o**]
do-PRT yesterday school-LOC walk-PRT come-PFV-PST-1/2/3

'I/ you/ he/ she/ Oroni seem(s) to have mistakenly come to school yesterday.'

Thus, it is proved that the embedded clause inside example (3) is indeed a finite CP and the subject-to-subject movement happens across this finite CP boundary.

All the diagnostics that have been used in the section above give us enough evidence that what we see in example (3), is indeed a case of true hyper-raising in Bangla. Now I will try to propose a solution for the syntactic derivation of such constructions.

4 The proposal

In this section, I will first discuss some of the previous analyses that talk about how hyper-raising (subject or object) occurs cross-linguistically. Along with this, I will examine if any of these approaches can be applied to the hyper-raising constructions in Bangla or not. A proposed solution will then be given to analyse these structures in the language drawing parallel with the A'-movement in *wh*-constructions in English. I will also give a working hypothesis about the remaining puzzle of the matrix verb in such constructions in Bangla.

4.1 Previous analyses

There have been several approaches to solve the puzzle of hyper-raising cross-linguistically. From non-phase or defective phase analysis (Ferreira 2000, 2004; Zeller 2006; *among others*) to phase deactivation (Halpert 2019), and movement to the phase edge (Tanaka 2002; Zyman 2017; Fong 2019 *among others*) – each of these try to answer the issues that arise when a subject (or object) is moved out of a finite embedded clause to the [Spec, TP] position of the matrix clause. For our current discussion, I will examine if the first two analyses can be applied in Bangla subject-to-subject hyper-raising seen in examples like (3)[11]. In a hyper-raising environment, the movement of the subject (or object) poses question on constraints such as: (i) Phase Impenetrability Condition (PIC, Chomsky 2000, 2001), (ii) Activity Condition (Chomsky 2001), and (iii) Case theory.

According to the non-phase or defective phase analysis, the subject is allowed to move out of the embedded clause due to the embedded CP lacking in some properties (finiteness, or agreement, or in some other categorial feature and so on (cf. Zyman 2023)). This type of analysis has been used for languages like Brazilian Portuguese (henceforth, BP; cf. Ferreira 2000, 2004; Martins & Nunes 2005, 2009, *among others*) where the movement of the subject out of finite CP is obligatory and not optional and the raised subject agrees with the matrix verb. In these hyper-raising constructions of BP, the embedded T has been argued to be ϕ-defective, thus unable to give nominative Case to the embedded subject. In contrast, the matrix T is ϕ-complete due to which the embedded subject can raise out of the defective CP phase to get its nominative Case checked by the matrix T. The defectiveness of this embedded T has been attested by the fact that in BP, the T head and its finiteness have been simplified throughout the time regarding its morphology (cf. ibid). This is not what we see in Bangla. As discussed in the examples from Bangla, the raised subject does not agree with the matrix verb in a hyper-raising environment. Along with this, the embedded T in Bangla hyper-raising constructions is also not ϕ-defective (see section 3.3, example (16)). Thus, the non-phase or defective phase analysis fails to work in Bangla.

In Brazilian Portuguese, one of the key arguments for the licensing of hyper-raising has been that the embedded nominal when raised, agrees with the matrix verb or raising

[11] The phase-edge analysis has been adopted for mostly hyper-raising to object position cross-linguistically. The current discussion revolves around subject-to-subject hyper-raising in Bangla and although the concept of phase-edge analysis may prove to be useful for such Bangla hyper-raising, I will not consider this for now and focus only on why both defective phase and phase deactivation fail to work in Bangla properly.

predicate. This is not always true for the Bantu languages. Zulu (Halpert 2012, 2015, 2019), along with other Luiya languages (a sub-group of Bantu languages[12]) like Logoori and Tiriki (Diercks et al. (2022)), shows 'non-agreeing hyper-raising' where the raised subject does not agree with the raising predicate. Halpert (2019) tries to solve this issue of 'agreeing' and 'non-agreeing' hyper-raising by postulating phase deactivation analysis. In phase deactivation analysis, an Agree relation is established between the embedded clause (finite CP or non-finite TP) and the raised subject. This Agree relation deactivates the phase boundary due to which the subject can be moved to the [Spec, TP] position in the matrix clause. She further suggests that the embedded CPs have phi-features and and these phi-features play crucial role in establishing two Agree cycles in Zulu hyper-raising constructions. She further provides evidence from Zulu that when the matrix T probe phi-agrees with the embedded CP, the (non-agreeing) **ku-** (17s noun class marker) appears on the matrix verb – thus establishing a non-agreeing hyper-raising in the language.

This account of phase deactivation by establishing an Agree relation between the raised subject and embedded CP has been adopted to several hyper-raising languages (cf. Lee & Yip (2024) for Vietnamese and Cantonese hyper-raising and Greeson (2023) for hyper-raising in English). The idea has been that the CPs (and non-finite TPs) in these languages bear certain ϕ-features which makes the CP a non-intervenor in hyper-raising instances. But this idea of CPs bearing phi-features does not work in Bangla.

(18) *Agreement between CP subject and verb*

*[je o bajar-e g-ech-il-o], amake khub obak kor-l-o
that 3SG.NOM market-LOC go-PFV-PST-3 1SG.ACC very surprise do-PST-3
'That he went to the market, made me very surprised.'

(19) *Agreement inside complex DP constructions*

[ei khobor-Ta [je o mara g-ech-e]] amake khub obak kor-l-**o**
this news-CLF that he/she die go-PFV-3 1SG.ACC very surprise do-PST-3
'The news that he has died, made me very surprised.'

Bangla does not have any object agreement, and a CP cannot occur in the subject position as shown in example (18). As seen in Zulu (Halpert 2012, 2015, 2019), in a complex DP construction, a complex DP in the subject position may show agreement with the verb. But as seen in example (19), we cannot claim that there is agreement between the complex CP subject and the verb *korlo* ('did') conclusively because of the third person agreement marker (which is also a default agreement marker in the language) on the verb. Thus, the phase deactivation account also cannot solve what happens in Bangla hyper-raising.

Apart from the issues that create a barrier in adopting these two hyper-raising analyses in Bangla, we will see in the next section why there may not even be a need to resort to these analyses of hyper-raising for our language.

[12]See Diercks et al. (2022) for more discussion on the classification of Luiya languages and how non-agreeing hyper-raising occurs in them.

4.2 What happens in Bangla?

Now that it has been shown what happens in hyper-raising constructions cross-linguistically and why the analyses that worked for those languages may not work for Bangla, I will try to show exactly what happens in Bangla hyper-raising constructions that set them apart from all the other languages. The difference between what happens in a canonical raising in a language like English, what happens in a typical hyper-raising constructions in languages like Brazilian Portuguese (Ferreira 2000, 2004; Martins & Nunes 2005, 2009, *among others*), Bantu languages like Zulu (Halpert 2012, 2015, 2019) etc., and compare it to what happens in Bangla has been shown schematically below.

(20) Canonical raising: [XP seem [$_{TP_{[NF]}}$ t$_{XP}$ . . .]]

(21) Hyper-raising: [XP seem [$_{CP}$ (C) [$_{TP_{[F]}}$ t$_{XP}$ . . .]]]

(22) Bangla hyper-raising: [XP seem [$_{CP}$ (*C) [$_{TP_{[F]}}$ t$_{XP}$ . . .]]]

One crucial difference that can be seen in Bangla, but is not seen in any of the hyper-raising languages, is the absence of the C head inside the embedded clause. As discussed in section 2 and 3, although the C head is non-overt in the embedded clause, it is indeed a CP. I argue that the answer lies in what happens in the *that*-trace effect seen in English *wh*-constructions. In both constructions, the C heads are non-overt or morphologically null – creating the assumption that the C head is syntactically deleted (Chomsky (2013, 2015)) in these structures. But there are issues that come with such a stipulation. I discuss them below and show why C deletion may not work for Bangla hyper-raising and to make our argument stronger for the absence of *je* inside the embedded clause, we might need to resort to a better solution.

4.2.1 C-deletion or something else?

There can be a parallel drawn between the absence of complementizer in English *wh*-constructions to what we see in Bangla hyper-raising constructions, arguing that both of these exhibit *that*-trace effect. In both of these constructions when we try to move an embedded nominal to a higher position in the matrix clause, the complementizer inside the embedded clause needs to be non-overt. Only then the movement is allowed. The only difference between the two is that the former is an instance of A'-movement, while the latter is an A-movement.

(23) a. *wh*-constructions in English → A'-movement

b. hyper-raising in Bangla → A-movement

Now, I first assume that there is an apparent complementizer deletion happening in hyper-raising constructions seen in example (3), similar to what can be seen in English *wh*-constructions where A'-movement of a *wh*-phrase is only allowed once the C head *that* is deleted (Chomsky 2013, 2015).

(24) a. Who$_i$ did you think [t$_i$ drew the picture on the board]?
b. *Who did you think [that t$_i$ drew the picture on the board]?

As Chomsky (2013, 2015) suggests, deletion of the overt C head *that* in example (24) weakens the phase and thus *wh*-phrases can move to the [Spec, CP] position in the matrix clause. I temporarily adopt this account of C deletion for the subject-to-subject hyper-raising at hand. In example (3), which is repeated in (25) below, I assume that the IC head *je* is deleted. This weakens the embedded CP phase boundary and as a result, the matrix T probe can look inside and raise the embedded nominal to matrix [Spec, TP] for it to satisfy the uD/uEPP feature on the matrix T, licensing what is known as hyper-raising in the language.

(25) **Oroni**$_i$ mon-e hO-y [(*je) **t**$_i$ gotokal skul-e eS-ech-il-o]
oroni.NOM mind-LOC be-HAB.3 IC yesterday school-LOC come-PFV-PST-3
'Oroni seems to have come to school yesterday.'

But this idea of C-deletion inside the embedded clause does not capture what happens in Bangla subject hyper-raising correctly. According to what Chomsky (2015) suggests, it can be inferred that in English *wh*-constructions, the matrix verb (such as *think, wonder*, etc.) selects a TP when the C head is syntactically deleted in the derivation. But this creates several issues for the hyper-raising constructions in Bangla seen in example (25) above. As shown in section 3.3, the embedded clause unfurls a CP projection regardless of the C head being present or absent inside it. So how do we exactly answer what happens in Bangla hyper-raising? I try to give an answer for it next.

4.2.2 Phase cancellation, external pair-merge, and hyper-raising in Bangla

Obata (2018) presents multiple issues that arise with the idea of C-deletion (cf. Chomsky 2013, 2015) to explain *that*-trace effect in English *wh*-constructions. She shows that if the C head is deleted in the syntax, it creates doubt on factors like *selection*, *legibility condition*, and *recoverability condition*[13].

It has already been discussed what happens if we assume that the matrix verb selects a TP, not a CP in Bangla (see section 3.3), giving us enough evidence how the 'selection condition' will be violated in Bangla hyper-raising constructions if the C head is deleted. On the other hand, the deleted C needs to be legible at LF, which cannot be supported if C is syntactically deleted as suggested by Chomsky (2013), creating issues with the 'legibility condition' both in English *wh*-constructions in example (24) as well as in the hyper-raising constructions in Bangla (see example (25)). Lastly, the deleted C in both constructions is never recoverable which is also in violation of the 'recoverability condition'.

[13]Obata (2018) shows both empirical and theoretical implications that underlie the assumption that C is syntactically deleted in English *wh*-constructions. The reader is suggested to look at Obata (2018) for a detailed discussion on the topic.

Obata (2018) gives a solution for these issues. Following Epstein et al. (2016), she suggests **external pair-merge of C to T**[14] for *that*-trace effect in English *wh*-constructions which cancels the phase boundary and licenses the movement of the embedded *wh*-phrase. This does not violate any of the stipulations mentioned above and gives us a neater and more unifying account for *that*-trace effect, restricting syntactic operations to Merge only. I summarize this idea in (26).

(26) **Phase cancellation through external pair-merge**: Under certain conditions, a phase CP is canceled iff the embedded C is externally pair-merged with the embedded T, licensing movements of the embedded nominal out of such phase.

For our hyper-raising cases in Bangla, I adopt this idea by extending the external pair-merge of C to T for the subject movement in examples like (25). In Bangla hyper-raising constructions, as we can see, the IC *je* sounds ungrammatical once raising occurs. Similar to English bridge verbs in *wh*-constructions, Bangla *mone hOy* ('seem'/ 'feel'/ 'think') in all environments (hyper-raising or non-hyper-raising) also takes a CP complement always (see example (14) and (15) for the non-raising counterparts of the matrix verb). I use labeling to show what happens in the derivation of a hyper-raising example such as (25) in Bangla. The derivation is as follows.

(27) Derivation of (25) *Oroni*$_{\text{i}}$ *mone hOy* [(****je***) t_{i} *gotokal skule eSechilo*], following Obata (2018):
[$_{\gamma}$ *Oroni*$_{\text{i}}$ *mone hOy* [$_{\alpha}$ t_i [$_{\beta}$ *gotokal skule eSechilo*]]]
Step 1: [$_{\alpha}$ *Oroni* <$T_{u\phi}$,C> [$_{\beta}$ t_{Oroni} … *eSechilo*]
(i) **External pair-merge** of <T,C> (ordered pair): C is affixed with T and this affixed C becomes invisible to further syntactic operations. The composite <T,C> becomes the new phase-head and all ϕ-features are transferred to T.
(ii) Internal pair-merge of β with the ordered pair <T, C>
(iii) NP *Oroni* raises from the domain of <T,C> (i.e., β) to α (in the Spec position of <T, C>).
Step 2: [$_{<\phi,\phi>}$ *Oroni* <$T_{u\phi}$,C> [$_{\beta}$ t_{Oroni} … *eSechilo*]
Labeling: α is labeled as <ϕ,ϕ> due to feature matching.
Step 3: [$_{<\phi,\phi>}$ *Oroni* <T,C>]
The domain of the new phase-head composite <T,C>, β, gets transferred to the interfaces.
Step 4: [$_{\gamma}$ [*mone hOy* [$_{<\phi,\phi>}$ *Oroni* <T,C>]]
The embedded nominal inside <ϕ,ϕ> is available for raising due to the composite <T,C> being the new phase-head (*Oroni* is in the phase edge position which is now visible to the matrix T probe).
Step 5: [$_{\gamma}$ *Oroni*$_i$ [*mone hOy* [$_{<\phi,\phi>}$ t_{i} <T,C>]]]
The embedded nominal raises to the matrix clause.

[14]External pair-merge is a 'structure-building operation' (cf. Epstein et al. 2016) where any combination of phrases or heads are adjoined together pre-syntactically.

The above step-by-step derivation through labeling gives us a clear picture of what happens in Bangla hyper-raising instances. Thus, not C deletion, but rather external pair-merge of <T,C> explains how the phase boundary becomes opaque for an embedded nominal to raise out of it, resulting in a subject-to-subject movement in examples like (25).

4.3 A conjecture about the matrix verb *mone hOy*

A puzzle still remains regarding the nature of the matrix predicate that is seen in Bangla hyper-raising constructions (see example (3)). I try to propose a tentative solution to this issue at hand.

One thing that has been argued to play a crucial role for hyper-raising across all the languages is the type of predicate that gets selected in the matrix clause. It has been shown that in raising and hyper-raising both, A-movement of the embedded nominal occurs when there is a perception verb like *seem*, *appear*, etc., in the matrix clause.

There has also been much discussion about what type of information a matrix verb must carry for it to allow hyper-raising in the language (Lee & Yip 2024; Greeson 2025, *among others*). According to Greeson (2025), it is the 'thematic properties' of the matrix perception verb that decides which predicates will allow hyper-raising and which will not. He argues that when the matrix verb gives a particular 'P(ERCEPTUAL)-SOURCE θ-role' to its external argument in the matrix clause, hyper-raising is blocked across the finite CP boundary. But when the verb fails to give this 'P(ERCEPTUAL)-SOURCE θ-role', hyper-raising can occur. Greeson (2025)'s analysis of the matrix verb is based on the evidentiality associated with the matrix verb. He follows Lee & Yip (2024)'s argument of bifurcating the raising predicates into two types according to what it 'encodes'. Lee & Yip (2024) state that in Vietnamese and Cantonese, if the matrix predicate 'encodes' indirect evidence, hyper-raising may occur. But when the matrix predicate 'encodes' direct evidence, the subject from the embedded CP fails to (hyper-)raise to the matrix subject position ('The Subject-Evidentiality Correlation'; cf. Lee & Yip 2024).

This direct-indirect bifurcation (cf. Lee & Yip 2024; Greeson 2025) which has been imposed on the matrix verb of hyper-raising and their non-hyper-raising counterparts does not really work for the matrix verbs in Bangla. As shown in section 2.3, the information carried by the two verbs in the two sets of examples (see example (8a) and (9)) is completely different. The verb in (8a), I suggest is what we can call the raising predicate in Bangla as opposed to the verb in example (9) which fails to license hyper-raising of the embedded nominal across clause-boundary.

Following Greeson (2025)'s analysis partially, I suggest that the raising predicate in example (8a) also lacks the ability to give 'P-SOURCE θ-role' to the raised NP in a hyper-raising example. But in Bangla, where the matrix verb *mone hOy* ('seem') in example (8a) is restricted in its tense morphology and the embedded clause also lacks an overt IC head *je*, this inability to give any θ-role is not due to 'indirect' or 'direct evidence' associated with the verbs in example (8a) and (9), but because the verb in (8a) has been grammaticalized[15]

[15]The grammaticalization incline of the verb *mone hOy* ('seem') as an evidential phrase needs to examined

as an *evidential phrase* in the language, frozen in its form. The other one still works as a canonical verb, being able to give a 'P-SOURCE θ-role' to its arguments as well as carrying various tense morphology on it. I suggest that the default agreement on the matrix predicate seen in examples like (3) which establishes a non-agreeing pattern between the raised subject and the matrix verb is also due to this nature of the raising predicate.

If the assumption that the matrix predicate *mone hOy* ('seem' or 'think' or 'feel') being two different forms (one an evidential phrase and the other a canonical verb) is correct, then this duality of the two forms makes it easier to identify which predicates will allow hyper-raising and which will not. A deeper understanding of this duality of *mone hOy* ('seem' or 'think' or 'feel') needs to be pursued separately and deserves further discussion.

5 Conclusion

In this paper, it has been shown that Bangla does have subject-to-subject hyper-raising, giving evidence through several tests from the language like island constraints, quantifier raising, idioms, lack of WCO effects and so on. It is further proven that the embedded clause is a finite CP and not anything smaller than that. Although all this evidence proves conclusively that the movement seen in examples like (3) in Bangla is an instance of true hyper-raising, we see that there is a crucial difference between Bangla hyper-raising and the ones seen in other languages. The absence of a non-overt or phonologically null complementizer head (IC *je*) inside the Bangla hyper-raising embedded clause is similar to what is observed in a *wh*-construction in English. The analysis follows Obata (2018) to explain this 'apparent' absence of C head in Bangla hyper-raising constructions and show through labeling that the external pair merge of C to T will give us the solution.

Acknowledgements

This paper has greatly benefitted from the guidance and comments of Tanmoy Bhattacharya, Sakshi Bhatia, Paroma Sanyal, Deepak Alok, Ojaswee Bhalla, alongside the numerous discussions that happened with my peers Adwitiya Dixit, Eshani Baishya, Jyoti Sharma at University of Delhi and the ICFL group at IIT-Delhi. I would also like to thank the audience of FASAL 15, especially Aswhini Deo for giving useful inputs regarding the topic.

References

Bayer, Josef. 1996. *Directionality and logical form: on the scope of focusing particles and wh-in-situ*. Dordrecht: Kluwer Academic Publishers.

Bayer, Josef. 1999. Final complementizers in hybrid languages. *Journal of Linguistics* 35(2). 233–271.

more thoroughly. But for our current purposes, we do not delve too deep into this and assume that the two verbs are fundamentally distinct where one allows hyper-raising and the other does not. See De Haan (2007) for grammaticalization of raising verbs in languages like English, German, and Dutch.

Bayer, Josef. 2001. Two grammars in one: Sentential complements and complementizers in bengali and other south asian languages. In P. Bhaskarorao & K. V. Subbarao (eds.), *The yearbook of south asian languages: Tokyo symposium on south asian languages-contact, convergence and typology*, 11–36. New Delhi: Sage Publications.

Bayer, Josef & Prabal Dasgupta. 2016. Emphatic topicalization and the structure of the left periphery: Evidence from german and bangla. *Syntax* 19. 309–353.

Bhattacharya, Tanmoy. 2001. The puzzle of bangla qc-internal clauses. *Snippets* 3. 6–7.

Bhattacharya, Tanmoy. 2002. Peripheral and clause-internal complementizers in bangla: A case for remnant movement. In *Proceedings of western conference in linguistics*, vol. 12, 101–113.

Bhattacharya, Tanmoy. 2015. Copula-less nominal sentences and matrix-c0 clauses: A planar view of phrase structure. In Laura E. Aldridge & Caterina Paolazzi (eds.), *The ucl working papers in linguistics*, vol. 27, 1–20.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In David Michaels Roger Martin & Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of howard lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken hale: A life in language*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2008. On phases. In Carlos Otero Robert Freidin & Maria-Luisa Zubizaretta (eds.), *Foundational issues in linguistic theory*, 00. Cambridge, MA: MIT Press.

Chomsky, Noam. 2013. Problems of projection. *Lingua* 130. 33–49.

Chomsky, Noam. 2015. Problems of projection: Extensions. In *Structures, strategies and beyond*, 3–16. John Benjamins Publishing Company.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.

De Haan, Ferdinand. 2007. Raising as grammaticalization: The case of germanic seem-verbs. *Rivista di Linguistica* 19(1). 129–150.

Diercks, Michael, Cliff Mountjoy Venning & Kristen Hernandez. 2022. Agreeing and non-agreeing hyper-raising in luyia: Establishing the em-pirical landscape.

Epstein, Samuel D, Hisatsugu Kitahara & T Daniel Seely. 2016. Phase cancellation by external pair-merge of heads. *The Linguistics Review* 33.1. 87–102.

Ferreira, Marcos. 2000. *Argumentos nulos em português brasileiro.* Campinas: Univ. Estadual Campinas MA thesis.

Ferreira, Marcos. 2004. Hyperraising and null subjects in brazilian portuguese. In *Mit working papers in linguistics*, vol. 47, 57–85. Cambridge, MA: MITWPL.

Fong, Suzana. 2019. Proper movement through spec-cp: An argument from hyperraising in mongolian. *Glossa: a journal of general linguistics* 4(1).

Greeson, Daniel. 2023. English hyperraising it seems is mediated by the θ-criterion. Workshop on Locality in Processing, Theory and Acquisition at UPenn.

Greeson, Daniel. 2025. Rethinking the connection between (hyper)raising and evidentiality. *48th Penn Linguistics Conference (PLC 48)* 31. 49–58.

Halpert, Claire. 2012. *Argument licensing and agreement in zulu*: Massachusetts Institute of Technology dissertation.

Halpert, Claire. 2015. *Argument licensing and agreement*. Oxford University Press.

Halpert, Claire. 2019. Raising, unphased. *Natural Language & Linguistic Theory* 37(1). 123–165.

Holmberg, Anders, Aarti Nayudu & Michelle Sheehan. 2009. Three partial null-subject languages: a comparison of brazilian portuguese, finnish and marathi. *Studia linguistica* 63(1). 59–97.

Huang, C.-T. James. 2022. Finiteness opacity and chinese clausal architecture. In *New explorations in chinese theoretical syntax: Studies in honor of yen-hui audrey li*, 17–77. John Benjamins.

Lee, Tommy Tsz-Ming & Ka-Fai Yip. 2024. Hyperraising, evidentiality, and phase deactivation. *Natural Language & Linguistic Theory* 42(4). 1527–1578.

Martins, AM & Jairo Nunes. 2009. Syntactic change as chain reaction: the emergence of hyper-raising. *Crisma, P* 144–57.

Martins, Ana Maria & Jairo Nunes. 2005. Raising issues in brazilian and european portuguese. *Journal of Portuguese Linguistics* 4. 53–77.

Obata, Miki. 2018. Eliminating c-deletion in the syntax: Structure-building by merge. *Koganei Journal of the Humanities* 14. 1–14.

Singh, Uday Narayan. 1980. Bole: An unresolved problem in bengali syntax. *Indian Linguistics* 41. 188–195.

Tanaka, Hidekazu. 2002. Raising to object out of cp. *Linguistic Inquiry* 33(4). 637–652.

Ura, Hiroyuki. 1994. *Varieties of raising and the feature-based bare phrase structure theory.* Vol. 7 of MIT occasional papers in linguistics. Distributed by MIT Working Papers in Lingustics.

Zeller, Jochen. 2006. Raising out of finite cp in nguni: The case of fanele. *Southern African linguistics and applied language studies* 24(3). 255–275.

Zyman, Erik. 2017. P'urhepecha hyperraising to object: An argument for purely altruistic movement. *Proceedings of the Linguistic Society of America* 2. 53–1.

Zyman, Erik. 2023. Raising out of finite clauses (hyperraising). *Annual Review of Linguistics* 9. 29–48.

Pronominal enclitics on floating flags in Sindhi and Brahui: Making a case for language contact with Balochi

SAMOPRIYA BASU, *Carleton University, Ottawa, ON, Canada*
NIRMAL RIAZ, *Karachi Urban Lab, Karachi, Sindh, Pakistan*

ABSTRACT

The present paper is the first of a series of planned studies looking at weak pronouns in Indo-Aryan and language contact centered around Balochistan. We introduce the weak pronoun inventories in Brahui, Balochi and Sindhi, discuss their historical origins, and provide a starting description of their main uses. In particular, the ability of floating case-flags to host objective pronominal enclitics in Sindhi and Brahui is emphasized and a likely origin of this construction in Balochi is suggested.

1 Introduction

The southern reaches of the Indo-Iranian frontier zone that runs roughly along the Sindh–Balochistan border in modern-day Pakistan is a region of great linguistic interest. Not only do the peripheries of the related but distinct Indo-Aryan and Iranic language families come into contact here through their representatives in Sindhi and Balochi, this region is also home to the divergent Dravidian language Brahui, famously isolated from the rest of its family in peninsular (and, in pockets, eastern) India. In addition, until the recent state-sponsored proliferation of Urdu throughout Pakistan, Sindhi served as a regional lingua franca and many border communities of Baloch and Brahui tribespeople speak it at near-native levels (Nazir Shakir Brahui, p.c.). Indeed, there are many Sindhi (and Saraiki) families that claim Baloch ancestry, and, conversely, several ethnic Baloch communities who speak a western variety of Sindhi called J̌aḍgālī or J̌aṭkī[1] as their mother-tongue. As such, one may naturally expect for there to be a significant amount of language contact in this region, yet the contact scenario here remains largely unexplored besides anecdotal comments here and there. The present study is an attempt to remedy this neglect and build toward a description of the extent of language contact among Sindhi, Balochi and Brahui, focusing first on the use of pronominal enclitics.

While Brahui is the only Dravidian language we know of to make use of enclitic weak pronouns side-by-side with full pronouns, Sindhi and Balochi are not as unique in their

[1]The convention followed in spelling non-English linguo- and ethnonyms here is as follows: For large languages or ethnic groups, standard spellings as have appeared previously in the literature are used, thus, *Balochi, Brahui, Sindhi* and *Saraiki* instead of *Balōčī, B(i)rāʔvī, Sindʰī* and *Sarāikī*. For smaller languages and dialects, however, a modified IAST transcription is used, thus, *J̌aḍgālī, Sulaymānī, Lāṛī, J̌ahlāvānī*, and so on. These are likely to be unfamiliar to many readers and risk being misinterpreted without diacritics, e.g., whether the "d" in an undiacriticized *Jadgali* is apical or laminal, or if the vowels are long or short. For languages with standard English names that differ from autoglossonyms, the English names are used, hence, *Persian* and *Bengali* rather than *Farsi* and *Bangla*.

respective families. Besides Sindhi, other western Indo-Aryan languages such as neighboring Saraiki and many western varieties of the Hindko-Punjabi continuum have sets of weak pronouns (enclitic or suffixal) as do Kashmiri and neighboring dialects. Within Iranic, weak pronouns are even more common, in fact, it is the lack of pronominal clitics that is outside the norm there, though the precise morphosyntax of their usage varies a great deal across languages and dialects, indeed, grammaticalizing to tense-aspect-triggered agreement markers in some cases. (West) Iranic pronominal enclitics have been studied in quite some detail, both from a comparative typological perspective (Mohammadirad, 2020) and their diachrony (Korn, 2009), as well as more probing formal analyses for specific languages, although the situation for Balochi in particular remains less explored (but see Dabir-Moghaddam (2008)). In contrast, weak pronouns in Indo-Aryan haven't received as much attention, except Kashmiri (Verbeke, 2017). They are mostly relegated to brief mentions in descriptions though some specialized studies have been inchoated (Butt, 2007; Wali & Koul, 1994; Jegorova, 1966; Khubchandani, 1969; Hook, 1987; Varyani, 1976/1977).

Despite the existence of cognate weak pronouns in languages like Kashmiri and Saraiki, we shall show in this paper that there is at least one usage of the Sindhi (objective) enclitic pronouns which is alien to other Indo-Aryan languages, but finds parallel in Brahui and Balochi. The rest of this paper is organized as follows: In §2, we begin with a working definition of what we mean by *weak pronouns* or *pronominal clitics & affixes*, and discuss their origins in the three languages under purview. The outlook is historical-comparative and fairly extensive as previous studies on weak pronouns in Indo-Aryan languages have glossed over this matter. We show that the systems have largely internal geneses and not simply borrowing of morphemes among neighboring languages. This section also provides brief language profiles to situate the study in context. §3 then describes the main uses of weak pronouns which are also shared with other western Indo-Aryan and West Iranic languages. §4 describes one specific usage of weak pronouns that set apart Sindhi from other Indo-Aryan languages and puts it closer typologically to Brahui and Balochi. We cautiously suggest a likely pathway wherein a (historical) calquing of a seemingly unremarkable Balochi construction may have led to the clitic-hosting floating flags in Sindhi and Brahui. Finally, §5 concludes the paper with discussion and exposition of limitations of the present study inviting further research in this area.

2 Weak pronouns — Form and genesis

By *weak pronouns*, we shall mean personal pronouns or pronominal possessors that are unaccented and morphologically bound, i.e., they cannot stand on their own and require (accented) hosts. These weak pronouns are, crucially, not agreement indexes, except when they have secondarily grammaticalized as such, in which case they are no longer weak pronouns synchronically speaking. Weak pronouns substitute for full pronouns in discourse and hence are usually, but not necessarily, mutually exclusive. They may be affixal or clitic. We do not dwell here on a precise criterion to disambiguate clitics from affixes, but loosely adopt the heuristic (though not every aspect of the definition) of Haspelmath (2023) that

affixes are selective of their hosts, while clitics are promiscuous, at least, on a surface level. As a consequence, both series of weak pronouns of Kashmiri are suffixes in our view, at odds with Wali & Koul (1994), who call them "clitics," but in agreement with Verbeke (2017), see wherein for more discussion on these definitional matters. For Balochi, Brahui and the objective weak pronouns (cf. §2.3) of Sindhi, the same criterion requires analyzing them as enclitics, while the ergative weak pronouns of Sindhi are suffixes.

2.1 Origin of pronominal enclitics in Brahui

Brahui is a divergent Dravidian language spoken exclaved away from its relatives, primarily in western Pakistan (Elfenbein, 1998), with smaller pockets in southern Afghanistan and southeastern Iran (Šeybānīfard & Zaršenās, 2020). Brahui speakers (at least males) are mostly bilingual with Balochi and possibly also in other languages depending on their location. Even in diaspora, Brahui is being passed down to children (Basu: own observation in Vancouver, Canada) and is not in any way threatened. Although it has traditionally (Krishnamurti, 2003) been classified in a "North Dravidian" node along with the closely related pair of Kurukh and Malto of eastern India, there is growing consensus among Dravidianists (Basu, 2024; Kobayashi, 2020/2021; McAlpin, 2003) that this classification is unwarranted and that Brahui forms its own subgroup in the family.

Elfenbein (1998) reports minimal dialectal differentiation in Brahui, but the most significant dialectal isogloss is reported to concern pronominal clitics. Ali & Kobayashi (2024) confirm dialectal faultlines along the use of pronominal clitics. Table 1[2] lists the pronominal enclitics of Brahui along with their corresponding full pronominal (oblique[3]) bases. Like most languages of the region, there are no dedicated third-person full pronouns in Brahui, instead, the distal demonstratives are used pronominally, but distinct third-person pronominal enclitics that cannot function as demonstratives or determiners do exist.

Table 1: Brahui pronominal enclitics and independent pronominal bases.

	enclitic	full pronominal base		enclitic	full pronominal base
1.SG	*=ka*	*kan-*	1.PL	(*=nan*)	*nan-*
2.SG	*=nā* ∼ *=nē*	*n(ē)-*	2.PL	(*=num*)	*num-*
3.SG	*=ta(n)* ∼ *=tē*	—	3.PL	*=tā*	—

The first- and second-person plural enclitics are reported by Andronov (1971) from Bray (1909), but the latter author explicitly notes they are only used in "some dialects" with examples from the speech of the "Nicharis." Later studies do not replicate these, which is why they have been parenthesized here. According to Elfenbein (1998), the first-

[2]Glossing conventions and abbreviations in examples adhere to the Leipzig Glossing Rules. Clitics are attached to their hosts by "=" and affixes by "-" as conventional. For consistency, we have taken the liberty of modifying — thus uniformizing — transcriptions of examples taken from other publications without distracting the reading with a warning each time.

[3]This caveat is only relevant for the 1.SG where the oblique stem *kan-* differs from the nominative *ī*.

and second-person singular enclitics only occur in the J̌ahlāwānī dialect. Ali & Kobayashi (2024) restrict only the first-person singular to J̌ahlāwānī. In their text corpus, no instance of *=ka* occurs,[4] but *=nā* ∼ *=nē* is amply attested. Notably, in the Iranian Brahui texts published by Barjasteh Delforooz (2020), a few occurrences of *=ka* do occur, showing its usage is alive in that variety, at least in possessive function. The third-person enclitics are abundantly used in all dialects.

The Brahui pronominal enclitics of the first- and second-persons are transparently derived from the corresponding pronominal stems. The origin of the far more commonplace third-person clitics are more mysterious. Already Bray (1909) had suggested derivation from the reflexive pronominal stem *tēn-*. Particularly relevant here is that Brahui shows an abnormal vowel in the form of the reflexive. From all other Dravidian languages, one obtains a reconstructable reflexive base **tan-* (NOM **tān*) in the singular and **tam-* (NOM **tām*) in the plural (Krishnamurti, 2003). The singular enclitic *=ta(n)* could, then, reflect an archaism in the modern language carrying the implication that grammaticalization of the expected base **tan-* as a pronominal enclitic happened before the irregular vowel-change. In fact, the existence of the variant *=tē* may even imply grammaticalization while the vowel-change was in effect. This still, however, leaves the plural enclitic *=tā* in need of an explanation.[5] One possibility is that it comes from the genitive-plural ending *-tā* latched onto the singular clitic, thus, older **=ta-tā* =ENCL.3.SG-GEN.PL which got haplologized to *=tā*, becoming formally homophonous with the bare genitive-plural ending. An analogous coinage of plural pronominal enclitics did take place in much of Iranic — though, notably, not including Balochi — at the onset of the Middle Iranic period as we shall see shortly.

2.2 Origin of pronominal enclitics in Balochi

Balochi is an Iranic language spoken in all the areas Brahui is, i.e., in a continuous band across western and southwestern Pakistan, southeastern Iran and southern Afghanistan, as well as a small outlying population in the Merv oasis of Türkmenistan.[6] Actually, the Brahui-speaking areas are enclaved within the larger Balochi-speaking zone, though Brahui is the predominant language in some areas within that zone, e.g., in northern Balochistan close to Quetta (Maryam Jamali, p.c.). Unlike Brahui, Balochi exhibits incredible dialectal variation throughout its vast home-range. The differences are so great across all facets of phonetics, morphology and syntax that it cannot be compactly summarized here (nor is it yet adequately documented). Traditionally, up to six dialect groups are recognized (Elfenbein, 1989), three of which — Makurānī, Raχšānī and Sulaymānī — are common knowledge among speakers themselves. But one can easily make a case for a much more fine-grained

[4]Brahui native-speaker and linguist Dr. Liaquat Ali confirms (p.c.) that *=ka* does exist in the Brahui of Pakistan despite its absence from his co-edited text corpus compiled from published stories.

[5]A derivation from the proximal demonstrative base *dā-* seems unlikely. Typically, it is not the proximal, but the distal demonstrative in *ō-* that is more commonly used pronominally, and, moreover, this choice makes it no easier to etymologize the plural in any case.

[6]There is also a significant Baloch presence in Oman (Collett, 1983), and, reportedly, smaller enclaves in other parts of Arabia and in eastern Africa.

dialectological analysis as done by Elfenbein (1980). The distinct far-western variety of the Koroš (Nourzaei et al., 2015) in southern Iran also deserves mention, as does Southern Bashkardi (Barbera, 2023), probably the closest relative of Balochi.

The internal classification of Iranic is fraught with unresolved complications, and especially so as far as the position of Balochi is concerned. The frequent reference in the literature to Balochi as "Northwest Iranic" belies the immense uncertainty surrounding not just the classification of Balochi as part of that group, but indeed the existence of such a genetic grouping in the first place. Nonetheless, the precise dialectological position of Balochi within Iranic will not matter for us here.

The Balochi pronominal enclitics are, for the most part, inherited from Proto-(Indo-)Iranic, but with significant changes to the syntax of their use. They are, thus, not recent innovations from full pronouns as in Brahui. In this, Balochi keeps with most other Iranic languages, where, too, the historical pronominal enclitics survive to varying extents.[7] However, one significant point of archaism in Balochi compared to most other western Iranic languages is that Balochi retains the inherited plurals which have been replaced by secondary pluralization of the singular enclitics by reflexes of the Old (Indo-)Iranic *a*-stem genitive-plural ending *-*ānām* in languages like Persian, Kurdish et c. Of course, -*ān* had already become the default oblique-plural marker in much of the family by the Middle Iranic period, including probably in Balochi, so, the extension to the clitic pronouns, which are necessarily oblique, is not surprising. Here we see a typological parallel to the situation hypothesized for the Brahui 3.PL enclitic pronoun in §2.1, but we have no reason yet to ascribe the development to contact of any sort. If anything, the exception of Balochi from this development within Iranic speaks against looking at language contact for an explanation.

This innovation began before attested Middle Iranic, as shown already by Middle Persian and Parthian, and progressed partway into eastern Iranic languages of the central steppes, including Sogdian, Chorasmian, Bactrian and Pashto.[8] The spread was not, however, as thorough as in Iran, and steppe languages like Ossetic and Hotanese conservatively keep the old paradigm, as do Ormuri and, partly, the Pamiri languages. The cases of Parachi and Yaghnobi are more complicated (and unclear).

Tables 2 and 3, respectively, show pronominal clitics across some Balochi dialects and in a few other western Iranic languages along with Avestan and Old Persian for historical reference. As mentioned before, note that Balochi inherits (cognates of) the plural clitics seen in Avestan, while in most other western languages, they have been supplanted by secondary plurals of the singular clitics. Unfortunately, the 1.PL and 2.PL clitics are not attested in the Old Persian corpus. We assume that the original plural enclitics were still preserved at the time and that the replacement by secondary plurals happened later. This seems true because Korn (2009) tables that =*n* existed for the 1.PL enclitic as a rare variant

[7]The languages in Turkic-dominated regions (viz., Zazaki and Kurmanji Kurdish in Turkey, and Tat and Northern Talysh in Azerbaijan) have largely lost the inherited proniminal clitic system; likewise Yidgha in far northern Pakistan whose speakers are fast shifting to Indo-Aryan Khowar (Haroon ur-Rasheed, p.c.).

[8]The case of Pashto is somewhat dubious. The syntax of Pashto pronominal enclitics is very conservative, but some of their forms are difficult to etymologize. We will not have reason to address Pashto in this paper.

alongside more common *=m-ān* in early Middle Persian implying that the two strategies were in competition centuries after the Achæmenid period.

Table 2: A sampler of Balochi pronominal clitics across dialect-groups.

	Southern Bashkardi	Korošī	Makurānī	Raχšānī	Avestan[9]
1.SG	*(=o)m=*	*=om*	*=un*	*=un*	*=mē*
2.SG	*(=e)t=*	*=et*	*=it*	(*=it*)	*=tē*
3.SG	*(=e)h=*	*=ī*	*=ī*	*=ī*	*=hē* ∼ *=šē*[10]
1.PL	*(=a)n=*	*=en*	*=in*	(*=in*)	*=nō*
2.PL	*(=o)χ=*	*=ō*	*=ō*	(*=ū*)	*=wō*
3.PL	*(=e)š=*	*=eš*	*=iš*	*=iš*	(OPrs. *=šām*)

Table 3: A sampler of western Iranic pronominal clitics.

	Gīlakī	Sorani Kurdish[11]	Zoroastrian Dari	Mīnābī	Persian	Old Persian
1.SG	*=əm*	*=m*	*(=)m=*	*=om*	*=am*	*=maiy*
2.SG	*=ət*	*=t*	*(=)d=*	*=et*	*=at*	*=taiy*
3.SG	*=əš*	*=y* (< **=haiy*)	*(=)š=*	*=i(š)*	*=aš*	*=šaiy*
1-PL	*=əm-an*	*=m-ân*	*(=)m-o=*	*=m-ån*	*=am-ån*	**=na*
2-PL	*=ət-an*	*=t-ân*	*(=)d-o=*	*=t-ån*	*=at-ån*	**=wa*
3-PL	*=əš-an*	*=y-ân*	*(=)š-o=*	*=š-ån*	*=aš-ån*	*=šām*

Within Balochi, there seems to be a consensus in the literature (with which we impressionistically agree) that the more west a variety is, the more productive is the use of pronominal clitics (Jahani & Korn, 2009), but we are not aware of any precise survey. In the far western variety Koroshi, clitics of all persons and numbers seem to be in use (Nourzaei et al., 2015). For the dialects of the Iranian Makurān coast, at least in the dialectal basis of what Jahani (2019) calls "Modern Standard Balochi," likewise, all clitics are in use. However, in the Afro-Balochi variety of coastal Iran studied by Korn & Nourzaei (2018, 2019), only the first and third-person enclitics are mentioned. In Sulaymānī Balochi (Elfenbein's "East Hill Balochi"), while older texts (Dames, 1922) attest at least the third-person clitics, we were unable to elicit any example from the modern speaker one of us is working with. The Balochi dialects of Afghanistan and Türkmenistan are (self-)identified within Raχšānī.

[9]Only the Younger Avestan GEN/DAT clitics are given here. Avestan and Old Persian (and Sanskrit) have separate ACC clitics as well, most completely in Elder Avestan, the cognates of which are reflexed distinctly from the GEN/DAT clitics in (at least some chronolects of) the eastern Iranic language Sogdian (Yosida, 2022).

[10]The variants here are phonologically determined according to the famous RUKI rule in Indo-Iranic and Balto-Slavic languages. In several later languages — already in Old Persian — the *š*-variants got generalized, but not everywhere, e.g., Xūrī, Kurdish and Balochi.

[11]The Akoi sub-variety of Central Kurdish seemingly retains the old first-person plural enclitic as *=nâ*, though the other two plurals are as in Standard Sorani (data in Meihami & Barzanji (2025)).

In his brief sketch of the former, Nawata (1981) only mentions third-person clitics. The dialect of the Merv oasis in Türkmenistan has been far better documented, where the latest description of Axenov (2006) reports only the third-person clitics in use. Sokolov (1956), half a century before Axenov, found also a first-person singular *=un*, but it was already very rare ("očenj redko") in his time. A full paradigm for this dialect cluster can, nonetheless, be salvaged from other sources (Jahani & Korn, 2009), but the exact distribution within the Raχšānī zone, as indeed elsewhere, is in need of precise mapping.

2.3 Origin of pronominal enclitics and suffixes in Sindhi

Though perhaps not to the same extent as Balochi, Sindhi too exhibits substantial dialectal variation within its territory.[12] Sindhi proper is spoken in the Pakistani province of Sindh, but the language spreads into the Rann of Kutch of northwestern Gujarat, India, as Kutchi (Kačč^{h}ī).[13] Over on the western side, the Lasī dialect of Sindhi spills over into southeastern Balochistan, and continues in patches as J̌aḍgālī throughout coastal Balochistan and into eastern Iran (Barjasteh Delforooz, 2008). Within Sindh, Jetley (1964) recognized four dialects (and Kutchi outside of Sindh): the Sirōlī dialect of northern Sindh, the Vičōlī dialect of central Sindh, which has been referred to as "standard" by Jetley (1964) and Khubchandani (2007), the Lāṛī dialect of the Indus delta, including the city of Karachi, and the Lasī dialect of Las Bela covering western Sindh and southeastern Balochistan.

Jetley (1964) also considers ^{h}aṭkī spoken in the Thar desert a dialect of Sindhi, but despite at least one crucial sound-change in common (*tr, dr* > *ṭ, ḍ*), a quick look at the data presented by Bhawnani (1979) suggests it is probably linguistically closer to Marwari.

In addition to J̌aḍgālī, another outlier Sindhi variety whose precise genealogical position within the Sindhi-continuum is difficult to ascertain without detailed study is Luwātī or K^{h}ōǰkī in Oman. A preliminary phonological sketch of this variety is by Salman & Kharusi (2012). Finally, there is Kholosi in southern Iran, which also appears to be "Sindhic," but separated from the Sindhi-continuum fairly early. It cannot be labeled a dialect of Sindhi, at least not in the current form. This language was discovered to scholarship by Anonby & Bahmani Hassan (2016) and currently being extensively documented (Rezaei, 2020; Arora, 2020–, 2022; Nourzaei, 2024, in prep.).

Despite promising early start during the British Rāǰ, Sindhi is the least studied of all major Indo-Aryan languages. Most primary studies are over half a century old at this point, if not more, and newer sketches are largely recycled off of those. For our purposes, weak pronouns (in the Vičōlī dialect) have been the focus of dedicated studies by Jegorova (1964);

[12] Of course, in the case of both languages, what dialectal variation is included under standard language labels and what excluded is largely a socio-cultural matter, not strictly linguistic.

[13] Kutchi lacks the characteristic Sindhi change of *tr, dr* > *ṭ(r), ḍ(r)*, thus resembling Gujarati in this one trait at least. Preliminary data on the language also suggests that it has developed a clusivity contrast akin to Gujarati (Gujarati INCL *āpṇe* v. EXCL *a̤me*, Kutchi INCL *pã̄* v. EXCL *asī̃*), but this is still just a conjecture at this stage, and perhaps not in all dialects. The Jadeja dialect described by Mukherjee (1992) has no clusivity identified. Nonetheless, Kutchi is unambiguously closer to Sindhi in morphology and lexicon than to Gujarati, and mostly intelligible to Sindhi speakers (Riaz: own observation).

Khubchandani (1969); Varyani (1976/1977), and also the grammar of Jetley (1964). In addition, they find brief mentions elsewhere, e.g., in standard grammars and sketches like Jegorova (1966); Zograf & Kogan (2011); Khubchandani (2007). For their use in the 19th c. and in compositions prior to that, see Trumpp (1872). Much of this older literature, however, is formulaic in the sense that paradigms are filled without a whole lot of analysis of the syntax and pragmatics of use, though occasionally gaps in paradigms are noted. Our elicited data is from the native Lāṛī dialect of one of the authors (Riaz) as spoken in Karachi. Lāṛī differs from Vičōlī in a few respects, such as the direct form of the first-person singular pronoun:[14] Lāṛī *$\widetilde{\hat{a}\mu}$* (< Sanskrit 1.SG.NOM *ahám*) against Vičōlī *mə̃* (developed from the Sanskrit oblique stem *ma-*). More specific morpho-syntactic differences haven't yet been identified, but they do not impede mutual intelligibility (Riaz: own observation).

Sindhi, like other Indo-Aryan languages with comparable morphology and unlike Iranic languages,[15] has at least two distinct sets of weak pronominal morphemes, though with some formal homophony across the paradigms. We call them *objective* and *ergative* weak pronouns and leave justification of these labels for the next section, where their uses are demonstrated. To these two may be added a *possessive* set used primarily with kinship nouns and partly similar to the objective clitics. The objective set in Sindhi is clitic, while the ergative set is affixal. We list the weak pronouns of Sindhi, along with comparanda from other Indo-Aryan languages in Table 4 below.

Table 4: Western Indo-Aryan OBJ and ERG weak pronouns compared to Sanskrit and Prakrit DAT/GEN pronominal clitics. Sanskrit and Prakrit ACC clitics not shown.

	Sindhi	Saraiki	Kashmiri	Poguli	Prakrit	Sanskrit
1.SG.OBJ	*=m*	*=m*	*-m*	*-m*	*=mē*	*=mē*
1.SG.ERG	*-m*	*-m*	*-m*	*-m*	×	×
2.SG.OBJ	*∅ ∼ =y*	*=(h)ĩ*	*-i*	*-t*	*=δē ∼ =tē*	*=tē*
2.SG.ERG	*∅ ∼ -y*	*-õ*	*-t*	*-t*	×	×
3.SG.OBJ	*=s*	*=s*	*-s*	*-s(ā)*	*=sē*	(*=ēna-*)
3.SG.ERG	*-ĩ*	×	*-n*	*-ni, -n^{y}e*	×	×
1.PL.OBJ	*=(h)ũ*	*=sē*	×	*-n*	*=ṇō*	*=nah*
1.PL.ERG	*-sĩ ∼ -sũ-*	*-sē*	×	*-n*	×	×
2.PL.OBJ	*=v*	*=vē*	*-v*	*-v*	*=vō*	*=vah*
2.PL.ERG	*-v*	*-vē*	*-v*	*-v*	×	×
3.PL.OBJ	*=n*	*=nē*	*-k^{h}*	*-nan, -n^{y}ən*	*=(ē)na-*	*=ēnā-*
3.PL.ERG	*-ũ*	×	*-k^{h} ∼ -h-*	*-ne, -n^{y}e*	×	×

The objective sets have fairly clear origins.[16] In all languages, they mostly continue

[14] Lasī and Kutchi agree with Lāṛī in having *$\widetilde{\hat{a}\mu}$*, while Sirōlī goes with Vičōli and has *mə̃*, cf. Sarāīkī *mæ̃*.

[15] Ossetic has multiple series of related pronominal clitics, but the system works very differently from the western Iranic enclitics we are concerned with in this paper.

[16] Among the Kholosi pronominal enclitics reported by Nourzaei (2024), which are objective and possessive, only the third-person forms are directly cognate to the rest of western Indo-Aryan.

the Sanskrit and Prakrit GEN/DAT enclitics, though some points need clarifying. All of the languages considered unambiguously reflex the 1.SG and 2.PL enclitics. Also reflexed consistently is the Prakrit 3.SG enclitic *=sē*, which doesn't occur in Sanskrit and Pali. According to some scholars, like Bubeník (1992), it is a conservatism in the Prakrits and later Indo-Aryan (lost by chance in dialects of Sanskrit attested), formally being cognate to Elder Avestan *=hōi̯* and Younger Avestan *=hē*, and further to Hittite *=ŝe*, Luwian *=si* and Greek *=hoi̯* (all 3.SG.DAT). However, a more realistic etymology is of secondary genesis in Middle Indo-Aryan, coinciding only by accident with the expected inherited cognate, as reported by Kümmel (to appear) following Scheller (1967) and von Hinüber (2001).

Although strengthened in various ways, the second-person singlular objective enclitics transparently descend from Sanskrit *=tē*, via the intervocalically weakened Prakrit *=δē*. In the Prakrits, there was positional allomorphy between *=δē* and *=tē* owing to well-known Middle Indo-Aryan sound changes, and this pattern likely continued into the New Indo-Aryan period, followed by leveling one way or the other. Poguli, with its conservative-looking *-t*, may be the single language that favored *=tē* in the leveling process, but it may also be brought over from the ergative paradigm (compare the Kashmiri ergative).

For the first-person plural, Kashmiri notably has a hole in its paradigm there, and Saraiki (as well as western Punjabi and Hindko varieties that have weak pronouns) has brought over the ergative suffix into objective use. Poguli is the only one that retains a reflex of Sanskrit *=nah̐*, while Sindhi's *=(h)ũ̃* has, in all likelihood, arisen from the Sanskrit first-person plural oblique stem *asmá-*; cf. attested Apabhramsha 1.PL.DAT/GEN *amha*.

Finally, the widespread *n*-forms of the third-person plural objective weak pronouns have long been known to continue the Sanskrit enclitic pronominal paradigm of *=ēna-*; already Trumpp (1872). They continue (contracted) in Pali and the Prakrits as *=na-*. In the older stages, it has a full paradigm (though defectively attested) across cases, numbers and genders. In the modern languages, however, it only survives in plural usage, probably because the singular slot had competition from *=sē*. The gender distinction uniquely maintained in Poguli no doubt carries forward the older system, but without further documentation of Poguli, nothing certain can be said of their exact history. The success of the *n*-forms here may have played a rôle in the ouster of the expected *n*-forms for the first-person plural in all languages save Poguli. Kashmiri *-k^h* is unexplained, although, notably, its allomorph in ergative use when followed by an objective suffix is *-h-*.

Since the objective paradigms of weak pronouns are well-explained by the Old and Middle Indo-Aryan GEN/DAT enclitics, it may be tempting to try and derive the ergative paradigms from the ACC enclitics. However, this doesn't quite work because, being very similar formally, they would yield identical outcomes in later languages.[17] Indeed, the *s*-forms of the first-person plural ergative suffix in both Sindhi and Saraiki have to have origins

[17]In fact, the accusative and genitive-dative paradigms started being partly leveled out in the early period itself. Only Elder Avestan maintained a full paradigm; Sanskrit and even Younger Avestan merged GEN/DAT and ACC for the plural enclitic pronouns. Besides, an origin of the modern ergative suffixes from the old accusative set would be highly unexpected in any case as ergativity itself arose (or at least fully established itself in the grammatical system) only toward the end of the Middle Indo-Aryan stage.

in the full pronouns, cf. 1.PL.DIR *asĩ*, 1.PL.OBL *asã*. The exact syntactic condition that led to *asĩ* becoming bound is not clear, nor is the fact that the shape of the suffix *-sĩ* resembles the direct *asĩ* rather than the oblique *asã*, which functions as the full ergative pronoun. The origin of the third-person ergative suffixes *-ĩ* and *-ũ* are even more perplexing.

Finally, Sindhi has a partial series of suffixal pronominal possessors having no equivalent in neighboring languages like Saraiki and Punjabi, nor in Kashmiri or Poguli. These partly resemble the objective clitics and mostly used to possess certain close kinship terms, though in the language of a couple generations ago, possession of a small class of other inalienables, like body-parts, was also possible (Khubchandani, 1969).[18] Possessive pronominal suffixes that exclusively possess close kinship items are an areal feature of the Indo-Aryan and Nuristani languages of the Hindukush, as noted by di Carlo (2008/2011), though certainly not all of them, e.g., not Khowar or Prasun, but there is reason to believe those are later innovations in mutual contact, unrelated to the Sindhi morphemes, at least directly.[19]

3 Functions of weak pronouns

Weak pronouns fill oblique rôles in Brahui, Balochi and Sindhi (and elsewhere). This means they can mark a variety of different kinds of objects, occasionally adjuncts, possessors, and also ergative agents if the language permits, but they can never stand in for direct-case full pronouns. In Brahui, alignment is always nominative-accusative, so, no agent or subject can be marked by pronominal clitics. In Sindhi and Balochi, ergative agents can be. However, cross-dialectally, Balochi shows such incredible diversity in alignment types, that uniformly applying this general principle there becomes tricky. Recall that, barring some minor phonological allomorphy, in Balochi and Brahui, there is just one set of weak pronouns each, while Sindhi has three, though with some formal overlap across paradigms.

In all three of Sindhi, Balochi and Brahui, weak pronominal when attached to nouns serve as their possessors. In Sindhi, this is more restricted than the other two as only certain kinds of inalienable possessions can be directly possessed by weak pronouns. In our data from the Lāṛī dialect, there are only cases of possession of close kinship terms, though older data in the literature also show possession of certain body parts. In Brahui (1) and Balochi (3), there seems to be no restriction on what is possessable, though in many varieties only the third-person clitic possessors are fully active.

(1) Brahui

a. Iranian Brahui (Barjasteh Delforooz, 2020, The parrot and the maina: 1)
dāsā, *rīš=ka* *pīun* *mass.*
now beard=1.SG.OBL white become.PRS.PFV.3.SG

[18] We interpret the difference between Khubchandani's (and earlier) description and ours as one of chronology, but it needn't be so. As stated before, Khubchandani's account was of his own Vičōlī dialect, while ours is of Riaz's Lāṛī dialect. A thorough comparison between these two (and other) Sindhi dialects with regards to the use of weak pronouns cannot be attempted at this point. (Kutchi seems to lack weak pronouns altogether.)

[19] The possessive suffixes of Pashai are used much more generally, not restricted to kinship terms, inalienables, or any other semantic class, though even there kinship nouns behave somewhat special (Lehr, 2014).

‘Now, my beard has turned white.’

b. Raχšānī Brahui (Kobayashi & Ali, 2022–, 12 (šōhan-nā rāg: 18’53))

pučč-ē hat-p-ēs-a, liχ-ē=nā
clothes-OBJ bring-NEG-PRS.2.SG-IPFV neck-OBJ = 2.SG.OBL
χal-ēv-a.
strike-PRS.1.SG-IPFV

‘(If) you don’t bring the clothes, I shall decapitate you (… strike your neck off).’

(2) shows that even in possessive phrases with seemingly lexicalized meaning, clitic-possession is valid.

(2) Sarāwānī Brahui (Ali & Kobayashi, 2024, khuriyā-nā bōḍ-ak: 4’56)

harāng ki ust=nā kašš-ē, čirrēng.
whither.REL COMP heart=2.SG.OBL pull-PRS.3.SG wander

‘Roam wherever you feel like (… your heart pulls).’

In Balochi, pronominal clitics are active more in Iranian dialects than elsewhere. Nonetheless, all dialects still seem to have at least the third-person clitics to whatever degree of productivity, some even the first-person ones. The second-person enclitics are the rarest (in contrast to Brahui where the first-person is rarest). The Koroshi example (3b) is interesting because it shows both clitic-possession and a clitic subject in one sentence.

(3) Balochi

a. Iranian Makurānī Balochi (Jahani, 2019, pg. 88)

zōr=on ča taī-g-ā gēš=ent.
strength=1.SG.OBL from 2.SG.GEN-PRED-OBL more=COP.PRS.3.SG

‘I am stronger than you (My strength is greater than yours).’

b. Koroshi (Nourzaei et al., 2015, pg. 56)

marō zahr=eš rētk-a mā χorāk=at.
today poison=3.PL.OBL pour.PST-PFV.PTCP into food=2.SG.OBL

‘They have poured poison into your food today.’

c. Afghanistani Raχšānī Balochi (Nawata, 1981, nakl)

zāg-ān=ē āi-rā gōr kurt-ant.
child-PL = 3.SG.OBL DIST.DEM.OBL-OBJ grave do.PST-3.PL

‘His children buried him.’

Jahani (2019) reports that pronominal clitics in Balochi cannot possess nouns marked by the oblique case suffix *-ā* except as objects of postpositions, which are all oblique-marked nouns, at least etymologically. This may be true for the Makurānī varieties she bases her standard on, but in Turkmenistani Raχšānī, at least, that seems allowed according to Axenov (2006), even productive.

In Sindhi, possessive weak pronouns sometimes trigger interesting stem-changes to the possessee. Kinship terms, at least, admit a stem-extension *-ņ-*, e.g., *bʰāu* 'brother', but *bʰāņŭ-s* 'his/her brother'; likewise, *sas* 'mother-in-law', but *sasŭņ-hē* 'your mother-in-law' (both examples from Lāṛī, but reproducible in Vičōlī and probably Lasī).[20] Since there's at least some variants of the possessive morphemes that don't fully coincide with the ergative or objective sets, we treat them as a distinct paradigm of suffixes. As the notation in (4) implies, our analysis sees the weak pronominal possessors as suffixes rather than clitics.

(4) Vičōlī Sindhi (Rahman, undated)

puṭŭ-m *ač*ʰ*-ē* *t*ʰ*-ō.*
son-1.SG.POSS come.IPFV-3.SG COP.PRS-M.SG

'My son comes.'

In all three languages, a possessive interpretation of pronominal clitics or suffixes is allowed by direct postposition to the possessee. External possession of the kind attested in western Punjabi and Saraiki (5) is generally not allowed in Balochi and Brahui. Punjabi and Saraiki, on the other hand, do not allow direct nominal hosts of their clitics. Attachment seems to be obligatorily to the finite verbal complex,[21] though there are some exceptions when finite verbs are ellipted as in (5a).

(5) Punjabic

a. Sargodha Punjabi (Bashir et al., 2019, from Wilson, 1999)
gʰar *kittʰē=nē?*
house where=3.PL.OBL
'Where is their house?'

b. Saraiki (Shackle, 1976, from Lashari, 1971)
d̄ūhē *dʰi.y-ã̄* *āi.y-ã̄* *hān=is.*
both daughter.OBL-F.PL come.PST-F.PL COP.PST.3.PL=3.SG.OBL
'Have both of his daughters come?'

In Brahui and Balochi, possessive use is entirely restricted to suffixing or encliticizing the possessor pronoun directly to the possessee, and this is mostly the case in Sindhi too.[22]

[20]Transcription of Sindhi is slightly modified from existing sources. In our experience, final *ultra-short* vowels are no longer pronounced except in western dialects spoken by the Baloch and Brahui. Sindhi is, thus, the latest (and last) Indo-Aryan language to lose old final vowels that were still alive well into the mid-20th c. In certain inflections, these vowels come back and we write them there with a breve.

[21]There is no study yet on the exact attachment patterns of pronominal enclitics in Punjabi and neighboring languages. In our cursory look, the obligatory host appears to be the first word of the prosodic domain of the finite verb, provided it exists. Thus, Butt (2007) cites an example ((2) in her paper) where it is the bound pre-verbal negator *nɛ=* to which the 3.SG enclitic *=s(u)* latches on to.

[22]Two instances in the corpus, one line apart in the same story (tuɣ: 90 & 91) and identical context, show that Brahui may require the (clitic) copula to be ordered before a clitic pronoun even if it means breaking up the possessive phrase: *aenō musiṭṭamīkō dē=ē=*$^{?}$*tē/nā* 'Today is his/your third day (waiting here).'

In the latter, post-verbal affixation with possessive interpretation is attested as well, but not in our data so far. The Balochi pattern is also what obtains in other western Iranic languages where enclitic possessors exist, but strikingly different to how the construction ancestral to this worked in Old and Middle Indo-Iranic, and still continued in Pashto. In the older stages, the GEN/DAT enclitics (and the ACC clitics as well) attached to whatever was the first constituent (the so-called *Wackernagel position*). In possessive constructions, attachment could be to the possessee itself, if the first constituent, or the possessee could follow it, or be completely disconnected.

Besides marking possessives, one of the chief uses of enclitic pronouns is marking oblique arguments and some kinds of adjuncts. Being a strictly nominative-accusative language, in Brahui, the subject or agent is never oblique and thus cannot be marked by a clitic pronoun, but it is the sole controller of verb-agreement. Objects, direct and indirect, do not trigger agreement, and may be cliticized to finite verbs as in (6).

(6) Brahui

a. J̌ahlāwānī Brahui (Elfenbein, 1998, 40a)
χal-k-us=ka!
strike-PST-2.SG = 1.SG.OBL
'You hit me!'

b. Raχšānī Brahui (Ali & Kobayashi, 2024, mēhr-anā kaedī-k: 61)
dā=tō kan-ā mehr=ē. num-ē
PROX.DEM = TOP 1.SG-GEN love = COP.PRS.3.SG 2.PL-OBJ
ti-f-ar-a=ta. padī d-ēv-a=ta.
give-NEG-PRS.1.SG-IPFV = 3.SG.OBL back take-PRS.1.SG-IPFV = 3.SG.OBL
'She is my love. I will not give her to y'all. I will take her back.'

Typically, once a participant has been introduced in a conversation, subsequent references are by enclitics (provided they exist) unless there is need to focus, in which case, full pronouns or demonstratives are used. In elicitation, at least, co-occurrence of a full pronoun and an enclitic with the same referent within a clause is strongly dispreferred (Liaquat Ali, p.c.), but we did find one rare instance in a publication, reproduced here as (7).

(7) Kalātī Brahui (Shazia et al., 2022, viii)
ō-t-ē iraɣ ēt-in-a=tā. yā
DIST.DEM-PL-OBJ bread give-PRS.1.PL-IPFV = 3.PL-OBL or
ma-t-av, ōχa pad čāe ēt-in-a=tā.
become-NEG.PST-3.SG that_much after tea give-PRS.1.PL-IPFV = 3.PL-OBL
'We shall give them bread. Or if not, after that much, we shall give them tea.'

Attachment to non-finite nominal verb-forms are more difficult to interpret since they may be arguments or possessors. In the examples in (8), alternative "possessor" reading of each sentence is provided in parentheses. Indeed, in (8a), that might be the only valid underlying interpretation since otherwise the clitic would have to stand for a subject.

(8) Brahui

a. Sarāwānī Brahui (Ali & Kobayashi, 2024, lāl malūk: 9'35)

kun-ing=nā=tō *mur mass.*

eat-INF = 2.SG.OBL = TOP far become.PRS.PFV.3.SG

'You are safe from being eaten (now). (Your being eaten is far for now.)'

b. Sarāwānī Brahui (Ali & Kobayashi, 2024, šōhan-nā rāg: 8'59)

bass-ur *hall-ing-kin=ta.*

come-PST.3.PL take-INF-BEN = 3.SG.OBL

'They came to arrest him. (They came for his taking.)'

Sindhi agrees with Brahui in that argument uses mostly require attachment to finite verbs. However, since Sindhi has ergative-absolutive alignment in the perfect tenses like many Indo-Aryan languages, besides (oblique) objects (9a), agents may also be marked by weak pronouns under the right tense-aspect conditions (9b). The two uses are, however, morphologically distinct. As mentioned in §2.3, Sindhi distinguishes objective pronominal enclitics from ergative pronominal suffixes. This also means that both can occur on verbs simultaneously, in which case, object clitics follow the ergative suffixes (9c).

(9) Sindhi

a. Lāṛī Sindhi

mũ̃=vaṭã̃ *kʰat pãhut=as.*

1.SG.OBL=from letter reach.PFV = 3.SG.OBJ

'He received a letter from me.'

b. Lāṛī Sindhi

kʰādʰ-ay, *yā kōnʰă kʰādʰ-ay?*

eat.PFV-2.SG.ERG or not eat.PFV-2.SG.ERG

'Have you eaten or not?'

c. Lāṛī Sindhi

sūf *ḏ̄in-am=ās.*

apple give.PFV-1.SG.ERG = 3.SG.OBJ

'I gave him/her an apple.'

The functional distinction between the ergative weak pronouns and objective weak pronouns is clear. Ergative weak pronouns stand in for strong pronouns as agents of transitive verbs in the perfect tenses, while objective weak pronouns are used for objects and adjuncts, including, as we shall see, postpositional arguments. The reason we interpret the ergative set as suffixes is because they necessarily attach to the finite verb, and, as far as we have checked, cannot be separated from it, while the objective set is more loosely bound accepting potential non-verbal hosts as well. In addition, the ergative set appears to be bound more tightly to the verbal stem compared to the objective set even if we compare them solely by

how they attach post-verbally. In Sindhi, like in Hindi, in ergative constructions, there is object-agreement with the verb unless otherwise blocked by a flagged object. Presence of an ergative suffix, interestingly, is also a trigger for blocking of agreement, but this doesn't amount to default agreement; agreement morphology is simply dropped.

(10) Sindhi

a. Lāṛī Sindhi

mɯ̃̄ *sūf* *kʰādʰ-ō.*
1.SG.OBL apple{M.SG} eat.PFV-M.SG

sūf *kʰādʰ-am.*
apple{M.SG} eat.PFV-1.SG.ERG

'I ate an apple.'

b. Lāṛī Sindhi

mɯ̃̄ *mākʰī* *kʰādʰ-ī.*
1.SG.OBL honey{F.SG} eat.PFV-F.SG

mākʰī *kʰādʰ-am.*
honey{F.SG} eat.PFV-1.SG.ERG

'I ate honey.'

Notice that the verb-forms in the second in each pair of sentences (10a & 10b) are identical, not indexing φ-features of the object, while in the absence of the ergative weak pronoun, we see clear object-agreement. As (11) shows, however, objective clitics do not suppress agreement morphology.

(11) Vičōlī Sindhi (Jetley, 1964, pg. 300)

hal-and-u=v.
go-FUT-3.M.SG = 2.PL.OBJ

'He will go with y'all.'

On the surface, in Balochi, it too looks like pronominal clitics marking oblique arguments attach to finite verbs at first glance. But careful probing shows that attachment is to the first element of the verbal complex, as in (12).

(12) Türkmenistani Raχšānī Balochi (Axenov, 2006, 88)

āy-rā *nazzīk-ēn* *āp-ā* *nišān=ī* *dāt.*
DIST.DEM-OBJ near-ATTR water-OBL sign = 3.SG.OBL give.PST.3.SG

'He showed him the nearby water.'

This is still different than the pattern common in some Iranic languages of Iran where attachment is to the first element of the verb phrase (Mohammadirad, 2020), but one has to

reserve judgment for all of Balochi until there is available better cross-dialectal data, especially for the dialects in Iran. Interestingly, in some of the latter, Dabir-Moghaddam (2008) presents data showing the clitics are grammaticalizing as agreement markers akin to what is seen in some varieties of Laki and Sorani Kurdish (Taghipour & Kahnemuyipour, 2019; Akkuş et al., 2025) and in the Pamiri languages.

In neither Balochi nor Brahui is there pronominal stacking on verbs the way it is possible in Sindhi and other Indo-Aryan languages like Saraiki and Kashmiri (though, of course, the stacked weak pronouns belong in different sets). Nor can pronominal clitics on verbs mark anything besides their direct and indirect objects in Brahui. However, Brahui does have a strategy to mark benefactive, comitative, allative and other pronominal adjuncts in a strategy shared with Sindhi that allows clitics to be hosted on case-flags. This is the subject of the next section.

4 Pronominal enclitics on floating flags

Without recourse to the theoretical complexities of a precise definition, in this paper, we use the term *flag* following Haspelmath (2019) for case-markers, whether they are analyzed as affixes or clitics or independent phonological words (adpositions), as long as they can license pronominal arguments. As we shall see, the morphemes in question are suffixes in Brahui and cannot stand on their own without a preceding host from which they cannot be split, clitics in Sindhi, and partially grammaticalized denominal postpositions in Balochi. *Flag* is, thus, a convenient cover term to refer to them all together, which we shall have need for. For Sindhi, we will also make use of *layering* terminology for the three diachronically[23] distinct *layers* of flags in Indo-Aryan languages identified by Masica (1993).

Unknown to other Indo-Aryan languages nearby with weak pronouns, in Sindhi, certain Layer 2 flags can host objective pronominal clitics and this is the only scenario when these flags do not need a substantival host preceding them. Thus, in (13), the second sentence has the same meaning as the first, but uses a pronominal enclitic on the dative flag =*kʰẽ*, which now needs no preceding host. Contrast, for example, Hindi =*kō*, which can never occur without a preceding host, or even Punjabi =*nũ* and Saraiki =*kũ* in languages with enclitic pronouns, but still do not support this construction.

(13) Lāṛī Sindhi

hună=kʰẽ *sūf* *ḏ̄in-am.*
DEM.DIST.SG.OBL = DAT apple give.PFV-1.SG.ERG

kʰẽ=s *sūf* *ḏ̄in-am.*
DAT = 3.SG.OBJ apple give.PFV-1.SG.ERG

[23] The distinction between layers is both diachronic and synchronically morphological, e.g., Layer 2 clitics show suspended affixation to (conjoined) noun phrases, while Layer 1 suffixes directly attach to the stem (and in languages like Hindi, effectively forming the stem). We don't go into any more discussion on this here as the layout is well-known to (at least) Indo-Aryanists from the influential book of Masica (1993), in Chapter 8 of which those unfamiliar may look.

‘I gave him/her an apple.’

Recall that these sentences are also equivalent in meaning to (9c), where the same objective clitic attaches to the verb. In that case, *(=)kʰē* doesn’t occur at all. The position of *kʰē=s* in the above sentence is not fixed and may be scrambled, but the crucial point is that it can occur utterance-initially.

The pragmatic consequences of word-order variation, or the difference between using *kʰē=s …ḏin-am* (13) versus *ḏin-am=ās* (9c) are not yet clear, nor indeed the discoursal frequency of use of each construction. Native-speaker intuition (Riaz) suggests these are equivalent, but subtle pragmatic differences, if any, are best studied from a diverse enough spoken corpus, which is currently non-existent for this language.

In our Sindhi data from the Lāṛī dialect, the most widely attested in this behavior is the dative *(=)kʰē*. Jetley (1964) and others note objective pronominal enclitics also attached to *(=)kʰɛ̃ā* ABL, *(=)sāṇ* COMM as well. The literary possessive *sand-* is unique in that its postpositional equivalent looks very different, *=ǰ-*.

(14) Vičōlī Sindhi (Jetley, 1964, pg. 320)

sāṇŭ=n *vaň!*
COMM = 3.PL.OBJ go.IMP.2.SG

‘Go with him!’

In Brahui, case-marking flags cannot stand without a preceding host. They are true suffixes. However, there is still a parallel to the Sindhi construction. Brahui uses what Ali & Kobayashi (2024) call a *dummy pronoun*, *ī-*, which serves as the host for nominal case-markers to which then attach the pronominal enclitics (15).

(15) Brahui

a. Sarāwānī Brahui (Ali & Kobayashi, 2024, lāl malūk: 17’06)
valdā dē tamm-ā *yā šām tamm-ā*, *harāṛē ki*
then day fall-PST.PFV.3.SG or night fall-PST.PFV.3.SG where.REL COMP
pā-s *bar-ēn-a* *ī-ā=nē.*
tell-SBJV.2.SG come-PRS.1.PL-IPFV DUMMY-ALL = 2.SG.OBL
‘Whether day breaks or night falls, wherever you tell us, we shall come to you.’

b. Sarāwānī Brahui (Ali & Kobayashi, 2024, lāl malūk: 17’06)
kasar-ā *īlum-k* *halk-ur* *ī-kin=ta*
road-ALL brother-NOM.PL take-PST.PFV.3.PL DUMMY-BEN = 3.SG.OBL
hullī-s.
horse-INDF
‘The brothers acquired a horse for him on the way.’

c. Sarāwānī Brahui (Ali & Kobayashi, 2024, šōhān-nā rāg: 10’37)

rāg-vālā *ambal=ham* *tudd=ē*
drama-PROPR friend=too together=COP.PRS.3.SG
ī-tō=tā.
DUMMY-COMM=3.PL.OBL
'The drama man was also with them.'

In one example, the dummy is attested with a full pronoun, but this is likely a secondary extension by analogy as this variety lacks first-person plural enclitics. It doesn't look like this pattern is at all widespread among native-speakers.

(16) Sarāwānī Brahui (Ali & Kobayashi, 2024, lāl malūk: 17'05)

nan-ā *laškar* *ǰang* *kē-k* *ī-tō* *(=)nan-ā.*
1.PL-GEN troop fight do-PRS.IPFV.3.SG DUMMY-COMM (=)1.PL-GEN

'Our husbands will scold us (… fight with us).'

Given that attachment of pronominal enclitics to flags detached from nominal cases isn't attested in other Indo-Aryan (nor Dravidian) languages, we speculate it may be a contact feature arising in the region, but details of this contact needs sketched out. Before suggesting a possible pathway, we briefly look at a surface parallel elsewhere in Dravidian. The family-wide situation in Iranic and Indo-Aryan was already briefly touched on in §2.

No Dravidian language other than Brahui has a system of pronominal clitics, but object-indexing on verbs is famously known in the Kūic[24] subgroup of the South-Central Dravidian languages. In these languages, the presence of any first- or second-person non-nominative clausal participant, including direct and indirect objects, possessors (17), experiencers and postpositional complements (even if ellipted), obligatorily triggers agreement on the clause-heading finite verb by a suffix Steever (1993) labels *transition particle* (TP).

(17) Pengo (Steever, 1993, from Burrow & Bhattacharya, 1970)

*āne*ŋ *nī* *keydi*ŋ *mrīn-ja-t-a*ŋ.
1.SG.NOM 2.SG.GEN hand press-TP-PST-1.SG

'I pressed your hand.'

Although one of the many allomorphs of this suffix is *-ta*, resembling the Brahui third-person pronominal enclitics, they index other persons, and Steever (1993) has shown that the present-day transition particle arose from contraction of a compound verb construction with **tar-/tā-* 'give to 1st or 2nd person' accounting well for the observed distribution.[25] Historically, this was the finite verb in the clause and, therefore, inflectional material today follow the transition particle. It was never a clitic to the entire verbal complex.

To explain the shared pattern in Sindhi and Brahui with regard to pronominal enclitics attaching to floating flags, we turn to Balochi. In Balochi, both pre- and postpositions

[24]This is the group made up of dialects of Kūi, Kūvi, Maṇḍa and Pengo, all spoken in eastern India.
[25]This verb continues in Brahui as *tir-/ēt-* but means 'give' without any person-restriction.

exist, but the postpositions are etymologically oblique-marked nouns while prepositions are an older layer mostly continuing Indo-Iranic Lokalpartikeln, as in other modern Iranic languages except Ossetic (Şahingöz & Basu, 2025). The general pattern is that Balochi varieties in Iran show more active use of prepositions while Pakistani varieties have more extensive use of postpositions.[26] In the Makurānī varieties represented in the standard of Jahani (2019), pronominal enclitics are allowed as objects of postpositions, but not of prepositions.[27] E.g., *poštā* is etymologically the oblique case of *pošt* 'back (body-part),' but used postpositionally, it means 'behind,' and governs a noun or pronoun in the genitive, and in (18), its complement is a pronominal enclitic postposed to it.

(18) Iranian Makurānī Balochi (Jahani, 2019, pg. 90)

pošt-ā=eš *čēr* *bay!*
back-OBL = 3.PL.OBL hide become.IMP.2.SG

'Hide behind them!'

Whatever the situation with prepositions may be, attachment to postpositions is simply a natural extension of the possessive usage since postpositions hosting pronominal clitics are etymologically (oblique) nouns possessed by their complements. In Sindhi, this is not so (at least not currently) since Layer 2 case-flags are purely functional, they have no lexical content of their own. In Brahui, this is even more so because case-markers are true suffixes there. Even for possession, clitics attach after case- and number-suffixes. Thus, it is in Balochi that postpositional attachment arises most naturally. Sindhi and Brahui possibly calqued this construction. In the case of Sindhi, oddly, detaching the otherwise bound case-flags from needed substantive hosts and in Brahui compensated by the use of a dummy pronoun to latch the case-endings to.

Of course, this diachronic trajectory is but hypothetical at this stage. There is no historical record of Brahui before the present-day, and its isolated position within Dravidian means that even comparative analyses are unlikely to yield any clue. We are more fortunate with Indo-Aryan, however, and there is need to examine pre-modern texts in Sindhi (and Saraiki) to determine how pronominal clitics, if attested, behaved historically. Also needed is more cross-dialectal data from both Balochi and Sindhi in order to clearly map out language-internal differences in weak pronoun usage.

5 Conclusion

In this paper, we have presented a preliminary study on the diachrony and usage of weak pronouns in Sindhi, Brahui and Balochi, and suggested plausible evidence of a micro-

[26] Native-speaker Maryam Jamali informs us that in her (southern) subvariety of the Sulaymānī dialect, even inherited prepositions like *až* 'from' have started being used postpositionally within the last 3 generations.

[27] In Koroshi, attachment to both postpositions and prepositions is attested. An example of the latter is given by Nourzaei et al. (2015) on pg. 35. However, the preposition in question, *az* 'from,' is borrowed from Persian, and whether this is truly prepositional hosting or special behavior attributable to the non-native origin of the host requires more data to judge.

Sprachbund centered at the Sindh–Balochistan border. Needless to say, this is hardly the last word on this topic and much more needs to be done. In particular, we have only pointed out in broad terms the main functions of weak pronouns in these languages. A thorough exploration of the pragmatics of usage of weak pronouns vis-à-vis their full counterparts remains a desideratum, as does an understanding of exact attachment parameters. This is especially true of Sindhi where multiple permutations of cliticization are possible with as-of-yet unknown difference in meaning. A sizable spoken corpus is sorely needed for such a study, however. Also needed is research on person-number restrictions on co-occurrence of the different weak pronoun sets. Indeed, in our estimation, some of the weak pronouns are falling out of use. This attenuation, no doubt, varies by dialect as well, and cross-dialectal variation needs to be taken into account for all languages, especially Balochi, where it is significant, but also Sindhi, where, in spite of a little more uniformity otherwise, one of the main variables of regionalism is in the use of weak pronouns.

The problem of language contact also needs more attention. One of the limitations of the present study is that Balochi, Sindhi and Brahui data from different dialects are pooled together to make an argument for contact, while a stronger case can be made by primarily looking specifically at regions of high multilingualism. This, for example, is the case of Sulaymānī Balochi, on which ongoing work with native-speaker Maryam Jamali has already revealed heavy influence of Sindhi, including direct morphological copying.

We hope that the present study serves to invite further attention to this region, hitherto-fore largely ignored in both Indological and Iranistic linguistic research. Sindhi remains the least studied of the larger Indo-Aryan languages despite having features seen nowhere else in its family, of interest to each of descriptivists, theoreticians and diachronicists.

Acknowledgments

The authors would like to thank Maryam Jamali and the Vancouver Baloch Community for their collaboration discussing examples from Balochi and Brahui. Unless otherwise noted, the Sindhi data used in this paper is Nirmal Riaz's own.[28]

Versions of this paper were presented at Formal Approaches to South Asian Linguistics on April 12, 2025, and at a DravLing meet on June 15, 2025. We thank our audiences for comments and suggestions, especially Masato Kobayashi, Liaquat Ali, Nazir Shakir Brahui, Gopalakrishnan Ramamurti, Sanford Steever, Rajesh Bhatt and Miriam Butt.

Thanks also to Masato Kobayashi for pre-publication access to his Brahui text corpus co-edited with Liaquat Ali, and to Ryosuke Masaoka, Sakura Ishikawa, Nazir Shakir Brahui, Fatemeh Sheybanifard, Carina Jahani, Agnes Korn, Maryam Nourzaei, Aleksandr Sergienko, Juho Pystynen, Martin Kümmel, Andrew Ollett and Elena Bashir for pointers to related literature and for sending us various relevant articles we didn't have access to.

[28] Other data sources (primarily for tables) include Barbera (2005) for Mīnābī, Barbera (2023) for Southern Bashkardi, Gholami & Pouladi (2024) for Zoroastrian Dari, Hook (1987) for Poguli, Nourzaei et al. (2015) for Koroshi, Bashir et al. (2019) for Saraiki. The rest of the tabled data are either cited in-text, or (re)produced from memory (Sanskrit, Avestan).

References

Akkuş, Faruk, David Embick & Mohammed Salih. 2025. *Case and the syntax of argument indexation: An analysis of Sorani Kurdish and beyond*. Oxford.

Ali, Liaquat & Masato Kobayashi. 2024. *Brahui texts: Glossed and translated short stories and folktales*. Research Institute for Languages and Cultures of Asia and Africa.

Andronov, Mixail Sergejevič. 1971. *Jazyk Braui*. Nauka.

Anonby, Erik & Mohebbi Bahmani Hassan. 2016. Shipwrecked and landlocked: Kholosi, an Indo-Aryan language in southwest Iran. *Studia Iranica* 43. 1–23.

Arora, Aryaman. 2020–. *Kholosi dictionary*. https://aryaman.io/kholosi/dictionary.html.

Arora, Aryaman. 2022. Historical phonology and other observations on Kholosi. https://aryaman.io/kholosi/Kholosi.pdf.

Axenov, Serge. 2006. *The Balochi language of Türkmenistan: A corpus-based grammatical description*. Acta Universitatis Upsaliensis.

Barbera, Gerardo. 2005. *Lingua e cultura a Minâb (Iran sud-orientale): Profilo grammaticale, testi e vocabolario*: Università degli Studi di Napoli "L'Orientale" dissertation.

Barbera, Gerardo. 2023. "The cannibal sister": Linguistic annotation of a South Bashkardi folktale. *Folia orientalia* 60.

Barjasteh Delforooz, Behrooz. 2008. A sociolinguistic survey among the Jadgal in Iranian Balochistan. In Carina Jahani, Agnes Korn & Titus Paul (eds.), *The Baloch and others: Linguistic, historical and socio-political perspectives on pluralism in Balochistan*, Dr. Ludwig Reichert.

Barjasteh Delforooz, Behrooz. 2020. Two Brahui texts with glossary and grammatical analysis. *Iranian Journal of Applied Language Studies* 12(1). 89–122.

Bashir, Elena, Thomas J. Conners & Brooke Hefright. 2019. *A descriptive grammar of Hindko, Punjabi, and Saraiki*. De Gruyter–Mouton.

Basu, Samopriya. 2024. Disproof of North Dravidian: Evidence from Indo-Aryan and Iranic Loanwords. Presented as an online talk to the DravLing working group.

Bhawnani, Murli D. 1979. *Descriptive analysis of Thari: A dialect of Sindhi language*: dissertation.

Bray, Denys de Saumarez. 1909. *The Brahui language: Introduction and grammar*. Reprinted (2010), Asian Educational Services.

Bubeník, Vít. 1992. On the use of pronominal clitics in Late Middle Indo-Aryan. *Wiener Zeitschrift für die Kunde Südasiens und Archiv für indische Philosophie* XXXVI(supplement, 17–18).

Butt, Miriam. 2007. The role of pronominal suffixes in Punjabi. *Architecture, rules, and preferences* 341–368.

di Carlo, Pierpaolo. 2008/2011. Two clues of a former Hindu Kush linguistic area. In Carol Everhard & Elizabeth Mela-Athanassopoulou (eds.), *Selected papers from the International Conference on Language Documentation and Tradition with a Special Interest in the Kalasha of the Hindu-Kush Valleys, Himalayas*, vol. 7–9, 101–114.

Collett, Nigel A. 1983. *A grammar, phrase book and vocabulary of Baluchi (as spoken in the Sultanate of Oman)*. Camberley.

Dabir-Moghaddam, Mohammad. 2008. On agent-clitics in Balochi in comparison with other Iranian languages. In Carina Jahani, Agnes Korn & Titus Paul (eds.), *The Baloch and others: Linguistic, historical and socio-political perspectives on pluralism in Balochistan*, Dr. Ludwig Reichert.

Dames, Mansel Longworth. 1922. *A text book of the Balochi language: Consisting of miscellaneous stories, legends, poems and Balochi–English vocabulary*. Superintendent, Govt. Print., Punjab.

Elfenbein, Josef. 1989. Balochi. In Rüdiger Schmitt (ed.), *Compendium linguarum iranicarum*, Ludwig Reichert.

Elfenbein, Josef. 1998. Brahui. In Sanford Steever (ed.), *The Dravidian languages*, Routledge.

Elfenbein, Josef H. 1980. *The Baluchi language: A dialectology with texts*. Indus publications.

Gholami, Saloumeh & Mehraban Pouladi. 2024. Linguistic insights from a bilingual letter: The Malati dialect of Zoroastrian Dari in Yazd; Part I: Transcription, translation, and linguistic structure. *Iran and the Caucasus* 28. 437–462.

Haspelmath, Martin. 2019. Indexing and flagging, and head and dependent marking. *Te Reo* 62(1 (Issue in Honour of Frantisek Lichtenberk)). 93–115.

Haspelmath, Martin. 2023. Types of clitics in the world's languages. *Linguistic Typology at the Crossroads* 3(2). 1–59.

Hook, Peter Edwin. 1987. Poguli syntax in the light of Kashmiri: A preliminary report. *Studies in the Linguistic Sciences* 17(1). 63–71.

Jahani, Carina. 2019. *A grammar of modern standard Balochi*. Acta Universitatis Upsaliensis.

Jahani, Carina & Agnes Korn. 2009. Balochi. In Gernot Windfuhr (ed.), *The Iranian languages*, Routledge.

Jegorova, Raisa Petrovna. 1964. Mestoimennyje ènklitiki dlja glagolov na Sindxi. *Indijskaja i iranskaja filologija*. 125–141.

Jegorova, Raisa Petrovna. 1966. *Jazyk Sindxi*. Nauka.

Jetley, Murlidhar Kishinchand. 1964. *Morphology of Sindhi: A descriptive analysis of Vicoli, the standard Sindhi dialect*: dissertation.

Khubchandani, Lachman. 1969. Pronominal suffixes in Sindhi. *Indian Linguistics* 23. 72–81.

Khubchandani, Lachman. 2007. Sindhi. In George Cardona & Dhanesh Jain (eds.), *The Indo-Aryan languages*, Routledge.

Kobayashi, Masato. 2020/2021. Viewing Proto-Dravidian from the northeast. *Journal of the American Oriental Society* 140(2). 467–482.

Kobayashi, Masato & Liaquat Ali. 2022–. Asyndetic conditional clauses in Brahui. In Raghavachari Amritavalli & Bhuvana Narasimhan (eds.), *The Oxford handbook of Dravidian languages*, Oxford.

Korn, Agnes. 2009. Western Iranian pronominal clitics. *Orientalia Suecana* 58. 159–171.
Korn, Agnes & Maryam Nourzaei. 2018. "Those were the hungry years": A glimpse of coastal Afro-Balochi. *Journal of the Royal Asiatic Society* 28(4). 661–695.
Korn, Agnes & Maryam Nourzaei. 2019. Notes on the speech of the Afro-Baloch of the southern coast of Iran. *Journal of the Royal Asiatic Society* 29(4). 623–657.
Koul, Omkar N. & Kashi Wali. 1997. *Kashmiri: A cognitive-descriptive grammar*. Routledge.
Krishnamurti, Bhadriraju. 2003. *The Dravidian languages*. Cambridge.
Kümmel, Martin J. to appear. The thematic genitive singular in Indo-Iranic. In Rachele Pierini, Birgit Olsen & Eugenio Luján (eds.), *Rethinking the branching order of the Indo-European language family through the diachronic morphophonology of the o-stem genitive singular ending*, Brill.
Lehr, Rachel. 2014. *A descriptive grammar of Pashai: The language and speech community of Darrai Nur*: University of Chicago dissertation.
Masica, Colin P. 1993. *The Indo-Aryan languages*. Cambridge.
McAlpin, David W. 2003. Velars, uvulars, and the North Dravidian hypothesis. *Journal of the American Oriental Society* 123(3). 521–546.
Meihami, Saman & Nian Barzanji. 2025. A case-driven account of the alignment in Sorani Kurdish. Presented as an online talk at the fourth North American Conference on Iranian Linguistics.
Mohammadirad, Masoud. 2020. *Pronominal clitics in Western Iranian languages: Description, mapping, and typological implications*: Université Sorbonne Nouvelle, Paris 3 dissertation.
Mukherjee, Sibasis. 1992. *Phonology and morphology of Jadeja dialect of Sindhi*: dissertation.
Nawata, Tetsuo. 1981. *Baluchi*. Institute for the study of Languages and Cultures of Asia and Africa.
Nourzaei, Maryam. 2024. Kholosi. In Charles Häberl (ed.), *Language diversity in Iran: New texts and perspectives from non-Iranian languages*, De Gruyter-Brill.
Nourzaei, Maryam. in prep. *Kholosi: A corpus-based grammatical description*: Uppsala universitet dissertation.
Nourzaei, Maryam, Carina Jahani, Erik Anonby & Abbas Ali Ahangar. 2015. *Koroshi: A corpus-based grammatical description*. Acta Universitatis Upsaliensis.
Rahman, Mutee U. undated. Computational linguistics issues and challenges: Sindhi pronominal suffixes, prodrop, and particles. Presentation.
Rezaei, Tahereh. 2020. First notes on the syntax of Kholosi as a heritage language in the south of Iran. https://microcontact.sites.uu.nl/wp-content/uploads/sites/210/2020/01/HLS1-Rezaei.pdf.
Salman, Amel & Nafla S. Kharusi. 2012. The sound system of Lawatiyya. *Journal of Academic and Applied Studies* 2(5). 36–44.
Scheller, Meinrad. 1967. Das mittelindische enklitikum *se*. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 81(1/2).

Shackle, Christopher. 1976. *The Siraiki language of Central Pakistan: A reference grammar*. School of Oriental and African Studies.

Shazia, Zahid Ali, Gulkhanda & Shahida. 2022. A morphosyntactic analysis of Kalati Brahui verbs. *Competitive Social Sciences Research Journal* 3(2). 74–87.

Sokolov, Sergej Nikolajevič. 1956. Grammatičeskij očerk jazyka beludžej Sovetskogo Sojuza. *Trudy instituta jazykoznanija akademičeskij nauk SSSR*. 154–192.

Steever, Sanford. 1993. *Analysis to synthesis: The development of complex verb morphology in the Dravidian languages*. Oxford.

Taghipour, Sahar & Arsalan Kahnemuyipour. 2019. Hybrid alignment in Laki agreement and the special status of clitics. In Katie Martin (ed.), *Proceedings of the* 37th *Annual Conference of the Canadian Linguistics Association*, .

Trumpp, Ernest. 1872. *Grammar of the Sindhi language*.

Varyani, Pritam L. 1976/1977. Pronominalized verbal constructions in Sindhi. *Bulletin of the Deccan College Post-Graduate and Research Institute* 36(1/4). 146–143.

Verbeke, Saartje. 2017. *Argument structure in Kashmiri: Form and function of pronominal suffixation*. Brill.

von Hinüber, Oskar. 2001. *Das ältere Mittelindisch im Überblick*. Österrichischen Akademie der Wissenschaften.

Wali, Kashi & Ashok Kumar Koul. 1994. Kashmiri clitics: The role of case and CASE. *Linguistics* 36(6). 969–994.

Yosida, Yutaka. 2022. *Sogudogo bunpō kōgi*. Rinsen.

Zograf, Georgij Aleksandrovič & Anton Ilič Kogan. 2011. Sindxi jazyk. In Tatjana I. Oranskaja, Julija V. Mazurova, Andrej A. Kibrik, Leonid I. Kulikov & Aleksandr Ju. Rusakov (eds.), *Novyje indoarijskije jazyki*, Rossijskaja akademija nauk: Institut jazykoznanija.

Şahingöz, Emine & Samopriya Basu. 2025. Introducing the Handbook of Ossetic Preverbs: An overview of the contextual change of dynamic verbs induced by preverbs. Presented as a talk at the tenth International Conference on Iranian Linguistics.

Šeybānīfard, Fāṭemeh & Zahreh Zaršenās. 2020. Towṣīf-e sāχtemān-e fe‘l dar zabānī Berāhūī-e Rūdbār-e Jonūb. *Pažūheš-e zabān pažūhī* 12(35). 177–204.

Re-visiting honorification in Hindi and the discourse particle *jii*

OJASWEE BHALLA, *Indian Institute of Technology, New Delhi*
DEEPAK ALOK, *Indian Institute of Technology, New Delhi*

ABSTRACT

The ungrammaticality of the honorific particle *jii* following a plural suffix in Hindi nominals has previously motivated a formal analysis that structurally posits complementarity between semantic plurality and honorification in the language. We propose an alternative analysis that maintains distinct structural positions for discourse particle *jii* and a separate Hon(orific) Phrase (that is relevant for the verbal agreement system and other nominal modifiers) in Hindi, each with its own semantic import to the nominal meaning. This is supported by empirical evidence like honorific mismatch scenario and data where various pluralizing strategies can be adopted based on the presence or absence of discourse particle *jii.* We also propose that this discourse particle has its own idiosyncratic restriction of selecting a specific and a maximal nominal complement, that controls the environment within which it can occur. The core idea that this paper presents is that along with co-opting of plural morphology to express honorification in Hindi, this system is sensitive to the presence of honour-lending discourse marker *jii* and the variable valuation of Hon – that directly impacts the interpretation and may lead to extra semantic effects.

1 Introduction: Honorification cues in Hindi

Honorification is a linguistically instantiated strategy of conveying the speaker's mental attitudinal state of respect (or lack of it) for a discourse referent – such as an argument or the addressee. This respect or honour is associated with an array of factors like the speaker's age, gender, kinship or any other social relation (with various degrees of intimacy or formality) with the discourse referent, their relative social standing (owing to professional stature) and the nature of the discourse context itself within which the utterance is located. This section discusses the honorification cues employed by Hindi (Indo-Aryan; SOV) speakers.

In the Hindi nominal domain, there are morphological and lexical cues for honorification whereas in the verbal domain – the verb's agreement with the argument reflects honorification status, when available.[1] The 2nd person (2P) pronominal form is morphologically distinct for a non-honorific (NH) referent (1a) in contrast to an honorific (H) referent (1b).[2,3] The demonstratives in their oblique form also cue the honorification

[1] Some of these points have also been noted in recent formal analyses of honorification in Hindi like Bhatt and Davis (2023) and Sinha (2023).
[2] The *tu/tum* form distinction for an NH 2P referent is due to an intimacy distinction between them, briefly discussed in section 5.2.
[3] Glossing is according to Leipzig glossing rules. Broad transcription adopted – where retroflexes are capitalized, nasalization is conveyed by N and long vowels are marked by two instances of the said vowel.

status of the nominal, with the regular singular oblique demonstrative *us* cuing the NH status (2a) and the plural oblique demonstrative *un* cuing H status (2b) for a semantically singular referent.[4] For the possessive in (3), the oblique form (homonymous with the plural form) cues the H status of the nominal whereas the direct form (homonymous with the singular form) cues the NH status. Hindi has a discourse particle *jii* which when marked on a nominal (4) indicates the speaker's respect for that discourse referent. Besides these, the standalone lexical words like *saab* 'sir/mister' or *mam* 'mam' can also follow a nominal to cue honour for them (5).

(1) a. tu/tum
2P.NH
b. aap
2P.H
'you'

(2) a. us neta
3P.OBL.NH politician
b. un neta
3P.OBL.H politician
'that politician'

(3) a. mera daamaad
my.NH son-in-law
b. mere daamaad
my.H son-in-law
'my son-in-law'

(4) Ravi-jii
ravi-JI

(5) Ravi-saab
ravi-SAAB

At the sentence level, Hindi speakers can infer the honorification status of a referent based on the verbal agreement pattern – with or without any honorification cue being made available in nominal domain. For the unmarked argument 'Ravi' in (6), the verbal agreement in (6a) cues NH status whereas the one in (6b) indicates the H status for Ravi vis-à-vis the speaker. While the verb in (6a) has the default agreement with a 3P singular argument, the verb in (6b) shows agreement with a 3P plural subject, even though the subject argument is semantically and formally singular. Because verbal agreement with an honorific (singular) subject parallels the verb agreement with a plural subject (7a) –

[4] Bhatt and Davis (2023) observe that the direct forms of the Hindi demonstrative *vo/ve* do not make a complementary distinction between taking singular versus plural referents (with *ve* being plural and *vo* being number neutral) and this leads to both of them co-occurring with honorific nominals.

honorific agreement has been termed as 'plural agreement' in the literature (Bhatt and Davis 2023, Sinha 2023). We agree with them on the claim that Hindi co-opts plural morphology to convey honorification.

(6) a. Ravi aaya hai
Ravi come.PFV.MSG be.PRS.3SG
'Ravi has come.'
b. Ravi aaye haiN
Ravi come.PFV.MPL be.PRS.3PL
'Ravi has come.' (speaker respects Ravi)

(7) a. Bachche aaye haiN
child.PL come.PFV.MPL be.PRS.3PL
'Children have come.'
b. Bachcha aaya hai
child.SG come.PFV.MSG be.PRS.3SG
'Child has come.'

The next section discusses the interaction of honorification with pluralizing strategies and outlines Bhatt and Davis' formal analysis of Hindi honorification based on a key empirical observation.

2 Honorification and Plurality

2.1 Interaction of Plural Suffix and Discourse Particle *jii*

Bhatt and Davis (2023) observe that Hindi does not permit the co-occurrence of its plural suffix *-aN* or *-e* (their oblique form *-oN*) and the discourse particle *jii* on a nominal, as highlighted by the data in (8).

(8) * maasTar-oN-jii aaye haiN
teacher-PL-JI come.PFV.MPL be.PRS.3PL
'(Respected) teachers have come.'

This language-specific pattern is contradicted in a neighboring Indo-Aryan language – Magahi – where the homonymous honorific discourse particle *jii* can co-mark a nominal with its pluralizing suffix *-an* (Alok and Bhalla 2024). Sentences (9a-b) provides an example of this robust Magahi pattern.[5] In addition to *jii*, Magahi has another discourse particle *-waa,* which has been previously analyzed as a familiarity or an intimacy marker for the language (Alok 2012, 2014). In Magahi, the auxiliary verb reflects the honorification status of the subject as well as the addressee. The morpheme *-au* in (9a) expresses a non-honorific subject (and a non-honorific addressee) and thus, the sentence

[5] Glosses specific for Magahi data: NHS- Non-Honorific Subject; HS- Honorific Subject; NHA – Non-Honorific Addressee; WA – discourse particle -waa.

coveys a meaning where the speaker does not respect the subject 'teachers'. On the other hand, the morpheme *–(a)thun* in (9b) marks speaker's respect for the subject 'teachers' (as well as NH attitude for the addressee).

(9) a. maasTar-jii-wa-an aavit hau
teacher-JI-WA-PL coming be.AUX.NHS.NHA
'The teachers are coming.'
Intended: Teachers are not respected by the speaker (i.e. the student's parents, who are angry with the teachers)

b. maasTar-jii-wa-an aavit hathun
teacher-JI-WA-PL coming be.AUX.HS.NHA
'The teachers are coming.'
Intended: Teachers are respected by the speaker and they are in a close relationship

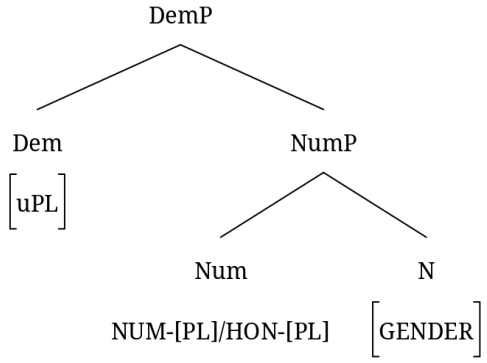
The key takeaways from this sub-section are that contrary to Hindi, in a Magahi DP, the honorificity markers and plural morphemes can come together and thus, they are not in complimentary distribution. Magahi allows different honorific markers to co-occur in the same DP and there could also be a potential honorific mismatch in the nominal and the verbal domain, leading to extra semantic effects. The *prima facie* complementarity between Hindi plural suffix and honorification cue *jii*, as in the case of (8), underlies the formal theoretical analysis proposed by Bhatt and Davis (2023). This is discussed in the next section.

2.2 Bhatt and Davis (2023)

Bhatt and Davis, motivated by data like (6b) and (8), interlink plurality and honorification in the derivational apparatus. They propose that the formative that contributes the semantics of plurality and the formative that contributes the semantics of honorification occupy the same syntactic position in the nominal spine – the Num(ber) head. They use '*' and 'Hon' as the labels for the two formatives respectively. The Num head, when it is occupied by *, gets a semantic plurality feature – the NUM-[PL]. This semantic plurality feature of the nominal is responsible for triggering the regular plural agreement on the verb. When the Num head is occupied by the Hon formative, it gets associated with a formal, non-semantic plurality feature – the HON-[PL]. This feature is responsible for triggering non-semantic plural agreement for singular honorified nominals (like in the case of 6b). The structure proposed by them is sketched in (10).

(10) Labelled Bracketing[6]

[6] [DemP
[Dem #uPL]



A direct outcome of this structure is that Hon and * are in complementary distribution. As a corollary, Num occupied by Hon can only take morphologically and semantically singular nominal complements. Bhatt and Davis analyze the ungrammatical N-PL-JI structures (like in example 8) by assuming particle *jii* to be an overt instantiation of Hon – which would rule out co-occurrence of both *jii* and the plural suffix *-oN* as they both occupy the same Num head position.

Bhatt and Davis' proposal provides an explanatorily adequate account of some empirical facts about honorification in Hindi. However, we present some data in the next section that cannot be accounted for within the syntax laid out by them – which motivates a re-investigation of the formal architecture.

2.3 Various pluralizing strategies and honorification – the broader picture

The structure proposed by Bhatt and Davis comes with an implicit consequence – as Hon can only select for singular nominals as complements, therefore Hindi does not permit honorification of plural referents. However, honorification and plurality are orthogonal notions and conceptually, should not be in a complimentary distribution. This line of thought gains traction once the full paradigm of honorification with various pluralizing strategies is observed in Hindi.

Example (11) exhibits singularity (of the nominal) – honorification interaction. The context in example (11) elicits sentences where a singular subject nominal gets overtly marked by *jii* (11c-d) or does not get marked by *jii* (11a-b), with each pair differing with respect to verbal agreement cuing honorification or not. This data shows that the verbal cue is sufficient criteria to express honorification for a discourse referent and the presence of particle *jii* is optional. For the sentence in (11c), we note a variable judgement with the sentence being partially acceptable for eastern Hindi variety speakers

---

[NumP
  [Num NUM\-\[PL\]\/HON\-\[PL\]]
  [N #GENDER]
]
]

(subject to contextual considerations) and being unacceptable for western Hindi variety speakers.

(11) Context: In the meeting of a village panchayat, the head of the panchayat observes the absence of female attendees and asks the male members where someone's wife is. To this, he hears the response:

a. devi ghar pe hai
lady.SG home LOC be.PRS.3SG
'The lady is at home.'

b. devi ghar pe haiN
lady.SG home LOC be.PRS.3PL
'The (respected) lady is at home.'

c. ?/* devi-jii ghar pe hai
lady.SG-JI home LOC be.PRS.3SG
'The (respected) lady is at home.'

d. devi-jii ghar pe haiN
lady.SG-JI home LOC be.PRS.3PL
'The (respected) lady is at home.'

In contrast to (11), its minimal pair in (12) explores plurality-honorification interaction. Grammaticality of non-*jii* marked plural nominals in (12a-c) exhibits that plural nominals can be honorified in Hindi. As for *jii*-marked plural nominals in (12d-f), the ungrammaticality of (12d-e) shows that N-aN-JI is not a permissible structure in Hindi. The grammatical sentence in (12f) shows that a *jii* marked nominal can be pluralized using a universal quantifier *sab* 'all' in Hindi.

(12) Context: In the meeting of a village panchayat, the head of the panchayat observes the absence of female attendees and asks the male members where all their wives are. To this, he hears the response:

a. devi-aN ghar pe haiN
lady-PL home LOC be.PRS.3PL
'The ladies are at home.'

b. sab devi-aN ghar pe haiN
all lady-PL home LOC be.PRS.3PL

c. sab devi ghar pe haiN
all lady home LOC be.PRS.3PL
'All the ladies are at home.'

d. * devi-aN-jii ghar pe haiN
lady-PL-JI home LOC be.PRS.3PL
'The (respected) ladies are at home.'

e. * sab devi-aN-jii ghar pe haiN
all lady-PL-JI home LOC be.PRS.3PL

f. sab devi-jii ghar pe haiN
all lady-JI home LOC be.PRS.3PL
'All the (respected) ladies are at home.'

We observed (post-FASAL) that particle *jii* marked nominal can be pluralized using two more strategies – using *log* 'people' after the N-*jii* complex (like example (13)) and for some, adding the plural suffix -*oN* to the N-*jii* unit is also acceptable (like in example (14)). We would like to note a judgement difference here – younger Hindi speakers find (14) more acceptable than elder Hindi speakers. These data points were raised in our SICOGG-27 conference paper (Bhalla and Alok 2025) and we direct the interested reader to the proceedings paper (published online).

(13) neta-jii-log baRhija kaam kar rahe haiN
politician-JI-people good work do.INF stay.PROG.MPL be.PRS.3PL
'Politicians are doing a good job.'

(14) maataa-jii-oN-ne khaana khaa liya
mother-JI-PL-ERG food eat.INF take.PFV.MSG
'The mothers ate the food.'

Interim summary:

The data above exhibits that singular referents, whether *jii* marked or not, can get honorific interpretation. For plural referents, the presence of *jii* particle appears to control the type of plurality indicating cue that nominal can get marked by. The non-*jii* marked plural nominals get honorific interpretation by virtue of the verbal agreement pattern. Additionally, the distributional pattern of particle *jii* is such that:

(i) Discourse particle *jii* cannot mark a nominal that has been marked by the plural suffix -*aN*.

(ii) *Jii*-marked nominal can be pluralized using a universal quantifier like *sab* 'all'.

Based on (i) and (ii), we propose that discourse particle *jii* controls for the pluralizing strategy that can be employed for a nominal it marks and that this particle cannot follow the regular plural suffix in a Hindi nominal.

3 Some other Issues

The structure proposed by Bhatt and Davis, repeated here as (15), stipulates that once the Num head is occupied by HON, the higher NP layer also gets valued (formally) as plural because of the formal [PL] feature of the Hon. Thus, data like (16) cannot be accounted for in their system where we have an honorified nominal (since *jii* is an overt instantiation of Hon for Bhatt and Davis) co-occurring with a formally singular demonstrative *us* 'that'.

(15) [Dem[uPL] [[Hon[PL] [[Num[SG] [N[GENDER]]]]]]]

(16) Context: A teacher has been caught by a police officer for some alleged crime. He asks the constable to bring her in his office for interrogation and says:

us	maasTarni-jii	ko	andar	bhejo
that.OBL.SG	teacher.F-JI	DAT	inside	send.IMP.2

'Send that teacher inside.'
Intended: Lack of respect for the (by-default respected) teacher

The utterance of (16) in the given discourse context has some extra semantic-pragmatic effects besides what is strictly entailed by that sentence. This sentence expresses the speaker's (the police officer's) manipulation of the social relation that exists between them and the marked nominal (teacher). The speaker bleaches or weakens the respect (cued by demonstrative *us*) that a teacher is otherwise assumed to have in the default social setting (owing to their profession, cued by the discourse particle *jii* following the nominal), possibly due to their anger or frustration at them/the situation. This type of honorific mismatch resulting in extra semantic effects was evidenced for Magahi in (9a) too.

Another potential issue with Bhatt and Davis' proposal is that they take the Hindi agreement system to be "oblivious to honorificity". For them, "there are no phi-features specifically associated with honorificity and we do not need to adjust the agreement algorithm to handle honorificity" (Bhatt and Davis 2023 pp:30). Keeping discourse-sensitive notions, like honorificity, out of the verbal agreement pattern complicates the analysis of data like (17) – where we have the honorific 2P pronominal *aap* 'you' co-indexed with *jii*-marked nominals in both parts of the minimal pair but the verbal ending distinctly co-vary.[7]

(17) a.

sasur-jii	aap	khaana	khaiye
father-in-law-JI	you.HON	food	eat.IMP

'Father-in-law, please eat the food.'

b.

papa-jii	aap	khaana	khaao
father-JI	you.HON	food	eat.IMP

'Father, please eat the food.'

Since the subject pronominal in both (15a) and (15b) have the same 2P honorific features, the formal machinery can account for the verb variation if we decompose the discourse-sensitive features at play that distinguish the speaker-father social relation from the speaker-father-in-law relation. This is attempted in section 5.2.

In the next section, we propose an alternative analysis to account for the honorification mechanism underlying the Hindi data discussed so far.

4 Our Proposal

[7] We thank an anonymous reviewer for bringing up the relevance of this distinction for the discussion at hand.

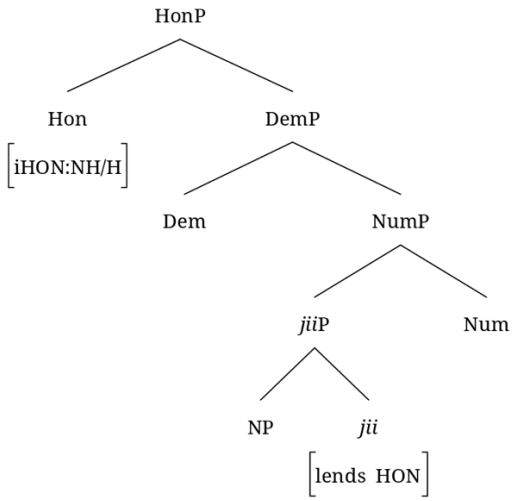
We adopt Alok (2020), who builds upon Portner *et al.*'s (2019) proposal, in assuming an Hon(orific) Phrase above all DPs in Hindi. The Hon head has an interpretable honorific feature – [iHON] – that establishes the honorification relation between the speaker and the referent of that specific DP. Assuming the syntactic instantiation of the speaker (and the addressee) in the higher left periphery of the clause, the [iHON] specifies either an H (honorific) relation or an NH (non-honorific) relation between a variable bound by the speaker co-ordinate in the left periphery and the denotation of the nominal complement it takes. The interpretation function applied to [iHON] when it is valued for the NH relation is specified in (18).

(18) $[|\text{iHON}|]=\lambda x.S_i \prec x$, where $\prec$ represents the hierarchical honorific relation

[iHON] is semantically interpretable for the nominal and is also the relevant formal feature that gets accessed by the verbal domain for the agreement operation. Thus, contrary to the existing analysis, we propose that the Hindi verbal domain is sensitive to both the phi features and the [iHON] value of the argument that it agrees with. We also propose that discourse particle *jii* is not the overt instantiation of Hon (even though it also has an honorificity-oriented meaning to contribute) and instead projects a distinct phrase of its own above the N. This structure is sketched in (19).

(19) Labelled bracketing[8]

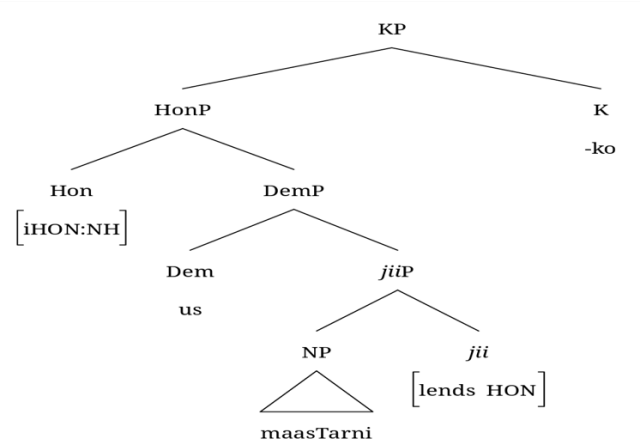
[8] [HonP
[Hon #iHON\:NH\/H]
[DemP
[Dem]
[NumP
[*jii*P
[NP]
[*jii* #lends HON]
]
[Num]
]
]
]



jii is a discourse particle that does not contribute to the truth-conditional meaning dimension but it has an honorification lending meaning core encoded in it that contributes to the expressive dimension. This particle does not have the formal feature that the agreement system is sensitive to. A noun that gets marked by particle *jii* conveys the default setting of respect being accorded to the marked nominal in that discourse context. However, it is not the sole contributor of honorification meaning since the value of Hon is also relevant. A nominal marked by *jii* with its [iHON] valued H represents the typical situation – where the honorification-oriented meaning contributed by both align. A nominal marked by *jii* with its [iHON] valued NH represents the honorific mismatch situation that leads to extra semantic effects – as was the case in (16) (see Porter *et al.* 2019, Alok 2020 for a detailed discussion of the dynamic property of honorification/politeness). The value of [iHON] conditions the morphological realization of other non-nominal elements in the NP. Since Hindi co-opts plural morphology for honorification, the H value of [iHON] selects for the plural form of Dem whereas the NH value of [iHON] selects for the singular form of Dem. We propose that the total interpretative import of a nominal is a composite of the honorification-oriented meaning by lexical items like *jii* and the formal [iHON] feature (cued by the form of nominal modifiers or the verbal agreement).

The derivation of data in (16) proceeds as follows: [iHON] gets valued NH – which then selects for the *us* form over the *un* form of the demonstrative; particle *jii* contributes its own meaning – lends honor to the nominal it marks (the default respect setting). This mismatch becomes acceptable when the speaker weakens the by-default respect conferred by *jii* because of some specific socio-pragmatic reason (like anger). The structure for this derivation is given in (20). Note that we put the KP (case layer) above HonP because case-marked nominals are not available for subject honorification in Hindi.

(20) Labelled bracketing[9]



An aside:

For non-case marked nominals, the direct forms of both the demonstratives *vo/ve* can co-occur with an honorific nominal (see footnote 4). In a non-honorific context, the number neutral *vo* can modify the nominal. We observe that for data like (21a) there is an acceptability difference. This sentence is not acceptable for the co-author who is a western Hindi variety speaker but is acceptable for the author who is an eastern Hindi variety speaker. The non-honorific interpretation becomes more easily available once we have an overt indication of the honorific mismatch in the DP – with the demonstrative *vo* that is ambiguous between NH/H marking. (21b) exhibits this. We think that this could be because of the influence of honorification paradigm in a speaker's other acquired languages. For the speakers whose other languages robustly exhibit honorification mismatches (like Magahi), they require fewer overt cues of mismatches.

(21) a. maiDam-jii aa rahi hai
madam-JI come stay.PROG.F be.PRS.3SG
'The teacher is coming.'

[9] [KP
[HonP
[Hon #iHON\:NH]
[DemP
[Dem us]
[*jii*P
[NP ^maasTarni]
[*jii* #lends HON]
]
]
]
[K \-ko]
]

b. vo maiDam-jii aa rahi hai
that madam-JI come stay.PROG.F be.PRS.3SG
'The teacher is coming.' *(intended-less respected)*

The next section deals with explaining the plurality-honorification interaction and the case for making Hindi agreement system not oblivious to discourse-sensitive information.

5 Resolving the other Open Issues

5.1 The case of Honorific Plural nominals

We agree with Bhatt and Davis (2023) and Sinha (2023) that Hindi does co-opt plural morphology to convey honorification. This aligns well with Wang's (2023) typological observations about honorification expression across world's languages.[10] However, she notes that plurality is one amongst the three features (indefiniteness and third person being the other two) that are the systems recycled by languages to express honorification (for a detailed discussion see section 6). By extension, honorification should not have a one-on-one mapping with plurality and our proposal outlined in the preceding section, where we dissociate an HonP and a *jii*P projection from the NumP projection for a Hindi DP, is in line with it. Such an analysis allows for both the singular and plural referents to be either honorific (with or without *jii* particle) or non-honorific (with or without *jii* particle) in the language. Table 1 gives an example for all these combinatorial possibilities.

***jii*P**	**NumP**	**HonP**	**Example**
-	SG	NH	Ravi aaya hai ravi come.PFV.MSG be.PRS.3SG 'Ravi has come.'
-	PL	NH	Bachche aaye haiN child.PL come.PFV.MPL be.PRS.3PL 'Children have come.'
-	SG	H	Ravi aaye haiN ravi come.PFV.MPL be.PRS.3PL 'Ravi has come.'
-	PL	H	maataa-eN aayiN haiN mother-PL come.PFV.FPL be.PRS.3PL 'Mothers have come.'
+	SG	NH	us maasTarni-jii ko bolo that.OBL.SG teacher-JI DAT say.IMP 'Tell that teacher'
+	PL	NH	* us maasTarni-jii-logo-N ko bolo that.OBL.SG teacher-JI-people-PL DAT say.IMP 'Tell that teachers that'

[10] We thank Rajesh Bhatt for bringing Wang's work to our attention.

			*us maataa-jii-oN ko bolo that.OBL.SG mother-JI-PL DAT say.IMP 'Tell that mothers that'
+	SG	H	Ravi-jii aaye haiN ravi-JI come.PFV.MPL be.PRS.3PL 'Ravi has come.'
+	PL	H	un maasTarnii-jii-log-oN ko bolo that.OBL.PL teacher-JI-people-PL DAT say.IMP 'Tell those teachers that' un maataa-jii-oN ko bolo that.OBL.PL mother-JI-PL DAT say.IMP 'Tell those mothers that'

Table 1: Number-Honorification Landscape in Hindi

The empirical observation about particle *jii* not occurring with a plural suffix (*-aN*) marked nominal (discussed in section 2.3, relevant data from (12) repeated below as (22)) still warrants an explanation.

(22) a. devi-aN ghar pe haiN
lady-PL home LOC be.PRS.3PL
'The (respected) ladies are at home.'
b. * devi-aN-jii ghar pe haiN
lady-PL-JI home LOC be.PRS.3PL
'The (respected) ladies are at home.'
b. sab devi-jii ghar pe haiN
all lady-JI home LOC be.PRS.3PL
'All the (respected) ladies are at home.'

What is the difference between *devi-aN-jii* and *sab devi-jii*? The bare plural *devi-aN* in (22a) gets a definite interpretation in Hindi, like the definite plural DP 'the ladies' in English. Brisson (2003), building up on previous literature that stated that quantification associated with definite plurals is somehow weaker than universal quantification, had proposed that definite plurals are 'non-maximal' in nature because the predication with them is evaluated as true even if most of the people and not all of the referents in the denotation set were part of the predication set (This was semantically derived based on an ill-fitting, context-dependent domain selection variable 'cover' in the restriction of the D operator). The quantifier 'all' has a 'maximizing' effect since it requires a good-fitting cover and thus rules out non-maximality as a domain-of-quantification effect.

We propose that *jii* can mark a nominal only if it is 'maximal' in nature. Singular referents (by default) and *sab* 'all' marked nominals both receive a maximal interpretation and thus, can be marked by *jii*. Bare plurals that are marked by the plural suffix *-aN* are non-maximal in nature, and thus not compatible with the particle *jii*. Other

pluralizing strategy (like adding *log* 'people'- discusses in section 2.3 example (13)) operates over the [N-jii] complex, once the selectional requirements of particle *jii* are met.

Along with maximality, this particle also looks for a specific reference of the nominal it attaches to. This can be gleaned from the data in (23) – it is acceptable with a 2P pronoun (specific for a discourse context), proper nouns, relational nouns like *daamaad* 'son-in-law' (that are specific owing to being defined from the speaker's perspective) but not okay with common nouns like 'boy' (23d) (but we note that this seems to be okay for Bhatt and Davis), since common nouns do not single out some individual in the world as their referent and need further informational support to receive a specific reference.

(23) a. aap-jii
b. Ravi-jii
c. daamaad-jii
d. * larke-jii
boy-JI

The fact that discourse particles have their own idiosyncratic properties that restrict the felicitous environment in which they can occur has been previously noted in Bhalla (2022) for Hindi discourse particle *-to*. Therefore, it is not surprising that discourse particle *jii* also imposes its requirement on its complement and selects for a maximal and specific nominal.[11]

5.2 Difference between father and father-in-law for the verb

Another issue raised in section-3 highlighted the relevance of discourse-sensitive notions for the verbal domain. Hindi has three pronominal forms in second person *tu/tum/aap* that have been discussed much in the literature. Our take is that these pronouns encode an honorific distinction (respect based on social factors like age) and an intimacy distinction – whether the person is close to the speaker (i.e. in an intimate relation with them) or a non-intimate relation. We propose the featural composition of these pronouns in (24). And the varied verb ending for imperatives in (25) exhibit that both these features have a role to play in the syntax. Thus, Hindi verbs are sensitive to both honorification and intimacy information of the speaker and the nominal referents. The minimal pair in example (17) instantiated the two feature combinations available for *aap* – with father being honorific and intimate for the speaker (usual case but need not be the norm) and father-in-law being the honorific and non-intimate counterpart for the speaker.

[11] In our post-FASAL work, we have modified the analysis about requirement of particle *jii* – from selecting a 'maximal' nominal complement to selecting an 'atomic' one. We refer the interested reader to Bhalla and Alok (2025).

(24) a. Tu [-HON], [+INTIM]
b. Tum [-HON], [-INTIM]
c. Aap [+HON],[+INTIM]
d. Aap [+HON],[-INTIM]

(25) a. Tu khaa
b. Tum khaao
c. Aap khaao
d. Aap khaiye
'You eat'

The next section concludes and locates the findings of this paper across the broader discussion happening around honorification across languages.

6 Conclusion and Ending Comments

This paper has motivated an empirically-driven, theoretical account for honorification in Hindi that pushes for an HonP for every nominal and a *jii*P above an NP to host discourse particle *jii*, when available. The formal feature of [iHON] is semantically interpretable for the nominal and formally relevant for the verbal agreement system. The H/NH value of [iHON] selects for the form of the nominal dependents and honorification (and thus, H value of [iHON]) co-opts plural morphology in the language. The discourse particle *jii* imposes its idiosyncratic requirement of selecting a specific and maximal nominal complement. It also controls for the plurality-indicating cue that can co-occur with it, allows *sab* but not *–aN*. And interpretation-wise, honorification is a composite build-up of values of the formal, interpretable [iHON] feature and any honorification lending lexical items in the DP – match of values gives regular semantics whereas mismatch gives extra semantics.

One support for our alternative analysis comes from the recent typological work by Wang (2023). She shows that the languages of the world recycle a specific set of values from a finite feature repository to morpho-syntactically express honorification. Specifically, [3rd Person] feature or [indefinite] feature or a [plural] feature is the only choice that a language ever makes when selecting a system to co-opt for honorification. She proposes that in contexts requiring respect, speakers adhere to a socio-pragmatic maxim of Taboo of Directness, by virtue of which they avoid specific and direct reference to the respected person. The common core underlying [3P], [Plural] and [indefinite] is that they are semantically unmarked choices in their respective feature sets and thus ideal candidates to express honorification, since they are all relative vague and less specific. Thus, Hindi makes a language-specific decision to recruits plurality for honorification while a language like Italian employs person mismatch feature to convey deference to someone in a context. Our proposal is in-line with this over-arching picture

of honorification since it establishes that in Hindi, an H valued [iHon] selects for plural morphology on the nominal dependents to covey the honorific status of the nominal (irrespective of the fact whether the referent of the nominal is singular or plural) and the verbal agreement also reverts to plural paradigm when it sees an H valued [iHON]. This language specific choice becomes visible at the time of vocabulary insertion. This proposal is thus, perspective-wise different from an approach that assumes a non-semantic [PL] feature to be the core of HON and views honorification itself as plural agreement, rather than a language-specific decision that Hindi makes.

And lastly, to wrap-up our discussion of honorification and plurality, there are other Indo-Aryan languages that keep the two distinct – Magahi and Assamese. Magahi data was discussed in (9a-b). The facts of that language are more complicated (we observed that Magahi *jii* co-occurs with the plural *-an* but we note that the plural marker comes with *-waa* but not alone with *jii*) but we kept the discussion brief in this paper. Assamese data is cited from Ali (2025) – work presented in FASAL-15 conference (see Ali, this volume, for a detailed discussion). Assamese has separate plural morphemes *-hɔt and -lʊk* – that are specified for the honorific status of the pronoun they can attach to. The pattern for 2P and 3P non-honorific and honorific pronouns is given in (26).

(26) a. tɔi → tɔhɔt
2P.NH 2P.NH.PL
b. tʊmi → tʊmalʊk
2P.H 2P.H.PL

Thus, across languages of the same language family, while Hindi does not permit the honorificity lending cue *jii* to co-occur with plural suffix *-aN*, Magahi is more flexible in permitting co-occurrence of *jii* with its plural suffix and Assamese has a systematic paradigm of plurality markers that are relativized to the honorific status of the nominal referent. Therefore, cross-linguistically also we receive independent evidence to push for an honorification account that is not contingent on a non-semantic [PL] feature that is in complementary distribution with [PL] feature, but rather for an account that is sensitive to the language-specific expression of honorification.

Acknowledgement

We thank Veneeta Dayal, Rajesh Bhatt, all the FASAL-15 participants and the anonymous reviewers for helpful comments and pointers. All errors are ours.

References

Ali, Joynal. 2025. Honorificity and Number Features are Distinct Processes: A case of honorificity and plurality in Assamese pronouns. Conference Presentation presented at the Formal Approaches to South Asian Languages (FASAL)-15, The University of Texas, Austin.

Alok, Deepak. 2012. *A language without articles: the case of Magahi*. New Delhi, India: Jawaharlal Nehru University Unpublished MPhil Dissertation.

Alok, Deepak. 2014. The Morpho-Syntax of Nominal particle-wa. *Indian Linguistics* 75(3–4). 39–44. (30 August, 2025).

Alok, Deepak. 2020. *Speaker and Addressee in Natural Language: Honorificity, Indexicality, and their Interaction in Magahi*. New Brunswick, USA: Rutgers University Doctoral Dissertation.

Alok, Deepak & Ojaswee Bhalla. 2024. The Syntax and Semantics of Honorific Mismatches in Magahi. Conference Presentation presented at the 38th South Asian Languages Analysis Roundtable (SALA-38), Nagaland University, Kohima, India.

Bhalla, Ojaswee. 2022. *'to'-marked XPs in Hindi*. New Delhi, India: Jawaharlal Nehru University, New Delhi Doctoral Thesis.

Bhalla, Ojaswee & Deepak Alok. 2025. The Hindi Discourse particle ji:, Honorificity and the Form-Meaning Interaction. In *Proceedings of the 27th Seoul International Conference on Generative Grammar (SICOGG-27) 2025*, 21–30. Seoul, South Korea: The Korean Generative Grammar Circle.

Bhatt, Rajesh & Christopher Davis. 2023. Number, Honor, and Agreement in Hindi-Urdu. In *Proceedings of the Formal Approaches to South Asian Languages (FASAL)-12*. University of Utah, Salt Lake City, Utah, USA.

Brisson, Christine. 2003. Plurals,'all', and the Nonuniformity of Collective Predication. *Linguistics and philosophy*. JSTOR 26(2). 129–184.

Portner, Paul, Miok Pak & Raffaella Zanuttini. 2019. The speaker-addressee relation at the syntax-semantics interface. *Language* 95(1). 1–36. https://doi.org/10.1353/lan.2019.0008 https://muse.

Sinha, Yash. 2023. Number morphology on honorific nouns. In *Proceedings of Formal Approaches to South Asian Languages (F)ASAL-12*, 202–222. University of Utah, Salt Lake City, Utah, USA.

Wang, Ruoan. 2023. Honorifics without [hon]. *Natural Language & Linguistic Theory*. Springer Science and Business Media LLC 41(3). 1287–1347. https://doi.org/10.1007/s11049-022-09563-0.

Tiwa *khúp*: A degree modifier in the verb phrase

VIRGINIA DAWSON, *Western Washington University*
ARI ROSE, *Western Washington University*

ABSTRACT

We examine the intensifier *khúp* in Tiwa (Tibeto-Burman; India), arguing that despite its broader distribution, it is a true modifier of degrees with the same semantics as English *very*. We attribute its broader distribution and apparent flexibility in readings to its syntax: while *very* is an adjectival modifier, *khúp* modifies verb phrases, allowing it to combine with a broader range of (often derived) predicates of degrees.

1 Introduction

Intensifiers within and across languages are not a unified class, neither semantically nor syntactically. While many intensifiers, for instance, operate directly on degrees, others do not (McNabb 2012, Beltrama & Bochnak 2015). For example, while English *very* is restricted to modifying gradable predicates like *tall* (Kennedy & McNally, 2005), *really* can also modify non-gradable predicates like *dead*. This contrast is shown in (1).

(1) a. Premica is {**very**, **really**} tall.
b. The lizard is {**#very**, **really**} dead.

Intensifiers also have different syntactic distributions (Bolinger 1972, McNabb 2012). For example, while *very* is largely restricted to modifying adjectives and adverbs, *really* has a much broader distribution. For example, *really* can modify verbs, while *very* can't, as shown in (2).

(2) This {***very**, **really**} interests me.

This level of variation raises a variety of questions that warrant detailed cross-linguistic study. What are the semantic strategies for intensifying predicates? How do these semantic strategies correlate (if at all) with their syntactic distribution?

In this paper, we add to the growing literature on intensification across languages by examining the intensifier *khúp* in Tiwa, a Tibeto-Burman language of Assam, India. At a first glance, it is immediately apparent that *khúp* differs significantly from English *very* in its distribution and the range of interpretations it receives. Specifically, while *khúp* can and does modify gradable stative predicates, as shown in (3) with the gradable stative verb *chu* 'tall', it also modifies eventive predicates, as shown in (4) with the verb *misâ* 'dance'.[1]

[1]Tiwa examples are given in the orthography used in Joseph's (2014) dictionary. Abbreviations are: ACC 'accusative', ALL 'allative', DAT 'dative', EX 'existential', FOC 'focus', GEN 'genitive', IPFV 'imperfective', NEG 'negation', NEUT 'neutral aspect', NMLZ 'nominalizer', PART 'discourse particle', PFV 'perfective', SG 'singular'.

(3) Rupson **khúp** chu-w.
Rupson KHUP tall-NEUT
'Rupson is very tall.'

(4) Maria **khúp** misâ-ga.
Maria KHUP dance-PFV
'Maria danced a lot.'

We argue that despite these apparent differences, *khúp* has exactly the same semantics as English *very*: it operates directly on degrees, and thus can only combine with gradable predicates. We argue that *khúp*'s broader distribution, rather than reflecting a semantic difference, can be explained entirely in syntactic terms. Specifically, while *very* largely modifies adjectives and adverbs, *khúp* is a VP modifier, allowing it to combine with a much wider range of (often derived) predicates of degrees. The Tiwa data thus show that *very*-like intensifiers that narrowly operate on degrees can appear in the VP, rather than being restricted to adjectival modifiers. We suggest that Tiwa's tendency to lexicalize gradable stative predicates as verbs, rather than adjectives, is the likely source of this syntactic variation in degree-modifiers.

The rest of this paper is structured as follows. In section 2, we provide some background on Tiwa, the data, and the distribution and interpretation of *khúp*. In section 3, we provide evidence that despite its relatively broad distribution, *khúp* can only modify predicates of degrees. We present our analysis in section 4, adopting Kennedy & McNally's (2005) analysis of *very* as a degree modifier, and showing how it captures the interpretation of *khúp* with the broader range of predicates it appears with. We also provide evidence for *khúp*'s syntactic position with the VP. We briefly conclude in section 5.

2 Background on Tiwa and *khúp*

Tiwa is a Tibeto-Burman language of the Boro-Garo subgroup (Post & Burling 2017), spoken by over 30,000 people primarily in West Karbi Anglong district, Assam, India.[2] The data presented here come from the first author's original fieldwork with four speakers of Tiwa in Umswai, Assam, between 2015 and 2023. The data come primarily from elicitation, with key semantic data gathered using the methodology of semantic fieldwork laid out by Matthewson (2004).

As introduced above, Tiwa has an intensifier *khúp* that in many cases gets translated by speakers as English 'very', with a typical degree-modifier reading. Indeed, *khúp* freely combines with gradable stative predicates like *chu* 'tall' (shown in (3) above), *kró* 'good', *chál* 'far' and *hóma* 'hungry', as shown in (5)-(7) below. Note that syntactically, these predicates are all verbs, rather than adjectives; Tiwa has very few true adjectives (Dawson, 2020). We will return to this point in section 4 below.

[2]The 2011 census lists 33,900 Tiwa speakers.

(5) Taglí mai **khúp** krói-do.
this.year paddy KHUP good-IPFV
'The paddy this year is very good.'

(6) Pe tes **khúp** chál-do.
3SG country KHUP far-IPFV
'That country is very far.'

(7) Ang **khúp** mai hóma-ga.
1SG KHUP rice hungry-PFV
'I am very hungry.'

However, unlike *very*, *khúp* has a much broader distribution. Most strikingly, it can appear with non-stative predicates like *misâ* 'dance', as shown in (4) above, and *krá* 'cry', as shown in (8).

(8) Pe korkhyá **khúp** krái-do.
3SG child KHUP cry-IPFV
'That child is crying so much.' [OM 2022.2.12]

Eventive predicates modified by *khúp* can receive a couple of different readings. Specifically, they can receive a reading in which there is a single event that can be measured in some way as exceeding a contextually determined standard, or one on which there are, contextually speaking, many events. For example, in (8), *khúp*, in combination with the imperfective marker *-do*, yields a reading on which there is a single ongoing event of crying that is particularly intense in the context. In contrast, in (9), the predicate is marked with 'neutral aspect' which, when combined with non-stative predicates, allows for habitual readings (Dawson, 2020). This sentence allows for a reading which conveys that the child cries frequently: it is the number of crying events that is noteworthy in the context, not their intensity.

(9) Pe korkhyá **khúp** krá-w.
3SG child KHUP cry-NEUT
'That child cries so much.'

A variety of examples with *khúp* modifying non-stative predicates are given in (10)-(14) below, illustrating the wide range of predicates that it can combine with.

(10) Ang **khúp** pe khûri-go phí-ga.
1SG KHUP that cup-ACC break-PFV
'I broke that cup so many times.'

(11) Ang **khúp** torgâ aw-ga.
1SG KHUP door open-PFV
'I opened the door so many times.'

(12) Ang **khúp** kó-ga.
1SG KHUP fall-PFV
'I fell down so many times.'

(13) Saldi **khúp** plaw-ga.
Saldi KHUP forget-PFV
'Saldi has forgotten so much.'

(14) Ang kashóng **khúp** pre-ga.
1SG dress KHUP buy-PFV
'I bought a lot of dresses.'

At first glance, these data suggest that *khúp* is not limited to operating on degrees. The readings in (4) and (9)-(14) in particular suggest that *khúp* can target events and entities, as well as degrees. Indeed, it's known that languages can use different semantic means to intensify predicates, which can lead to a wider range of uses. While intensifiers like *very* operates directly on degrees, intensifiers like Hebrew *mamaš*, English *really*, Washo *šemu*, and Italian *-issimo* likely operate on contexts (McNabb 2012, Beltrama & Bochnak 2015). This different core semantics leads to a wider distribution for these intensifiers than degree-operators like *very*. The data presented in this section suggest that such an analysis might be appropriate for *khúp* as well. We argue in the next section, however, that despite its wider distribution, *khúp* does directly operate on degrees, and further that it can and should be analyzed as having the same semantics as *very*.

3 *khúp* is a degree modifier

In section 2, we showed that *khúp* appears with a wider range of predicates than *very* does. Specifically, in addition to appearing with gradable stative predicates, it also appears with a wide variety of non-stative predicates. However, *khúp* is not as unrestricted in its distribution as other intensifiers like English *really* and Hebrew *mamaš*. Most notably, unlike those other intensifiers, *khúp* cannot be used to emphasize the truth of a proposition. This can be seen clearly in the dialog in (15). Here, speaker (a) asserts that Samsing went home, but speaker (b) disagrees and asserts the opposite. While an intensifier like *really* can be used in contexts such as these to emphasize the truth of speaker (a)'s original assertion, *khúp* is judged infelicitous.

(15) a. Samsing nó-jîng lí-ga.
Samsing house-ALL go-PFV
'Samsing went home.'

b. Cha, lí-wa-n' cha.
NEG.EX go-NMLZ-GEN NEG.EX
'No, he hasn't gone.'

a. # **Khúp** lí-ga.
KHUP go-IPFV
Intended: 'He *really* did go.'

Similarly, speakers also judge *khúp* infelicitous when it is combined with a predicate like *thi* 'die', at least in an out of the blue context. Instead, the only reading that sentences like (16) can receive is the impossible one on which the subject died many times, in line with the multiple-event readings described in section 2.

(16) # Rupson **khúp** thi-ga.
Rupson KHUP die-PFV
Intended: 'Rupson really did die.'
Only reading: 'Rupson died many times.'

Note in contrast that if the subject is not singular,[3] *khúp* freely combines with *thi* 'die', as it easily allows a reading on which the number of creatures that died exceeds contextual standards, as shown in (17).

(17) Khúgri-râw **khúp** thi-ga.
dog-PL KHUP die-PFV
'So many dogs died.'

We've seen that *khúp* has a broader distribution than *very*, but a narrower distribution than non-degree-based intensifiers like *really.* Taken together, what all the predicates that *khúp* can modify have in common is that they are gradable: all the predicates that *khúp* can modify can also appear in comparatives, which is a standard diagnostic for gradability (Kennedy & McNally, 2005; Beltrama & Bochnak, 2015). This includes lexically gradable stative predicates (i.e., those that are inherently type $\langle d, et \rangle$), like *chu* 'tall', but also includes eventive predicates and stative predicates that are (presumably) not lexically gradable. For example, the non-stative verb *lí* 'go' can both appear in comparative constructions, as shown in (18), as well as be modified by *khúp*, as shown in (19). The comparative compares the number of times that the subject went to Guwahati; the effect of *khúp* is to signal that the subject went to Guwahati many times.

(18) Ang Guwahati-jíng konông-a khúli **parâ** lí-ga.
1SG Guwahati-ALL other-DAT than more go-PFV
'I went to Guwahati more than the others.'

(19) Ang Guwahati-jíng **khúp** lí-ga.
1SG Guwahati-ALL KHUP go-PFV
'I went to Guwahati so many times.'

[3]That is, the subject is either explicitly marked plural, or a bare non-human noun (which are number neutral in Tiwa; Dawson & Gibson to appear).

Similarly, the possessive construction, which is presumably not lexically gradable in the way a predicate like *chu* 'tall' is, can appear in both comparatives, as in (20), and modified by *khúp*, as in (21). In the comparative, the quantity of possessed objects are being compared. With *khúp*, this quantity is being signaled as high, in the context.

(20) Context: The consultant is asked who has more eggs, Lastoi or Saldi?

Lastoi-ne **parâ** tudí tong-o.
Lastoi-GEN more egg exist-NEUT
'Lastoi has more eggs.'

(21) Rupson-e **khúp** tudí tong-o.
Rupson-GEN KHUP egg exist-NEUT
'Rupson has a lot of eggs.'

Comparatives like those in (18) and (20) involve quantification over degrees. Given *khúp*'s semantics as an intensifier, its inability to get veridicality readings like the ones attempted in (15) and (16) above, and the fact that it patterns with comparatives in what predicates it modifies, we assume that *khúp* operates directly on degrees.

Before turning to the details of our analysis, it's worth noting that while *khúp* is a degree modifier, and can combine with any predicate that can appear in a comparative, *khúp* itself does not have comparative semantics. This is evident from the infelicity of sentences with *khúp* in scenarios like the one in (22). In this context, Rupson is objectively taller than the people he is around, but only by one centimeter. This small difference in height is not enough to license *khúp*, as shown in (22a), but it is enough to license a comparative, as shown in (22b).[4]

(22) Context: Rupson is 135cm, and he's in a group with other people between 130 and 134cm.

a. # Rupson **khúp** chu-w.
Rupson KHUP tall-NEUT
'Rupson is very tall.'

b. ✓ Rupson **parâ** chu-w.
Rupson more tall-NEUT
'Rupson is taller.'

Rather than expressing comparison between the degree to which the predicate holds true and some contextually determined standard, *khúp* instead requires that the predicate hold to a much higher degree than that standard, just like English *very*.

[4]Note that 'crisp judgment' data like these provide clear evidence that comparative constructions in Tiwa do involve quantification over degrees (unlike some other languages; Beck et al. 2009). See Dawson 2020 and Dawson 2021 for discussion and analysis of comparatives in Tiwa.

4 Analysis

In the previous section, we established that *khúp* shows one of the key restrictions that *very* does; namely, it can only combine with gradable predicates. In section 2, however, we saw that *khúp* has a wider distribution than *very*, in that it can appear with a wide range of non-stative predicates. In this section, we develop an analysis that explains *khúp*'s distribution, arguing that semantically, it conveys that a gradable predicate holds to a degree significantly higher than the contextual standard, just like *very,* but that it differs in its syntax. Specifically, we argue that while *very* is largely restricted to modifying adjectives and adverbs, *khúp* modifies VPs, allowing it to appear with a much wider range of gradable predicates than *very*.

4.1 Basic semantic analysis

Given that *khúp* is an intensifier that (i) can only combine with gradable predicates, and (ii) requires that the predicate hold to a degree significantly higher than the contextual standard, we assume that *khúp* has the same semantics as a modifier of predicates of degrees like *very*. Specifically, we adopt for *khúp* Kennedy & McNally's (2005) analysis of *very* in proposing that *khúp* combines with a type $\langle d, et \rangle$ gradable predicate and yields a predicate which holds to a degree significantly above a contextually determined standard. A formal definition for *khúp* is given in (23), following Beltrama & Bochnak (2015).

(23) $[\![\text{khúp}]\!]^c = \lambda G_{\langle d,et \rangle}.\lambda x.\exists d[G(x, d) \wedge d >! \text{standard}(G, c)]$

This analysis straightforwardly results in the desired reading for gradable stative predicates, like *chu* 'tall', as shown in (24), which provides the truth conditions of (3), repeated here from the introduction. Sentence (3) holds true if there is some degree to which Rupson is tall that greatly exceeds what counts as tall in the context.

(3) Rupson **khúp** chu-w.
Rupson KHUP tall-NEUT
'Rupson is very tall.'

(24) $[\![(3)]\!]^c = \exists d[\text{tall}(\text{Rupson}, d) \wedge d >! \text{standard}(\text{tall}, c)]$

A key supporting piece of evidence for this approach comes from *khúp*'s behavior with maximum-standard predicates. As Kennedy & McNally (2005) note, *very* sounds degraded when it combines with a maximum-standard predicate. This follows from the analysis above: maximum-standard predicates are true only if they hold to the maximum degree on their scale; it's therefore odd to modify these predicates to say they hold to some significantly higher degree than that. Just like *very, khúp* is judged somewhat degraded when it combines with maximum-standard predicates. Example (25) shows that *khúp* is judged felicitous with the minimum-standard predicate *ler* 'wet'. In contrast, the examples in (26) show that *khúp* is judged degraded with the maximum-standard predicates *rán* 'dry' and

porê 'full'.[5]

(25) Pe ré **khúp** ler.
that cloth KHUP wet
'That cloth is very wet.' [minimum-standard]

(26) a. ?? Pe ré **khúp** rán-ga.
that cloth KHUP dry-PFV
?? 'That cloth is very dry.' [maximum-standard]

b. ?? Hêbe khûri **khúp** porê-ga.
this cup KHUP full-PFV
?? 'This cup is very full.' [maximum-standard]

Speakers instead have offered alternative ways to emphasize how completely the predicate holds true, like the sentences in (27), which use focus marking and a discourse particle.

(27) a. Pe ré rán-ga**-lô-bó**.
that cloth dry-PFV-FOC-PART
'That cloth is completely dry.'

b. Hêbe khûri porê-ga**-lo-bó**.
this cup full-PFV-FOC-PART
'This cup is completely full.'

Note that data like these also provide evidence against an analysis of *khúp* as an intensifier that operates on contexts (cf. McNabb 2012, Beltrama & Bochnak 2015): such intensifiers, including *really,* readily combine with maximum-standard predicates.

4.2 Non-stative predicates

Predicates like *chu* 'tall' are inherently gradable: they are type $\langle d,et\rangle$ and have lexically-determined scales. Not every predicate is lexically gradable, however, including the eventive predicates that we saw *khúp* combines with in section 2. While predicates like *misâ* 'dance' and *lí* 'go' are not lexically $\langle d,et\rangle$, they can appear in comparative constructions, which require a type $\langle d,et\rangle$ predicate (Cresswell 1977, von Stechow 1984, among many others). That such predicates can appear in comparatives suggests that there is some silent degree operator present in comparative constructions that derives a type $\langle d,et\rangle$ predicate from what was otherwise not a predicate of degrees. We assume that this same silent degree operator that is present in comparative constructions is available for predicates that combine with *khúp,* which likewise combines with a $\langle d,et\rangle$ predicate.

We leave the compositional details of this silent degree operator to future work, as they are relevant for degree-constructions more broadly in Tiwa. For our purposes, it is enough

[5] As in English, Tiwa speakers give variable judgments on these, rather than judging them completely infelicitous. We assume there's a pragmatic halo effect, where what counts as dry and full is variable according to context (Kennedy & McNally 2005).

to note that whatever derives a $\langle d, et \rangle$ predicate from a non-gradable predicate in comparatives and other degree constructions can serve the same purpose for *khúp*. Whatever the compositional details, the resulting semantics for non-stative predicates in comparative constructions and with *khúp* are comparable, as shown in (28) and (29).

(18) Ang Guwahati-jíng konông-a khúli **parâ** lí-ga.
1SG Guwahati-ALL other-DAT than more go-PFV
'I went to Guwahati more than the others.'

(28) $[\![(18)]\!]^c$ = ∃d[go(I, G., d-times) & ¬go(others, G., d-times)]

(19) Ang Guwahati-jíng **khúp** lí-ga.
1SG Guwahati-ALL KHUP go-PFV
'I went to Guwahati so many times.'

(29) $[\![(19)]\!]^c$ = ∃d[go(I, G., d-times) & d >! standard(go(G.), c)]

In both cases, the number of times that the subject went to Guwahati serves as the gradable component that the comparative and *khúp* can target.[6]

4.3 Differing distributions

We have argued above that *khúp* and *very* have the same core semantics. We have also seen, however, that *khúp* has a much broader distribution than *very.* Specifically, while *khúp* can combine with any gradable predicate – including those that are lexically gradable, and those that are derived – *very* cannot. Instead, *very* can (for the most part) only occur with lexically gradable predicates like *tall.* Rather than attributing this difference in distribution to semantics, we propose instead that it is syntactic: while *very* modifies adjectives (and adverbs), as represented in (30), *khúp* modifies verb phrases, as represented in (31).

(30) Rupson [$_{VP}$ is [$_{AP}$ **very** tall.]]

(31) Rupson [$_{VP}$ **khúp** chu-w.]
Rupson KHUP tall-NEUT

This syntactic difference explains the differing distributions of *very* and *khúp*, and in particular explains why *khúp* has a much broader distribution than *very:* a much wider range of predicates can appear in verb phrases than in adjective phrases, and they allow for whatever operator is responsible for deriving gradable predicates from non-gradable ones.

The most obvious piece of evidence that *khúp* is not an adjectival modifier is that the vast majority of predicates that it combines with do not contain an adjective; instead, they are verbal. This includes most gradable stative predicates, like *chu* 'tall' and *phung* 'fat', which inflect for tense and aspect, and do not appear with the copular verb (Dawson, 2020). For example, we saw that *chu* 'tall' is inflected with neutral aspect in (3). Example (32) shows *phung* 'fat' inflected with perfective aspect, yielding an inchoative reading.

[6]Given the available readings, this analysis would ideally be implemented in an event semantic framework, but we leave that to future work.

(32) Sonasing **khúp** phung-ga.
Sonasing KHUP fat-PFV
'Sonasing has become very fat.'

Most gradable stative predicates in Tiwa are syntactically verbs, but there are a small number of true adjectives which cannot inflect for tense and aspect, and which occur with the copular verb when in predicative position (Dawson, 2020). True gradable adjectives in Tiwa, like *kumún* 'well', can be modified by *khúp*, as shown in (33).

(33) Ang **khúp** kumún hóng-do.
1SG KHUP well COP-IPFV
'I am very well.'

We assume that in these cases, *khúp* is modifying the entire VP which contains the adjectival predicate and the copular verb *hóng* 'be'.[7]

It's clear that *khúp* is not restricted to modifying adjectives. Instead, we've claimed that it modifies the verb phrase. There are two main pieces of evidence that *khúp* is a verb phrase adjunct, rather than merging higher in the structure (e.g., in AspP or TP). The first is that default word order places *khúp* between the subject and any (non-subject) internal arguments of the verb (i.e., on the left edge of the VP). While more marked word order variations are possible, we assume these are derived by scrambling.

The second piece of evidence that *khúp* is located in the VP comes from a split in how *khúp* interacts with internal vs. external arguments. Specifically, only internal arguments can be targeted in deriving a degree predicate that *khúp* modifies, suggesting that *khúp* merges before external arguments. This contrast is clearly seen in the different readings available to unergative and unaccusative predicates modified by *khúp*. As shown in (34), when *khúp* modifies unaccusative predicates like *phi* 'come', there are several distinct readings available. On one reading, *khúp* is interpreted as intensifying the number of coming events that a single dog participates in. On the other reading, which is key for us, it is interpreted as intensifying the number of dogs that participate in the coming event.

(34) Khúgri$_i$ [$_{VP}$ **khúp** t_i phi-do.]
dog KHUP come-IPFV
✓ 'The dog is coming a lot.'
✓ 'Many dogs are coming.' [unaccusative predicate]

In contrast, when *khúp* modifies unergative predicates like *shúng* 'bark', as in (35), it can only receive a reading on which a single dog barks a lot; *khúp* cannot yield a reading that targets the subject.

(35) Khúgri [$_{VP}$ **khúp** shúng-do.]
dog KHUP bark-IPFV

[7]Note that the copular verb can be omitted in Tiwa in present tense. This has happened in example (25) above, whose main predicate is a syntactic adjective *ler* 'dry'. We assume that the copula is syntactically present in these cases, if unpronounced.

✓ 'The dog is barking a lot.'
✗ 'Many dogs are barking.' [unergative predicate]

This contrast between unaccusative and unergative predicates is explained if *khúp* adjoins in the VP, merging before any external arguments, but able to affect the interpretation of internal arguments.

We've argued in this section that the distributional differences between *khúp* and *very* follow from their syntax. While *khúp* is a VP modifier, and thus can occur with a wider array of predicates, *very* modifies adjectives and adverbs. Minimal pairs like those in (36)-(38) suggest that this limitation of *very* is a syntactic restriction, not a semantic one.

(36) a. This [$_{VP}$ is [$_{AP}$ (very) interesting to me.]]
b. This [$_{VP}$ (*very) interests me.]

(37) a. I will [$_{VP}$ get [$_{AP}$ (very) tired.]]
b. I will [$_{VP}$ (*very) tire.]

(38) a. My dog [$_{VP}$ is [$_{AP}$ (very) barky.]][8]
b. My dog [$_{VP}$ (*very) barks.]

While the (a) and (b) sentences in these examples contain the same base predicate and have very nearly the same meaning, *very* can only appear with the predicate in its adjectival form.

5 Conclusion

In this paper, we've argued that despite its apparent flexibility in the types of predicates it appears with, *khúp* operates directly on degrees. Specifically, we have argued that it has the same semantics as a degree-modifier like *very*, but that it differs in its syntactic placement. This analysis captures both *khúp*'s similarities to and differences from *very*: its identical semantics ensures it can only combine with gradable predicates and sounds degraded with maximum-standard predicates, while its differing syntactic position explains why it combines with a far wider array of gradable predicates, including those that are derived.

Overall, the Tiwa data show that degree-modifiers of the *very*-type can appear in verb phrases, suggesting that *very*'s distributional restrictions follow from syntactic constraints rather than semantic ones. That *khúp* has such a different syntax from *very* is not surprising, as the vast majority of lexically gradable stative predicates in Tiwa are syntactically verbs. Together with an independently necessary silent operator that derives degree predicates that can be used in comparatives, *khúp*'s status as a verb phrase modifier accounts for its ability to modify eventive predicates as well as gradable stative predicates.

[8]This is a naturally occuring example, found at https://forums.digitalspy.com/discussion/334362/barking-dog (last accessed on August 18, 2025).

Acknowledgements

Thanks to our Tiwa teachers Bibiana Maslai, Mary Maslai, Olphindro Malang, and the late Juliana Maslai. Thanks also to audiences at Western Washington University and FASAL 15, as well as two anonymous FASAL abstract reviewers, for helpful comments and feedback. Any errors are ours alone.

References

Beck, Sigrid, Sveta Krasikova, Daniel Fleischer, Remus Gergel, Stefan Hofstetter, Christiane Savelsberg, John Vanderelst & Elisabeth Villalta. 2009. Crosslinguistic variation in comparison constructions. *Linguistic Variation Yearbook* 9(1). 1–66.

Beltrama, Andrea & Ryan Bochnak. 2015. Intensification without degrees cross-linguistically. *Natural Language and Linguistic Theory* 33. 843–879.

Bolinger, Dwight. 1972. *Degree words*. De Gruyter.

Cresswell, Max J. 1977. The semantics of degree. In Barbara Partee (ed.), *Montague grammar*, 261–292. New York: Academic Press.

Dawson, Virginia. 2020. *Existential quantification in Tiwa: Disjunction and indefinites*: University of California, Berkeley dissertation.

Dawson, Virginia. 2021. Disjunction as alternatives: evidence from phrasal comparatives. In Rachel Soo, Una Y. Chow & Sander Nederveen (eds.), *Proceedings of the 38th West Coast Conference on Formal Linguistics*, 158–168. Somerville, MA: Cascadilla Press.

Dawson, Virginia & Royce Gibson. to appear. The pragmatics of optional plural in Tiwa. In *Proceedings of WCCFL 42*, Cascadilla.

Joseph, U.V. (ed.). 2014. *Tiwa-English dictionary with English-Tiwa index*. Shillong: Don Bosco Centre for Indigenous Cultures. Associate editors: Horsing Kholar, Juliana Maslai, Alfred Maslai, Bibiana Maslai, and Simon Mithi.

Kennedy, Christopher & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81(2). 345–381.

Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70(4). 369–415.

McNabb, Yaron. 2012. Cross-categorial modification of properties in Hebrew and English. In Anca Chereches (ed.), *Proceedings of Semantic and Linguistic Theory 22*, 365–382. Linguistic Society of America.

Post, Mark W. & Robbins Burling. 2017. The Tibeto-Burman languages of Northeast India. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 213–242. Routledge 2nd edn.

von Stechow, Arnim. 1984. Comparing semantic theories of comparison. *Journal of Semantics* 3(1). 1–77.

The particle '*em*' in Gujarati

SAMPADA DESHPANDE, *UT Austin*

ABSTRACT

This paper investigates the discourse particle *em* in Gujarati, which derives a specialized utterance-final discourse use across declarative and interrogative clauses from its origin as distal demonstrative in the language. I show that *em* contributes a unified discourse function: it foregrounds the prejacent proposition, marking it salient relative to a set of alternatives. However, its discourse effects vary depending on the clause types and contextual conditions it occurs in. In declaratives, *em* serves as a specification strategy that selects a particular alternative from a set of alternatives that over-answer the primary QUD. This move triggers and answers a related QUD to the primary QUD, thereby already answering the primary QUD. In interrogatives, *em* can be used to seek neutral confirmation or convey the speaker's negative epistemic bias towards the prejacent. *em* declaratives cannot be felicitously used to answer the primary QUD and *em* interrogatives are felicitous only when the context provides direct evidence for the prejacent and the speaker is negatively biased towards it. The analysis is framed within a commitment-based model of discourse structure (Farkas & Bruce, 2010; Roberts & Rudin, 2024), which captures how *em* adjusts the interlocutors' commitments to the propositions on the Table. I propose that *em* questions signal the speaker's assessment of whether the contextual evidence for the prejacent is adequate to warrant committing to it. This is formalized by introducing an adequacy value within the representation of evidence while placing the issue $\{p, \neg p\}$ on the conversational table.

1 Introduction

This study presents an analysis of the discourse particle *em* in Gujarati, which derives its specialized utterance-final discourse use from a distal demonstrative (1) in the language.

(1) [Context: Arun and Bina are shopping at the mall. At 5:00 pm, Bina leaves to go for a walk in the park next to the mall. At 5:30 pm, Mina comes to the mall and asks Arun where Bina went. Arun responds]

bina **em** baʤu ʧal-e ʧʰe.
Bina DEM side walk-IPFV be.PRS

Bina is walking on that side. (*em* as a demonstrative)

em is literally translated as *like that* in English, but in its utterance-final use: in interrogatives, it has a function similar to the same polarity tag (see Malamud & Stephenson (2015)) as in (2) and can be translated as 'you say'. In declaratives, it has a function similar to (but

not exactly) the reformulation marker (Cuenca, 2003) as in (3a) and can be translated as 'that is' as in (3b).

(2) It's raining, is it?

(3) a. ...in determining the point on the scale between them which is valid for the case in question. To be more precise, the extent to which a text is translatable...

b. This book has everything, everything good, that is. [1]

Rather than occurring in interrogative clauses per say, *em* gives rise to an interrogative strategy that emerges from the combination of a declarative with a falling pitch contour and a tag with a rising pitch contour, much like the tag-questions in English, which are classified as non-canonical questions.

em is used in polar questions (4) and declaratives (5), but not in wh-interrogatives.[2] A cross-linguistic observation is that *em*'s demonstrative counterparts in the neighboring Indo-Aryan languages, such as Hindi *vəisa* and Marathi *təsə* do not appear to exhibit these discourse uses. Thus, *em* presents a unique test case for theories of discourse particles.

(4) rija g^{h}ər=e gə-ji *em* (↑)
riya home=LOC go-PFV.F.SG em
Riya went home, Is it/ Did she?.

riya g^{h}ər gə-ji *vəisa* (↑) (Hindi)
riya g^{h}ər=i ge-li *təsə* (↑) (Marathi)

(5) rija g^{h}ər=e gə-ji, mira=na, *em* (↓)
riya home=LOC go-PFV.F.SG mira=GEN em
Riya went home, Mira's, that is.

I show that the utterance-final particle *em*, has a unified primary function across clause types: to foreground the prejacent proposition. Additionally, it has different discourse functions depending on the clause type in which it occurs. In declaratives, it can be used to pick out an entity from a set of alternatives as a specification move, as in (5). The *em* strategy is only felicitous when used to to trigger and answer a QUD related to the primary QUD. In interrogatives, it is used for neutral confirmation of the prejacent proposition or it is used to convey that the speaker is biased against the prejacent proposition. The felicity of utterance-final *em* is conditioned by the prior discourse context, which must provide direct evidence for prejacent and the speaker's negative epistemic bias. The use of an *em* question is associated with a diversity of pragmatic flavors of rhetoricity, skepticism, surprise and confrontation.

[1] All English examples are from COCA: Corpus of Contemporary American English- https://www.english-corpora.org/coca/

[2] The Gujarati data come from the following sources: (a) Leipzig Corpora Collection- https://corpora.uni-leipzig.de/en?corpusId=guj_newscrawl_2014_1M (b) Concordance: https://www.sketchengine.eu/gujarati-web-corpus/; (c) My own judgements with verification from other Gujarati speakers.

2 Empirical profile of *em*

The different discourse effects of *em* arise due to the interaction between the conventionally encoded meaning of *em*, varying contextual settings, and the conventional meaning associated with the clause types of interrogatives and declaratives.

2.1 Declaratives

Declarative clauses combined with the utterance-final *em* exhibit the pragmatic effect of a specification move / clarificatory move where *em* highlights the prejacent as offering a more specific answer than what is strictly required. By doing so, the speaker guides the listener to the speaker's preferred /intended interpretation of the discourse move, ie., narrowing the range of relevant alternatives and avoids potential for misinterpretation. An *em* declarative triggers and answers a related QUD and in effect indirectly answers the primary QUD.

2.1.1 Specification move / Clarificatory move

em is used to contrast the set of alternatives, where it flouts the maxims (Grice (1975)) of (1) quantity, where the speaker intentionally provides more information than minimally necessary, and (2) relevance, where the speaker chooses to answer a related QUD and indirectly answers the primary QUD. *em* is used to highlight and select the least expected answer within the contextually salient set of alternatives to the question under discussion rather than the expected answer[3]. In such contexts, the use of *em* signals that the speaker anticipates a potential misunderstanding and controls the interpretation of the response accordingly. It behaves similarly to the English *I mean*, as an explicit repair or specification strategy (6a). In (6b), use of another discourse particle in the language, *hə̃* (which behaves like the English *okay*, is demonstrated. *hə̃* is also used to select the unexpected answer to the QUD, but this strategy does not over-answer, but rather directly answers the QUD. It is important to note that the pragmatic considerations on the speaker's perspectives differ in the use of *em* versus *hə̃*. *em* can be felicitously used in contexts where the speaker preempts potential misinterpretation of the discourse move whereas *hə̃* is felicitously used in contexts where the speaker is aware of the divergence in the speaker's and addressee's attitude towards the proposition and hence this strategy is used to convey speaker attitude.[4]

(6) [Context 1: Bina loves khaman, and has bought some from a special shop. Mina informs Arun:]
[Context 2: Arun thinks Bina has no taste in food. Mina wants to prove Arun wrong by conveying that Bina shops at the best sweet shop in town:]

[3]This function of highlighting the alternative that is least expected as opposed to the most expected alternative as an answer to the QUD might relate to the property of *em* being a distal demonstrative in the language

[4]The empirical and theoretical complexity of this particle warrants a dedicated in-depth investigation to be undertaken in separate future work.

a. bina=e kʰəməɳ lav-ja, balubʰai=na, *em*
bina=ERG khaman.PL bring-PFV.PL balubhai=GEN.PL em
Bina brought khaman, from Balubhai's (shop), that is.
Context 1: [✓] and Context 2: [✗]

b. bina=e kʰəməɳ lav-ja, balubʰai=na, *hə̃* [5]
bina=ERG khaman.PL bring-PFV.PL balubhai=GEN.PL hə̃
Bina brought khaman, from Balubhai's (shop), okay.
Context 1: [✗] and Context 2: [✓]

In context 1, *em* provides specific information that strictly over-answers the question under discussion (QUD). The primary QUD is: 'What did Bina buy?' and the expected minimal answer would simply state the object 'kʰəməɳ'. But instead, when the speaker adds further information (here, Balubhai's shop) appending it with *em*, this move triggers a related QUD: 'Where did Bina buy it from?' and answers that. By doing so, the *em* declarative answers the primary QUD indirectly. In context 2, where the source of the 'kʰəməɳ' (Balubhai's shop) is at-issue and necessary for answering the primary QUD, *em* is infelicitous. This contrast shows that *em* is felicitous only when it contributes additional discourse content beyond a direct answer to the primary QUD, in other words, *em* is licensed only when it is used to indirectly address the primary QUD, not when it directly resolves it.

The example below illustrates the same pattern, *em* is felicitous only in context 1 where the response is an over-answer to the primary QUD. By mentioning 'the time of flight' adds specificity beyond merely asserting that the speaker will be traveling. In context 2, the response directly addresses the primary QUD and is infelicitous. Thus, the specification move with *em* selectively foregrounds secondary/ related information in response to discourse needs.

(7) [Context 1: Bina is Arun's student and asks him if they can meet to discuss her work the next evening. Arun responds:]
[Context 2: Bina and Arun are friends and Bina knows that Arun is traveling soon. She asks him when he is traveling and Arun responds:]

ma=ri kale flaiʈ ʧʰe, 6 wagje, *em*
1.SG=GEN.3.F.SG tomorrow flight be.PRS 6 o'clock em

I have a flight tomorrow, at 6 o'clock, that is.
Context 1: [✓] and Context 2: [✗]

In (8a) *em* signals that the speaker wants the preceding content to be interpreted in a specific way. By using the *em* strategy, the speaker ensures that 'he doesn't know me' is interpreted narrowly as in, 'that he did not know her from before, rather than implying that 'they don't know each other at all' which would be the broader interpretation.

[5]Utterance-final expressions like Hindi *ʈʰik hai na* and Marathi *bərə/ bərə ka* seem to have the same discourse effect.

(8) [Context: Mina is asking Bina if she and Arun know each other well. Bina says:]

a. e mə=ne nət^{h}i oɭək^{h}-to, pehla=t^{h}i, *em*
3.SG.M 1.SG=DAT/ACC NEG know-IPFV.3.SG.M earlier=from em
He doesn't know me, from before, that is.

b. e mə=ne nət^{h}i oɭək^{h}-to, *eʈle*, pehla=t^{h}i
3.SG.M 1.SG=DAT/ACC NEG know-IPFV.3.SG.M means earlier=from
He doesn't know me, I mean, from before.
[alternative strategy- more productive]

This section described the use of *em* in declarative clauses, a strategy used for making a specification/clarificatory discourse move.

2.2 Interrogatives

em does not occur in interrogative clauses, but rather it creates an interrogative strategy by combining a declarative with a falling pitch contour and a tag (*em*) with the rising pitch contour, much like tag-questions in English, which are analyzed as non-canonical questions.

2.2.1 Confirmatory *em* with an explicit prejacent

When the prior discourse provides explicit evidence for the prejacent proposition, *em* can be used in order to seek confirmation from the addressee as to whether the prejacent proposition is true.

(9) [Context: Arun and Bina are at a party with loud music. Mina, a mutual friend of Arun and Bina, just texted Bina that she urgently needs to go to Mumbai the next day. Bina tells Arun this and he asks to clarify:]

e kale mumbəi ʤa-j ʧ^{h}e ***em***?
3.SG tomorrow mumbai go-IPFV be.PRS em

She is going to Mumbai tomorrow, is she?

em is felicitous in a neutral context where the speaker has no prior positive or negative epistemic bias and requests to confirm if the prejacent proposition is true. The speaker has no commitment to p or $\neg p$.

2.2.2 Confirmatory *em* with an implied prejacent

The particle *em* sometimes, but not always, behaves like the English Same Polarity-tags. Specifically, *em* is not felicitous in contexts where the prejacent is implied, and not explicitly available in the prior discourse unlike the English Sp-tags.

Consider the 'Blushing/Innuendo context' from Malamud & Stephenson (2015), where the prejacent *p* is implied in the prior discourse and an uninformed speaker makes a guess about the hearer's commitment and in these contexts, SP-tags are permissible. The addressee's judgement is at issue, not the speaker's. This suggests that the SP-tags involve independent commitments of the addressee and may/may not involve dependent commitments of the speaker.

(10) [**Blushing/Innuendo context:** A and B are gossiping. A doesn't know anything about B's neighbor. B says, blushing, 'You've GOT to see this picture of my new neighbor!' Without looking, A replies:]

a. A: He's attractive, is he?

b. # A: He's attractive, isn't he?

c. # A: He's attractive.

em is unavailable in such a context by itself where the prejacent is implied and not explicitly provided by the context. In these contexts, *em* needs to be supported by other expressions as they connect the prejacent to the larger QUD operative in the previous discourse, thus making the prejacent available for the particle *em* to combine with. This is demonstrated below:

(11) [Context: Asha and Bina had made plans to go to a movie tonight and had purchased tickets the previous week. Asha tells Bina that she is going to a cafe with Dhruv instead. Bina is surprised and asks:]

tu e=ne ɖeʈ kər-e ʧʰe ***em***?
2.SG 3.SG=DAT/ACC date do-IPFV be.PRS em

You are dating him, are you?

In (11), the absence of explicit prior evidence about 'dating' renders the utterance-final *em* infelicitous. However, *em's* felicity can be restored when used with the wh-question *kem* in (12a), the discourse markers *eʈle* in (12b) and *to* in (12c). [6]

(12) a. **kem** tu e=ne ɖeʈ kər-e ʧʰe ***em***?
why 2.SG 3.SG=DAT/ACC date do-IPFV be.PRS em
Why? You are dating him, are you?

[6]It is also felicitous when used with hybrid- interrogative strategies (see Deo (2023)), but this needs further exploration.

(1) tu e=ne su ɖeʈ kər-e ʧʰe *em*?
2.SG 3.SG=DAT/ACC what date do-IPFV be.PRS em
You are dating him, are you?

b. **eʈle** tu e=ne ɖeʈ kər-e ʧʰe ***em***?
means 2.SG 3.SG=DAT/ACC date do-IPFV be.PRS em
Meaning, You are dating him, are you?

c. **to** tu e=ne ɖeʈ kər-e ʧʰe ***em***?
so 2.SG 3.SG=DAT/ACC date do-IPFV be.PRS em
So, You are dating him, are you?

In this subsection, it can be observed that *em* can be felicitously used in interrogatives for confirmation of the prejacent in contexts where the speaker is unbiased/neutral towards the prejacent, given that the evidence for the prejacent is explicitly provided by the prior context or by necessarily using explicit discourse connectives in the contexts where the prejacent is implied. Now, let us explore the use of *em* in contexts where the speaker is biased in the next subsection.

2.2.3 *em* and Bias

There is another layer of complexity in the felicitous use of *em* when we consider how it is affected by varying speaker's epistemic bias (positive, negative) and when the speaker is unbiased (Goodhue (2022); Sudo (2013)).

(13) [Context 1: Arun is disinterested in cooking as an activity and his friend Bina is aware of this. One day, Bina enters the house to see Arun cooking in the kitchen and says:] (Epistemic Bias: $\neg p$)
[Context 2: Bina has no knowledge of Arun's interest/ disinterest in cooking as an activity. One day, Bina enters the house to see Arun cooking in the kitchen and says:] (Unbiased speaker)
[Context 3: Arun has always been good at cooking as an activity and his friend Bina is aware of this. One day, Bina enters the house to see Arun cooking in the kitchen and says:] (Epistemic Bias: p)

tu rasoi kər-e ʧʰe ***em***?
2.SG cooking do-IPFV be.PRS em

You are cooking. Are you?
Context 1, Context 2: [✓], Context 3: [✗]

Based on the data, a generalization that can be drawn is that the use of *em* requires contexts involving either the speaker's negative epistemic bias or when an unbiased speaker has direct evidence for the pejacent. *em* can never be used in contexts where the speaker is positively biased towards the prejacent.

2.2.4 Other Discourse Effects

Besides confirmation-seeking, *em* can also be used to express sarcasm, rhetoricity[7] or a confrontational stance. In these uses, *em's* function is not to confirm whether the prejacent proposition *p* is true, but rather to verify addressee's commitment to the prejacent proposition.

(14) [Context: A teacher is correcting answer-sheets. They come across an answer-sheet and suspect that the student has plagiarized from ChatGPT and confronts the student:]

a ɖʒəwabo te potte ləkʰ-ja ʧʰe *em*?
DEM.PROX answer.3.PL 2.SG.ERG yourself write-PFV.3.PL be.PRS em

You have written these answers yourself. Have you? [Rhetorical, Sarcastic]

Here *em* is not used to confirm whether 'it is the case that the addressee has written the quiz responses themselves' but to signal that the speaker is biased towards $\neg p$ wants the addressee to retract their commitment to *p*. *em*, when used in these contexts, functions to challenge the addressee's prior commitments.

(15) [Context: A lot of things have gone wrong with Arun on the day of the utterance and Bina is aware of it. They both witness the thunderstorms and heavy rains looking out from the window. Arun, getting frustrated, says:]

aɖʒe vərsad bi pəɖ-se *em*?
today rain also fall-3.FUT em

It is going to rain today too. Is it? (on top of everything) [Rhetorical, Frustration]

Here, *em* is used to convey the speaker's preference for $\neg p$. The addressee's commitment to *p* or $\neg p$ is of no relevance in this context.

3 Proposal

Based on the various discourse effects of the utterance-final *em* across the clause types of declaratives and interrogatives as observed in the previous sections, I now try to unpack how these effects arise through the interaction of some conventionally encoded meaning of *em*, the conventional meaning associated with the clause types, and the varying contextual conditions. I propose a preliminary contribution of *em* by situating it within commitment-based discourse models (Farkas & Bruce, 2010; Roberts & Rudin, 2024) which track commitments, open issues, and projected common grounds.

[7] I follow Farkas (2025) analysis of rhetorical questions.

3.1 Contextual conditions for (in)felicity

Before we look at the conventionally encoded meaning of *em*, it is important to separate the contextual conditions under which *em* can felicitously occur. As noted, *em* can occur in the presence of negative speaker bias as in (14) and negative speaker preference as in (15). *em* is also felicitous in contexts where the speaker is unbiased (9). Hence, it must be the case that speaker bias and speaker preference are contextual discourse parameters and not conventionally encoded by *em* itself.

3.2 Interaction of *em* with meanings of the clause types

The idea I want to explore is that *em* uniformly points to the prejacent (like a demonstrative) and thereby foregrounds the proposition expressed by the prejacent. It is this foregrounding function that interacts differently with the conventional contribution of the clause-types because these clause types have distinct default commitments and discourse functions thus giving rise to the different discourse effects.

3.3 Notions adopted

To model the assumptions of how the *em* questions update the discourse, I adopt the Table model of discourse (Farkas & Bruce (2010)) within which a context is defined in terms of four components described below:

Components of the Table model Farkas & Bruce (2010)

(16) a. DISCOURSE COMMITMENTS: For all discourse participants x, there is a set DC_x of propositions that x has publicly committed to.
b. TABLE: T is a push-down stack of Issues (sets of propositions)
c. COMMON GROUND: cg is the intersection of the discourse commitments of each participant.
d. PROJECTED SETS: ps set of possible common grounds.

In Farkas & Bruce (2010)'s system, context updates for various conversational moves (assertions or questions) are defined in terms of how they update the interlocutors discourse commitments, how they update the Table. For example, if the speaker asserts a proposition p, then p is added to the discourse commitments of the speaker, to the top of the Table and to the projected set (ps) . If the addressee accepts the assertion, then that next move in the discourse removes p from the Table and adds it to the common ground (cg). In contrast, a polar question would add $\{p, \neg p\}$ on the Table and would create projected sets containing p as well as ones containing $\neg p$.

Maxim for commitment and projection: Roberts & Rudin (2024)

Rudin (2018, 2022) notes that most work on pragmatic reasoning focuses on propositional content, ie., drawing inferences about what propositions the speaker intends to convey

based on their choice of the sentence and how it relates to what alternative choices they could have made. In contrast, he shifts the focus of pragmatic reasoning to discourse moves instead, where inferences are derived based on the speaker's choice either to make a commitment or not and from how they shape the direction of the conversation. He provides formalizations of these underlying pragmatics of the Table model where he introduces various maxims for commitment and projection. The maxim relevant for my analysis is the maxim of viability where projecting common grounds is subject to this maxim.

- Viability: Violated by any move that adds a set including p to T, where $\cap DC_{\mathrm{X}} \cap p = \phi$ for some interlocutor X, or $DOX_{\mathrm{X}} \cap p = \phi$ for some interlocutor X.

As described in Rudin (2022), "This maxim states that a cooperative agent shouldn't project a possible future for the conversation that contradicts any interlocutor's commitments (or private beliefs), as that possible future state of the common ground would either be unreachable, as common ground is blocked by the incompatible commitment, or uncooperative, as it would not accurately reflect the beliefs of the interlocutors."

Hence, a cooperative agent should not place a proposition on the Table if it is incompatible with any interlocutor's commitment or private beliefs. This maxim is important for my analysis of the *em* questions and to describe the way in which they update the discourse context in order to reason as to why it is viable for the speaker to put p on the table despite their bias for $\neg p$. The maxim is useful for understanding the pragmatic reasoning of the discourse move made by an *em* question.

Evidenced possibilities: Farkas & Roelofsen (2017)

Farkas & Roelofsen (2017) note that tag-interrogatives are distinct from polar interrogatives as they propose that the former signal that the speaker has access to some *evidence* for the highlighted alternative and are infelicitous in contexts in which such access to evidence for the highlighted alternative cannot be assumed. They thus propose that the representation of a discourse context should include, not only the list of possibilities that x (interlocutor) has publicly committed to, *commitments(x)* but also, a list of possibilities for which x has signaled to have some evidence. This list is denoted as *evidence(x)* and the possibilities that it contains is referred to as *evidenced possibilities*.

The notion of evidenced possibilities is useful for my analysis of *em* questions, as they too are felicitous only in contexts where the speaker has access to direct evidence for the prejacent proposition (highlighted alternative) provided by the prior discourse context.

3.4 Declaratives

This section will discuss the previous analysis of a default declarative update and proposes an idea of how the *em* declarative updates the discourse.

3.4.1 Previous Theories

Farkas & Bruce (2010)

A declarative sentence adds the asserted proposition to the speaker *Sp's* discourse commitments DC_{Sp}, places it on the Table T, and adds it to the projected set *ps* (Farkas & Bruce (2010)). Given below are the basic conventional discourse effects of a declarative sentence and how the context update could look like.

(17) Default Declarative update:

DC_{Sp}	Table	DC_{Ad}

ps: $\{s_1\}$

$\longrightarrow$

DC_{Sp}	Table	DC_{Ad}
p	$\{p\}$	

ps: $\{s_1 \cup \{p\}\}$

Having looked at the default declarative update, the next section describes the *em* declarative update and how it differs from the default declarative update.

3.4.2 Analysis of *em* declaratives

When the default assertion is accompanied by *em*, it still adds p to the speaker's discourse commitments $DC_S p$. The contribution of *em* is to make salient (through pointing) a more specific answer than what is strictly required and guides the listener to the intended meaning. An *em* declarative has the discourse effect of specification/clarification where a related QUD is triggered and answered with a detailed/ more specific answer and this discourse move then in effect answers the primary QUD indirectly. An *em* declarative flouts the maxims (Grice (1975)) of (1) quantity, where the speaker intentionally provides more information than minimally necessary and (2) relevance, where the speaker chooses to answer a related QUD and indirectly answers the primary QUD.

An idea of the context updates of default assertions followed by *em* with a falling intonation is that it is a complex discourse move where there are two updates that are happening as part of the strategy, first that answers the primary QUD and second that triggers a related QUD and answers it, which is an over-answer to the primary QUD to avoid misunderstanding the first answer, hence the specification. Modeling the context update for this complex discourse move is beyond the scope of this paper and is left for future work.

3.5 Interrogatives

This section will discuss the previous analysis of a default polar interrogative update as well as the non-canonical same polarity tag interrogative updates. I propose an analysis of how an *em* interrogative updates the discourse.

3.5.1 Previous Theories

Farkas & Bruce (2010)

They describe the default polar interrogative places the question's denotation $\{p, \neg p\}$ on the Table T, and projects both p and $\neg p$ into the Projected set *ps* for the addressee to

resolve. The speaker will have a dependent commitment based on the addressee's commitment on the Issue I (Gunlogson (2008)) once resolved. Given below is how a context update of a default polar interrogative would look like.

(18) Default Polar Interrogative update:

DC_{Sp}	Table	DC_{Ad}

ps: $\{s_1\}$

$\longrightarrow$

DC_{Sp}	Table	DC_{Ad}
	$\{p, \neg p\}$	

ps: $\{s_1 \cup \{p\}, s_1 \cup \{\neg p\}\}$

Given below are some of the previous which works have analyzed the English 'same-polarity tags' by building on the (Farkas & Bruce (2010)) Table model.

Malamud & Stephenson (2015)

They propose that by asserting p with an SP-tag (same-polarity tag), there is no change made to the speaker's current DC_{Sp} or projected commitments $DC_{\mathsf{Sp}*}$, or the current or projected CGs (common grounds). But rather, the SP-tag utterance adds p to the Table and to the addressee's projected commitments DC_{Ad} which indicates that the speaker is making a guess about the addressee's beliefs. Upon the addressee's acceptance of the move, p gets added to the addressee's commitment set. This is shown as:

(19) SP tag interrogative update:

DC_{Sp}	DC^*_{Sp}	Table	DC_{Ad}	DC^*_{Ad}

$\longrightarrow$

DC_{Sp}	DC^*_{Sp}	Table	DC_{Ad}	DC^*_{Ad}
		$\{p\}$		p

As an SP-tag projects a commitment of the addressee, rather than the speaker, it is expected that SP-tags are felicitous only when the addressee's judgment is at issue, but is infelicitous when the speaker is conveying their own stance and/or seeking agreement.

Roberts & Rudin (2024)

They present a differing view. By using the positive polarity tag questions, the speaker commits to p (via the assertion) but then questions p (via the tag), setting up a pragmatic tension. This triggers an antipresupposition, forcing an interpretation where the speaker's commitment is dependent, that is, based on external evidence (see Gunlogson (2008) for dependent commitment). These type of questions are used to express surprise or skepticism towards the evidence for the proposition in the anchor. This non-canonical context update they present is shown as:

(20) SP tag interrogative update:

DC_{Sp}	Table	DC_{Ad}

ps: $\{s_1\}$

$\longrightarrow$

DC_{Sp}	Table	DC_{Ad}
p_d	$\{p, \neg p\}$	

ps: $\{s_1 \cup \{p\}, s_1 \cup \{\neg p\}\}$

The next sub-section discusses how the *em* questions differ from that of the English SP-tags and describes an analysis of the *em* interrogatives.

3.5.2 Analysis of *em* questions

The Gujarati *em* questions, even though they seem similar to the English same polarity tag questions, behave differently. *em* questions used to seek confirmation, convey speaker's negative epistemic bias, preference, skepticism, rhetoricity, confrontational stance and surprise. An analysis that captures all of these discourse functions and effects of *em* questions is required.

Felicity conditions:

Condition Type	**Condition**
Evidence for the prejacent proposition	Must be provided by the prior discourse and must be direct
Speaker epistemic bias	Negative (expects ¬p)
	Neutral / unbiased

Table 1: Felicity conditions for *em* questions

I propose that for Gujarati *em* questions: the speaker does not commit to *p* and neither is it added to addressee's projected commitment set. The speaker is either seeking confirmation or is simply signaling a pragmatic conflict between their epistemic bias/preference and the evidence that is provided by the prior context. By using an *em* question, the speaker is signaling that they are evaluating whether the evidence, presented by the context, for the prejacent proposition is sufficient / adequate in order for them to add *p* to their discourse commitments. This intuition holds uniformly across contexts involving an unbiased speaker, as well as those in which the speaker is biased towards ¬ *p*, wherein they would either be open to potentially revising their beliefs or they prefer for the addressee to modify their commitment for *p*.

Context update of *em* questions:

To model the context update associated with *em* questions, I borrow the notion of *evidence* from (Farkas & Roelofsen (2017)) in addition to the key components of the *table model* (Farkas & Bruce (2010)) and the *viability maxim* (Roberts & Rudin (2024)). Building on these, I introduce the notion of ***adequacy value*** that the speaker employs to update the discourse context. Given the above mentioned intuitions, it will be assumed that *evidence(x)* is not just a list of possibilities but a list of pairs $< p, a >$, where *p* is a possibility and *a* is the adequacy value: if the speaker considers the evidence as adequate for them to add the prejacent proposition to their discourse commitment, the adequacy value is **'+'** and if the speaker considers the evidence as inadequate for them to add the prejacent proposition to their discourse commitment, the adequacy value is **'-'**.

(21) *em* question update:

Evidence: {<p, + / ->}		
DC_{Sp}	Table	DC_{Ad}
	{p, ¬ p}	

ps: $\{s_1 \cup \{p\}, s_1 \cup \{\neg p\}\}$

Discourse effect of *em* questions:

Basic effect:

- The denotation of the question $\{p, \neg p\}$ is added to the *table* and the *projected set*.

Special effect:

- $<p, a>$ is added to *evidence(x)*

Cases where speaker is biased towards ¬p and receives direct evidence for *p* from the context, by using an *em* question strategy, the speaker signals that they consider the evidence to be inadequate for them to revise their belief and puts both *p* and ¬p on the table and in the projected set. The speaker must have a reason, which is the epistemic bias / preference for ¬p, to present ¬*p* as viable (see the Viability maxim by Roberts & Rudin (2024)) despite the evidence for *p*. Cases where the speaker is unbiased, the function of the *em* question is to confirm whether the prejacent proposition is true, the speaker here considers the evidence as being adequate for them to undertake a dependent commitment.

Aside:

A potential alternative analysis of utterance-final *em* could be that it results from ellipsis of a speech report. In Gujarati, both *evu* and *em* can function as complementizers that introduce speech reports (as illustrated in 22a). Given this, one might hypothesize that utterance-final *em* simply reflects a truncated speech report construction, where the matrix verb (such as "say") has been elided, leaving behind the complementizer. However, the empirical facts argue against this ellipsis hypothesis. Specifically, only *em*, and not *evu*, is felicitous in utterance-final position (as shown in 22b). This asymmetry indicates that utterance-final *em* is not an elliptical remnant of reported speech. Rather, it has developed an independent discourse function and should be analyzed as a discourse particle that helps structure speaker commitments and discourse flow.

(22) [Context: Arun is asking Bina about how Riya is doing. Bina tells him that Riya is happy and he says:]

a. rija $k^huʃ$ $ʧ^he$ ***evu / em*** kəh-e $ʧ^he$
riya happy be.PRS evu/em say=IPFV.3.SG be.PRS.3.SG
He is saying that Riya is happy, Is he?

b. rija $k^huʃ$ $ʧ^he$ ***#evu / em***
riya happy be.PRS evu/em
Riya is happy, # like that / is she?

The next section provides a comprehensive summary of the main empirical and theoretical findings of this paper and outlines avenues for future research.

4 Summary and Future Directions

This paper has investigated the Gujarati discourse particle *em*, which derives its specialized utterance-final discourse use from a distal demonstrative in the language. It occurs in declaratives and interrogatives. It has a unified function of foregrounding the prejacent proposition acorss clause types. However, depending on the clause type to which it attaches and contextual conditions, *em* clauses have additional discourse functions.

In Declaratives, *em* can be used to pick out a particular entity from a set of contextually available alternatives, as a specification move. This strategy triggers and answers a related question to the primary QUD and in effect indirectly answers the primary QUD. It provides an answer to the related question, which is an over-answer (more specific) to the primary QUD in order to avoid any misunderstanding and guide the discourse towards the intended meaning.

In Interrogatives, *em* can be used for either neutral confirmation of the prejacent proposition or to convey that the speaker is biased against the prejacent proposition. The felicity of such *em* questions is conditioned by the prior discourse context, which must provide direct evidence for the prejacent proposition as well as the speaker's negative epistemic bias towards it or the speaker's dis-preference of the prejacent proposition. *em* questions convey a range of pragmatic flavors of rhetoricity, skepticism, surprise, confrontation depending on the discourse context and speaker's stance towards the prejacent proposition. They differ from the English SP-tags both in discourse functions and felicity conditions.

The preliminary theoretical proposal describes that the contribution of *em* would be modeled best within commitment-based discourse models (see Farkas & Bruce (2010) and Roberts & Rudin (2024)). For Gujarati *em* questions, it is proposed that the speaker does not undertake commitment to p and neither do they add it to the addressee's projected commitment set. They signal that the speaker is evaluating whether the evidence, presented by the discourse context, for the prejacent proposition is adequate for them to add p to their discourse commitments. This is modeled using the components of the *Table model* (Farkas & Bruce (2010)), the notion of *evidence* (Farkas & Roelofsen (2017)) and introducing the notion of adequacy value, which the speaker employs to update the discourse context, which captures the speaker's assessment of wether the available evidence is adequate to commit to the prejacent proposition. An *em* question updates the discourse context by placing the issue $\{p, \neg p\}$ on the table while simultaneously encoding the speaker's evaluation of the adequacy of the evidence for p.

The idea to be explored further in order to formalize the effect of the conventionally encoded meaning of *em* and the discourse update is that *em* declaratives make a non-canonical assertion update where it still adds the prejacent to the speaker's discourse commitments but highlights it among alternatives, leading to over-answering related QUDs. An *em* declarative is a complex discourse move comprising of two updates happening as part of the strategy. The first update is the one that answers the primary QUD and the second is that it triggers and answers a related question to the primary QUD, which is an over-answer to the primary QUD to avoid misunderstanding the first answer.

Along with this, various other avenues for future research emerge from the present study.

- Intonation and *em*: To investigate the prosodic properties of utterance-final *em*. In particular as to how varied intonation would distinguish the uses of *em*, in interrogatives, to seek confirmation, convey various distinct pragmatic flavors of sarcasm, confrontation, rhetoricity, surprise, skepticism via phonetic production and perception experiments.
- Role of Connectives: To explore how discourse connectives like *kem* (the wh-word 'what'), discourse markers like *eʈle* ('means') and *to* ('so') to license *em* when the prejacent is implied, probably helping our understanding of presupposition, anaphora, and QUD structure in Indo-Aryan languages.
- Hybrid Interrogatives: To examine *em's* behavior in hybrid interrogative constructions.

Acknowledgements

I thank Ashwini Deo for her constant guidance and insightful discussions. I am also grateful to David Beaver, Hans Kamp, Rajesh Bhatt, Deniz Rudin, members of the UT Austin SynSem seminar group and FASAL-15 participants for valuable comments and pointers. All errors are mine.

References

Beaver, David I & Brady Z Clark. 2008. *Sense and sensitivity: How focus determines meaning* Explorations in Semantics. West Sussex, United Kingdom: Wiley-Blackwell, a John Wiley & Sons, Ltd.

Bhatt, Rajesh & Veneeta Dayal. 2020. Polar question particles: Hindi-urdu kya. *Natural Language & Linguistic Theory* 38. 1115–1144.

Biezma, María, Miriam Butt, Farhat Jabeen & Benazir Mumtaz. 2025. Urdu/hindi polar kya as an expression of uncertainty. *Natural Language & Linguistic Theory* 43. 1417–1468.

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2019. *Inquisitive semantics* Oxford Surveys in Semantics and Pragmatics. Oxford, United Kingdom: Oxford University Press.

Cuenca, Maria-Josep. 2003. Two ways to reformulate: a contrastive analysis of reformulation markers. *Journal of pragmatics* 35(7). 1069–1093.

Cuenca, Maria Josep & Carme Bach. 2007. Contrasting the form and use of reformulation markers. *Discourse studies* 9(2). 149–175.

Deo, Ashwini. 2023. Hybrid interrogatives in Hindi-Urdu. Talk presented at FASAL 2023.

Deo, Ashwini. 2025. Take on this commitment: the particle bara in (Marathi) (Indo-Aryan). In *Proceedings of sinn und bedeutung*, vol. 29, 386–403.

Farkas, Donka. 2020. Canonical and non-canonical questions Available online: https://semanticsarchive.net/Archive/WU2ZjIwM/questions.pdf.

Farkas, Donka. 2025. Rhetorical questions revisited. In *Annual meeting of the chicago linguistic society*, vol. 60 1, 139–173.

Farkas, Donka & Kim B Bruce. 2010. On reacting to assertions and polar questions. *Journal of semantics* 27(1). 81–118.

Farkas, Donka & Floris Roelofsen. 2017. Division of labor in the interpretation of declaratives and interrogatives. *Journal of semantics* 34(2). 237–289.

Goodhue, Daniel. 2022. Isn't there more than one way to bias a polar question? *Natural Language Semantics* 30(4). 379–413.

Grice, Herbert P. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Speech acts*, vol. 3 Syntax and Semantics, 41–58. Leiden, The Netherlands: Brill.

Gunlogson, Christine. 2008. A question of commitment. *Belgian Journal of Linguistics* 22(1). 101–136.

Malamud, Sophia A & Tamina Stephenson. 2015. Three ways to avoid commitments: Declarative force modifiers in the conversational scoreboard. *Journal of semantics* 32(2). 275–311.

Roberts, Craige. 2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and pragmatics* 5. 1–69.

Roberts, Tom & Deniz Rudin. 2024. Polarity reversal questions and the semantics of prosodic incorporation. In *Proceedings of sinn und bedeutung*, vol. 28, 737–755.

Rudin, Deniz. 2018. *Rising above commitment*: University of California, Santa Cruz dissertation.

Rudin, Deniz. 2022. Intonational commitments. *Journal of Semantics* 39(2). 339–383.

Scheffler, Tatjana & Sophia A Malamud. 2023. "won't you?" reverse-polarity question tags in american english as a window into the semantics-pragmatics interface. *Linguistics and Philosophy* 46(6). 1285–1327.

Sudo, Yasutada. 2013. Biased polar questions in english and japanese. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, vol. 28, 275–295. Leiden, The Netherlands: Brill.

Left-adjoined correlatives as specificational pseudoclefts in Bangla

SRABASTI DEY, *Autonomous University of Barcelona*

ABSTRACT

This paper looks at specificational pseudoclefts as discussed in Akmajian (1970) and Higgins (1973) and compares them with correlative structures in Bangla, following Srivastav's (1991) analysis for Hindi. It is shown that right-adjoined relative structures in Bangla cannot constitute specificational pseudoclefts while left-adjoined correlatives can. This is because the particular semantics and syntax of correlatives attributed by Srivastav are compatible with the characteristics given for specificational pseudoclefts in the literature. This also addresses the debate over base-generation vs derivational analyses of specificational pseudoclefts. Lastly, it also explains the obligatory occurrence of the quirky copula holo in these constructions.

1 Introduction

Following Akmajian (1970), pseudocleft sentences can be characterised as copular constructions which have a semantic variable, often instantiated by a wh-clause in English, and a focus item that provides the 'value' for the variable, as in (1)[1]. Pseudoclefts in Bangla correspond to correlative structures where there is a relative clause with a relative marker and a main clause usually initiated by a demonstrative (2)[2].

(1) What Herman bought was that tarantula.

(2) ja Herman kinechhe ta holo oi TaranTula-Ta
REL Herman bought DEM COP that tarantula-CL
'What Herman bought was that tarantula'.

Though pseudoclefts in English are usually identified by the presence of a wh-relative clause, (Higgins 1973:64) notes that sentences with 'noun antecedents' like (3) should also be considered pseudoclefts. These sentences have a noun modified by a relative clause (with or without an optional relative marker like *that*) as their subjects and are thus internally headed as opposed to the wh-clause antecedents in (1) and (4). Such sentences can also be ambiguous with respect to specificational and predicational interpretations[3] like conventionally known pseudoclefts such as (4) is.

[1] I have used a Romanized style for the transcriptions, and the following characters in the examples can be interpreted as the corresponding sounds: ch - tʃ ; chh - tʃʰ ; T - ʈ ; sh - ʃ ; o - o ; O - ɔ.

[2] The form of the relative marker and demonstrative show concord with respect to honorificity and number specifications of the head noun so their forms might be different in other examples.

[3] In the predicational sense, we get a description for the discarded object; in the specificational sense, we learn which object was discarded. (see Akmajian 1970:175)

(3) The thing that John threw away was a valuable piece of equipment.

(4) What John threw away was a valuable piece of equipment.

The observation that sentences like (3) should be brought under the realm of pseudoclefts carries over to Bangla (5) in a very obvious way because in this language the internally-headed 'variable' constituents obligatorily require a relative marker in the antecedent clause.

(5) *(je) jiniS-Ta John hariyechhe she-Ta holo ek-Ta mulloban shOronjam
REL object-CL John lost DEM-CL COP one-CL valuable equipment
'The thing that John lost is a valuable equipment.'

In (6) we have the externally headed version of (5), corresponding closely to the English wh-pseudocleft.

(6) ja John hariyechhe ta holo ek-Ta mulloban shOronjam
REL John lost DEM COP one-CL valuable equipment
'What John lost is a valuable equipment.'

Thus, pseudoclefts of both types manifest as correlatives in Bangla. However, it must be noted here that the ambiguity between the specificational and predicational interpretations in sentences (3) and (4) is not seen in case of Bangla owing to the quirky copula *holo*. Only specificational correlatives can carry this copula, as indicated by their incompatibility with an additional adjective in (7).

(7) ja John hariyechhe ta holo ek-Ta mulloban shOronjam, (*bipOjjOnok-o)
REL John lost DEM COP one-CL valuable equipment dangerous-EMPH
Intended: What John lost is a valuable equipment, and important too.

(8) ja John hariyechhe ta ek-Ta mulloban shOronjam, bipOjOnnok-o
REL John lost DEM one-CL valuable equipment dangerous-EMPH
'What John lost is a valuable equipment, and important too.'

Correlatives can appear in three possible word orders in Bangla. I apply the nomenclature in Srivastav (1991) for Hindi correlatives to Bangla as they correspond to Bangla structures as well.

(9) je meye-Ta gaan shunchhe she lOmba
REL girl-CL music listening DEM tall
'The girl who is listening to music is tall.' [left-adjoined]

(10) she meye-Ta lOmba je gaan shunchhe
DEM girl-CL tall REL music listening
'The girl who is listening to music is tall.' [right-adjoined]

(11) she meye-Ta je gaan shunchhe lOmba
DEM girl-CL REL music listening tall
'The girl who is listening to music is tall.' [embedded]

Interestingly, only the left-adjoined correlative is grammatical with the specificational sense. The purported right-adjoined version (29) is ungrammatical, with or without the quirky copula.

(12) ja Herman kinechhe-e ta holo oi TaranTula-Ta
REL Herman bought-3 DEM COP that tarantula-CL
'What Herman bought was that tarantula.'

(13) *ta (holo) oi TaranTula-Ta ja Herman kinechhe-e
DEM COP that tarantula-CL REL John bought-3
'That tarantula was what Herman bought.'

In this work, I will address primarily try to answer two questions: (i) why are specificational pseudoclefts only expressed through left-adjoined correlatives and not the other types of correlatives in Bangla; (ii) where does the Bangla data figure in the literature on pseudoclefts, and predication in general. For the first question, we will analyse the syntactic and semantic attributes of the different correlative structures in Bangla, as given in Srivastav (1991) vis-a-vis specificational and predicational characteristics. Regarding the second point, we will look into the transformation vs base-generation approaches to pseudoclefts as exemplified in Akmajian (1970), Higgins (1973), among others, and put the data and the discussion in the perspective of Minimalist theories on predication (Adger & Ramchand 2003, Den Dikken 2006).

2 Background

As one might have noticed, I have referred to the copula in the Bangla examples above as a *quirky* copula. This is because it is not a regular copula by any means as we will see in the next sub-section. Besides specificational pseudoclefts, they have been found to be obligatory in root equative clauses (14a) and optional in root specificational (14b) and identificational (14c) constructions. They are incompatible with root adjectival predicational constructions (14d).[4]

(14) a. Peter Parker *(hol-o) Spiderman
Peter Parker COP-3 Spiderman
'Peter Parker is Spiderman.'

b. Gora-r lekhok (hol-o) Tagore
Gora-GEN writer COP-3 Tagore
'The writer of Gora is Tagore.'

c. o (hol-o) John
3 COP-3 John
'He is John.'

[4]See Dey (2023) for more details on the Bangla data

d. meye-Ta (*holo) klanto
girl-CL-3 COP tired
'The girl is tired.'

This copula has also been attested in the corresponding form in Odia and descriptions of this copula have been given in terms of 'focus' or 'emphasis' (Dasgupta 2006 and Thompson 2003). However, there has been no formal analysis of its structural position as far as I know.

2.1 The quirky coopula

This copula raises questions about its form and structural position for at least the following two reasons. First, the copula can only occur between DP1 and DP2 and cannot occur clause-finally. This is curious because Bangla has SOV order canonically, though it allows other orders pertaining to information structure constraints. Second, it takes the past form (*holo*) or the progressive form (*hochchhe*) of the verb 'be' *(hOwa*), though they are interpreted in the present tense in this position. If they appear in the default final position they would be interpreted as *becoming* (15b) or *became* (15c), as expected, though they sound odd owing to pragmatic reasons. The main point is that the last two examples definitely do not yield the identity relation depicted in (15a).

(15) a. Peter Parker ho-l-o / ho-chchh-e Spiderman
Peter Parker be-PERF-3 / be-PROG-3 Spiderman
'Peter Parker is Spiderman.'

b. Peter Parker Spiderman ho-chchh-e
Peter Parker Spiderman be-PROG-3
'Peter Parker is becoming Spiderman.'

c. Peter Parker Spiderman ho-l-o
Peter Parker Spiderman be-PERF-3
'Peter Parker became Spiderman.'

Since *holo* and *hochchhe* are interchangeable in (15a) as far as I can tell, I will only use *holo* in my work to refer to the quirky copula for the sake of simplicity.

2.2 Base-generation and extraction approaches

Before getting into the analysis of the correlatives qua pseudoclefts in Bangla, I will briefly discuss the literature on English pseudoclefts. Broadly speaking, there have been two approaches to specificational pseudocleft constructions in the early days of generative syntax, as summarised in Den Dikken (2007): transformation and base-generation. I have mostly referred to the analyses in Higgins (1973), who argues for an exclusively base-generation approach, and Akmajian (1970) who claims that both extraction and base-generation are valid ways of deriving pseudoclefts.

In the base-generated approach, the relative clause (i.e. the 'variable') is deemed to emerge in the surface subject position itself.

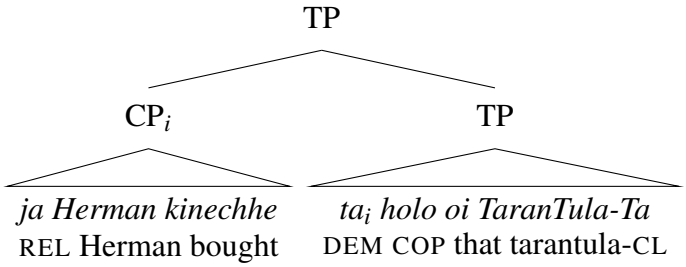
> ...pseudo-cleft sentences of all types are generated by the base rules in a form essentially identical to their surface structure form. In particular, the phrase which appears to the right of the copula is generated in that position by the base rules. (Higgins 1973:150)

Thus, there is no c-command relation between the pre- and post-copular constituents in this representation. He characterised the predicate constituent of the specificational pseudocleft to embody 'in some way the content or constitution' of the relative clause subject.

The analysis of Hindi left-adjoined correlatives in Srivastav (1991) can be looked at as an independent example for base-generated pseudoclefts qua correlatives: examples such as (2), repeated below, are shown to be base-generated in the CP-IP order (17).

(16) ja Herman kinechhe-e ta holo oi TaranTula-Ta
REL John bought-3 DEM COP that tarantula-CL
What Herman bought was that tarantula.

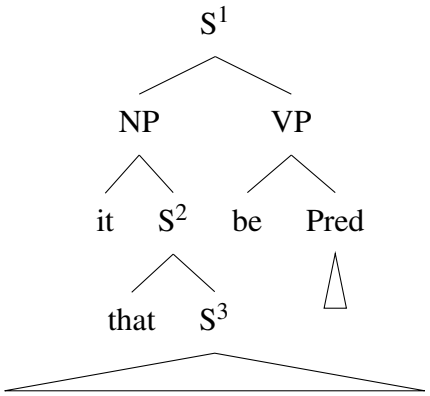
(17)



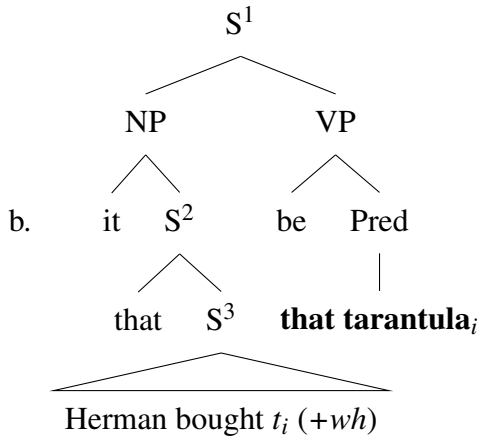
The left-adjoined correlative has different semantic and syntactic characteristics from the other two, which we will discuss in the following sections.

The transformational approach to pseudoclefts itself involved two possibilities: deletion and extraction, and we will only discuss the latter as it is more relevant here. The extraction approach was pursued in Akmajian (1970) among others. In this approach, the post-copular position is base-generated empty (18a); the 'value' constituent emerged within the embedded pre-copular constituent and ended up in the post-copular position in the surface structure following an extraction transformation (18b). The relative clause structure of the 'variable' is yielded from a separate *wh* transformation on the trace of the extraposed element.

(18) a.



b.

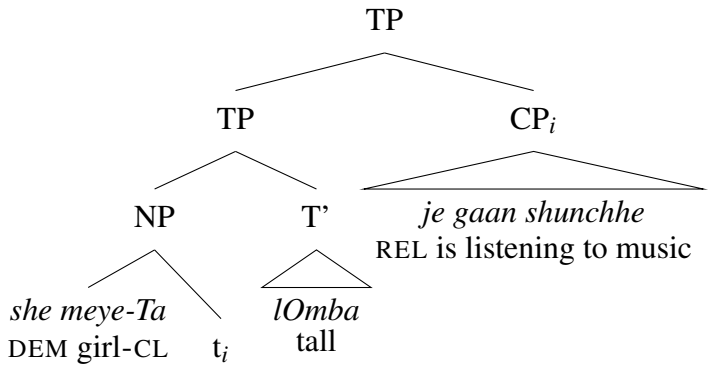


Though this would not be admissible within GB or Minimalist theories owing to violation of locality conditions (the extracted constituent would not be able to C-command over its trace), there have been other attempts to overcome this structural deficiency, as pointed out in (Den Dikken 2007: 386). What is pertinent to the current discussion is that as opposed to the base-generated structure, the value [focus] constituent is extracted to its surface position in the latter analysis. While Akmajian (1970) argued that both are admissible in the grammar and pointed out specific syntactic contexts where one would be favoured over another. Higgins (1973) maintained that only the base-generated approach is syntactically coherent.

Apropos the extraction analysis: incidentally, Srivastav's analysis for Hindi correlatives shows that the right-adjoined correlatives [a Bangla equivalent is provided in (10), repeated below] are actually derived following the right extraposition of the relative clause from a lower complement position (this underlying structure is the so-called embedded correlative). The details are applicable in a straightforward way to Bangla.

(19) she meye-Ta lOmba je gaan shunchhe
DEM girl-CL tall REL music listening
'The girl who is listening to music is tall.'

(20)



(21) she meye-Ta lOmba je gaan shunchhe
DEM girl-CL tall REL music listening
'The girl who is listening to music is tall.' [right-adjoined]

Thus, in a way, Bangla seems to attest both mechanisms posited to account for pseudoclefts – base-generation [as seen for left-adjoined correlatives in (9) and 17] and transformation [as seen for right-adjoined correlatives in (10 and 20)]. But as pointed out earlier, specificational pseudoclefts can only render themselves through the [base-generated] left-adjoined correlatives [(29), repeated below]:

(22) *ta holo oi TaranTula-Ta ja Herman kinechhe
DEM COP that tarantula-CL REL John bought
'That tarantula was what Herman bought.'

It is then clear that specificational pseudoclefts in these languages exclusively choose the base-generated structure though extraposition is also available in the languages. Thus, it becomes an interesting empirical point of departure to weigh in on this classical debate over the theory of pseudoclefts. However, current approaches towards predication present other considerations that need to be taken into account before we attempt an analysis of specificational pseudoclefts in Bangla and we will look into those in the next section.

3 Theory of Predication

Since the analysis of Italian copular constructions in Moro (1997), it has been a standard approach to collapse all copular constructions into either canonical order (23) or inverse order structures (24) – the predicates are indicated in bold.

(23) The picture of the wall is **the cause of the riot**.

(24) **The cause of the riot** is the picture of the wall.

Within this general view, specificationals (and equatives) are seen as inverse copular constructions where the surface subject is actually the underlying predicate which has undergone movement. For instance, Den Dikken 2006 shows how the precopular element in

equatives like (25) are actually reduced relative clauses with a covert relative marker a pro-predicate, which need to be obligatorily moved to the Spec of a functional head (such as T in English) that would license the pro-predicate.

(25) Cicero is Tully.

(26) [TP [Pred PRO-PREDICATEϕ [CP Op$_i$ [Cϕ [RP Cicero [RELATORϕ t$_i$]]]]$_j$ [T′ RELATOR$_k$=be [RP Tully [t$_k$ t$_j$]]]] [Den Dikken 2006:73]

The subject of the predication is an external argument mediated by a Relator, which can be instantiated by some functional category in Den Dikken's account. Thus, the subject and the predicate are in an asymmetrical relationship making a base-generated equative relation untenable, unless the copula is endowed with equative semantics in itself. Indeed, one of the difficulties for the view that considers predication as the primitive relation is accounting for the referential post-copular element in equatives. This challenge extends to specificational pseudoclefts as well, since the post-copular 'value' entity is usually referential. Den Dikken's analysis maintains an underlying predicational structure for equatives and specificational pseudoclefts without implying that the post-copular element is not referential.

On the other hand, there are accounts such as Heycock & Kroch (1999) which characterise specificational pseudoclefts as equatives. They do not assume the existence of an equative copula either but they achieve equative semantics through type-shifting operations. Thus, there is no requirement of an inversion analysis of equatives or specificational pseudoclefts for them.

In her 2012 paper, Heycock however maintains that inversion takes place in specificational constructions but raises doubts about the predicate status of the inverted constituents. She argues that the inverted elements behave somewhat like arguments and certain inversions can only take place when there is a type match as shown by the ungrammaticality of (28). She takes this requirement for type match an indication for an equation relation. Thus, she maintains the equation relation even in the inverted account of specificationals.

(27) Mary is the second kind of nurse

(28) *The second kind of nurse is Mary. (Heycock 2012:225)

I have followed the broad approach of considering predicationals as the underlying structure of all copular constructions, while other types are derived through inversion or some other operation as in Adger & Ramchand (2003), Den Dikken (2006), Heycock (2012) etc to pursue the inversion approach towards specificationals. Specifically, I will assume an underlying predicational structure for all copular construction and that inverted predicates move to the Spec of a functional category to be licensed, following Den Dikken (2006).

Under this view, there would naturally be a resistance in analysing specificational pseudoclefts as base-generated structures as Higgins and Akmajian suggested. However, we saw in the previous section that Bangla only uses the base-generated correlative clause for specificationals, seemingly exemplifying their position. In the next section, I will show that left-adjoined correlatives in Bangla can still be maintained to be base-generated without abandoning the predication-first stance.

4 Specificational pseudoclefts as left-adjoined correlatives

In the context of Bangla, following Srivastav, we have seen three possible surface orders for correlatives qua pseudoclefts. An empirical question was raised: why only left-adjoined correlatives are suitable for specificationals.

(29) *ta (holo) oi TaranTula-Ta ja Herman kinechhe
DEM COP that tarantula-CL REL John bought
'That tarantula was what Herman bought.'

Conversely, one can ask why are right-adjoined correlatives barred from expressing specificational pseudoclefts in Bangla. In order to address both these related questions, we will review Srivastav's syntactic and semantic analysis of the correlative structures. My attempt will be to show that left-adjoined correlatives independently embody semantic attributes that capture certain qualities of specificationals, which the right-adjoined correlatives lack.

The structural descriptions of the left-adjoined and right-adjoined structures yield different constraints in certain syntactic environments, supporting the analysis of different underlying structures. The different underlying syntactic structures also correspond to different semantic characteristics, which seem to be relevant for predication / specification. First, I will briefly demonstrate two empirical contexts Srivastav uses for Hindi and replicate them in Bangla to provide evidence for the different underlying syntactic structures in Bangla as well. Then I will show how the structure and the semantics of left-adjoined correlatives map to specificational pseudoclefts.

Srivastav points out that the left-adjoined correlative clauses can have an internally-headed relative clause (the head noun in bold) while the right-adjoined and embedded correlative clauses cannot. The following examples from Bangla corroborate with her observation for Hindi correlatives.

(30) je **meye-Ta** gaan shunchhe she meye-Ta lOmba
REL girl=CL music listening DEM girl-CL tall
'The girl who is listening to music is tall.' [left-adjoined]

(31) she meye-Ta lOmba je (***meye-Ta**) gaan shunchhe
DEM girl-CL tall REL girl-CL music listening
Intended: The girl who is listening to music is tall. [right-adjoined]

(32) she lOmba je (***meye-Ta**) gaan shunchhe
DEM tall REL girl-CL music listening
Intended: The girl who is listening to music is tall. [embedded]

This is a fallout of the fact that relative clauses in the embedded and right-adjoined correlatives are posited to emerge as complements to the noun (20); thus they do not contain the noun themselves. On the other hand, the relative clause in left-adjoined correlatives (17) emerge as independent CPs, and contain a nominal position.

This difference has an effect on the predication vs specificational divide seen between the left-adjoined and right-adjoined correlatives. The right-adjoined clause clearly behaves like a relative clause in English, yielding a predicative clause, unlike the one in left-adjoined correlatives.

The syntactic difference between right-adjoined and left-adjoined correlative structures that she points out is that the latter (34) cannot drop the demonstrative in the main clause (cf. 33).

(33) meye-Ta lOmba je gaan shunchhe
girl-CL tall REL music listening
The girl who is listening to music is tall. [right-adjoined]

(34) *je meye-Ta lOmba meye-Ta gaan shunchhe
REL girl-CL tall girl-CL music listening
Intended: The girl who is listening to music is tall. [left-adjoined]

This difference speaks to the semantic difference between these two structures. She points out that the obligatory presence of a demonstrative in the main clause of a left-adjoined relative is a result of the the relative marker being a quantifier that has to quantify over a variable (the demonstrative in this case) in the main clause. The relative marker in the right-adjoined correlative does not have this requirement since the relative clause here modifies the noun as shown above, thus not giving rise to a quantificational relationship. This brings us to the semantic attributes of the relative clauses in the respective correlative structures given by Srivastav: "internally-headed relatives are generalised quantifiers, while ordinary relatives can be set-denoting terms".

The CPs in right-adjoined correlatives are sets that intersect with the set denoted by the noun phrase [remember they are extraposed from the nominal complement position]. For instance, a relative clause like *meye-Ta je gaan shunchhe* 'the girl REL music listening' would have the depiction as in Figure 1.

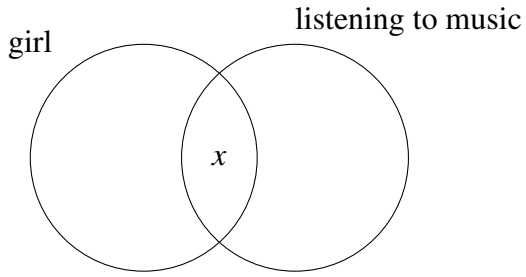


Figure 1: Set intersection of *girl* and *listening to music* in (33)

On the other hand, the relative clause in left-adjoined correlatives are not base-generated as a complement to a noun. Srivastav shows that the CP in these constructions is a set of

sets[5]. Thus, in a sentence like (30) the relative marker would yield a set of sets containing the intersection of the nominal set and the predicate *listening to music*. The entire clause is deemed true if the main clause clause is a set that is identical to any set within the set of sets in the CP (Figure 2).

$$\{x\} \in \{\{x\} \ldots\}$$

Figure 2: Relation between main clause and relative clause in left-adjoined correlatives

The fact that specificationals are compatible with correlatives that embody 'a set within a set of sets' is not puzzling. However, one might still ask why right-adjoined structures cannot express specification in Bangla. My conjecture for that is that the moved relative clause fails to take scope over the nominal, owing to independent scope restrictions in the language, thus failing to mediate a quantificational relation. The inability of the right-extraposed element to bind over lower variables can be shown independently as in (35). This in turn indicates that establishing a binding / quantificational relation between the value and variable constituents of a specificational pseudocleft is crucial in Bangla.

The following example show that rightward scrambled quantifier expressions fail[6] to take scope over lower constituents (as shown for Hindi in Bhatt & Dayal 2019):

(35) taar$_i$ bhai merechhe [prottek-Ta lok-ke]$_i$
3-GEN brother beat.PST each-CL man-ACC
Intended: His$_i$ brother beat [each man]$_i$.

The inference that Bangla has a strong quantification requirement for specificationals is compatible with informal descriptions of the (matrix or pseudocleft) specificational relation given in the literature. In the beginning of the article we already came across Akmajian's characterisation where he calls the relative clause a 'variable' and the post-copular entity the 'value'. Higgins describes specificationals as list denotation and points out that actual sentences with lists like (37) are 'very similar in meaning' to (36):

(36) What I bought was a punnet of strawberries and a pint of clotted cream.

(37) I bought the following things: a punnet of strawberries and a pint of clotted cream.
[Higgins 1979: 155]

Though I do not have a formal way of connecting the 'set within a set of sets' depiction of left-adjoined correlatives to the above observations on specification pseudoclefts yet, I would argue that they correspond closely. However, indirect evidence that specificationals are sensitive to scope effects comes from negation scope interaction in specificational and

[5]She assumes a type-shifting function to that takes an entity and abstracts over its properties, like in the case of generalised quantifiers.

[6]Leftward scrambled quantifier expressions are able to take scope

predicational sentences, as pointed out in Higgins (1973) and also Declerck (1988). For instance, only the value or the focus constituent can be negated in specificational sentences, resulting in a contrastive negation interpretation and never a neutral negation interpretation (cf. the predicational example in (40):

(38) ?The best candidate is not John.

(39) The best candidate is not John, but Bill.

(40) The best candidate is not tall.

In other words, specificational sentences cannot take wide scope, as Declerck points out, though predicationals can have both narrow scope and wide scope negation.

The other interesting data point in (41) is from Den Dikken (2007) who quotes these examples from Declerck (1988) during a discussion on exhaustivity implicature of specificationals. What we see here is that if exhaustivity is an entailment in a relative clause and there is only a unique value that can satisfy it then the post-copular constituent in a specificational blocks *only*, as seen in (41b).

(41) a. The one who murdered Smith is my neighbour. [ambiguous]

b. The one who murdered Smith is only my neighbour. [only predicational]

This could be taken as an indication there the post-copular constituent here is inherently sensitive to the quantificational requirement imposed by the relative clause, and thus blocks the extraneous *only*.

There are works (Heycock 2012 and Selvanathan 2016) that admit inversion in specificational contexts, while positing a relation of equation between the two terms. However, if one wants to maintain the stance that the semantics of equation is difficult to admit in the grammar (see Adger & Ramchand 2003, Roy & Shlonsky 2019 for discussion) then the Bangla data can be used to suggest that quantificational scope can be explored as the relevant factor for specificationals.

However, though we find base-generated specificational pseudoclefts in Bangla I will not suggest that specification can be ascribed a primitive function like that of predication. As I will show in the next section, it still requires predicate inversion inside the main clause to yield a specificational structure.

5 Inversion in specificational correlatives

There is some evidence from Bangla data and precedence in the literature to suggest that predicate inversion takes place within the main clause in the specificational correlatives. Essentially, the demonstrative is base-generated in the predicate position in specificational pseudoclefts. On the other hand, the demonstrative emerges in the surface subject position in predicational correlatives. We have precedence of this kind of analysis in specificational and predicational English *it*-clefts in Den Dikken (2013). While the *it* in the specificational clause (42) cannot exert CONTROL (since it emerges in the predicate position and is not referential), the *it* in a predicational clause (43) can:

(42) it$_i$ was Ryan who murdered Brian, besides PRO$_i$ being a bad guy.

(43) it$_i$ was an interesting meeting I went to last night, despite PRO$_i$ being poorly organised.

The specificational and predicational pseudoclefts in Bangla also exhibit similar differences with respect to the referentiality of the demonstrative in the main clause. A specificational clause like (44) cannot tolerate a referential distal deictic demonstrative like *o-Ta* in the main clause while a predicational one (45) can[7]:

(44) *Bheem je kaaj-Ta niyechhe o-Ta holo shikhOkOta
Bheem REL work-CL taken 3PRON-CL COP teaching
'The job Bheem that has taken up is teaching.'

(45) ?Bheem je kaaj-Ta niyechhe o-Ta koThin
Bheem REL work-CL taken 3PRON-CL difficult
'The job that Bheem has taken up is difficult.'

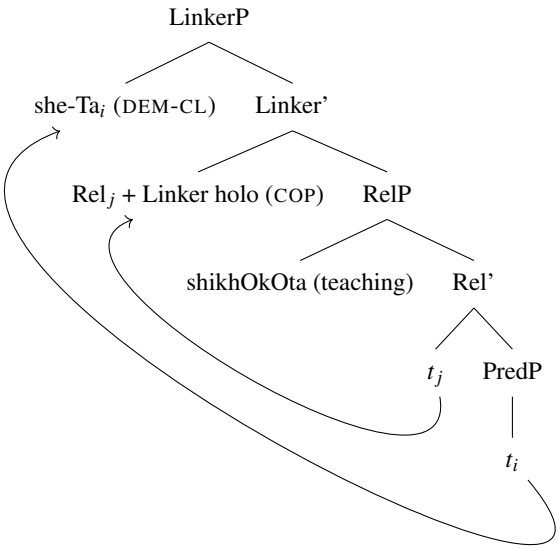
The crucial evidence for predicate inversion in Bangla specificational pseudoclefts would be the requirement of the semantically vacuous copula that occurs clause-medially here:

(46) Bheem je kaaj-Ta niyechhe she-Ta *(holo) shikhOkOta
Bheem REL work-CL taken DEM-CL COP teaching
The job that Bheem has taken up is teaching.

I have tried to show in Dey (2023) how the obligatory quirky copula in Bangla matrix equatives is a syntactic reflex of the Linker projection to enable inversion of the underlying predicate since Bangla lacks overt tensed copulas in predicational copular constructions (in a language like English, T serves as the Linker). Since the quirky copula is obligatory in specificational pseudoclefts as well, I will infer that predicate inversion applies in this context too. The underlying predicate position of the demonstrative in specificationals thus explains the ungrammaticality of (44). In (47) we see inversion of the demonstrative in the main clause of a specificational sentence like (46).

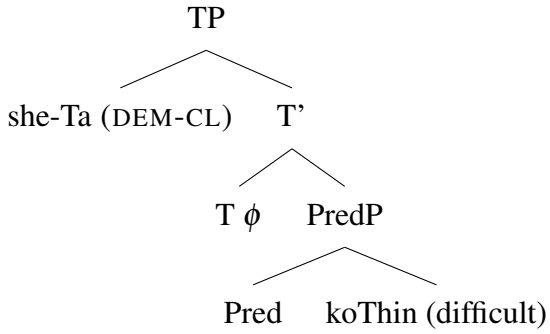
[7]The predicational example (45) is slightly odd: as Bagchi (1994) notes, 'in a Bangla correlative structure an anaphoric correlative pronoun is preferred over a diectic one'. But when it comes to specificationals as in (44) the deictic demonstrative is completely out.

(47)



The demonstrative in a predicational correlative like (45) emerges in the surface subject position (48), thus (somewhat) allowing referential demonstratives and not triggering the presence of the quirky copula *holo*[8].

(48)



There is possibly another way to interpret this data – scrambling. It has been argued in Heycock (2012) that certain traits of predicate inversion with respect to information structure can be explained in terms of scrambling in a language like German. Data from Hindi is particularly instructive to explore this idea further. This is because Hindi, unlike Bangla, has an overt tensed copula in predicationals but it still requires the copula to obligatorily occur clause-medially in specificationals (50).

(49) Bheem-ne jo kaam lia hai vo hai apne bareme kitab likhna
Bheem-ERG REL work taken be.PRES.3 DEM COP.3 self about book to.write
The job Bheem has accepted is to write a book about himself.

(50) *Bheem-ne jo kaam lia hai vo apne bareme kitab likhna hai
Bheem-ERG REL work taken be.PRES.3 DEM self about book to.write COP.3

[8]I am not committed to the maximal projection of the finite clause in predicational correlatives being a TP

The job Bheem has accepted is to write a book about himself.

The predicational structure bears the copula clause-finally as expected:

(51) Bheem-ne jo kaam liya hai vo uske liye peeDa ka karan hai
Bheem-ERG REL work taken be.PRES.3 DEM him for pain GEN source COP.3
The job Bheem has accepted is a source of pain for him.

Acknowledgements

I would like to express my gratitude to Xavier Villalba, Marcel den Dikken, and the audience at FASAL-15 for their comments and suggestions at various stages of this work. I thank Anjasi NN and Vipul Vivek for the Hindi data judgements. Finally, I acknowledge and appreciate the financial support provided by the project PID2023-150347NB-I00 and grant 2022 FI_B 00477 for this work.

References

Adger, David & Gillian Ramchand. 2003. Predication and equation. *Linguistic Inquiry* 34(3). 325–359.

Akmajian, Adrian. 1970. *Aspects of the grammar of focus in english*: Massachusetts Institute of Technology dissertation.

Bagchi, Tista. 1994. *Theoretical perspectives on word order in south asian languages, csli lecture notes* chap. Bangla correlative pronouns, relative clause order and D-Linking. CSLI, Stanford, California.

Bhatt, Rajesh & Veneeta Dayal. 2019. Rightward scrambling as rightward remnant movement. *Linguistic Inquiry* 50(1). 197–218.

Chomsky, Noam. 1972. *Studies on semantics in generative grammar* chap. Deep Structure, Surface Structure, and Semantic Interpretation. De Gruyter Mouton.

Dasgupta, Probal. 2006. The positive polarity copula in Bangla: an item of unexplored origin. *Indian Linguistics* 67(1-4). 61–64.

Declerck, Renaat. 1988. *Studies on copular sentences, clefts and pseudo-clefts*. Berlin, New York: De Gruyter Mouton.

Den Dikken, Marcel. 2006. *Relators and linkers: The syntax of predication, predicate inversion, and copulas*. The MIT Press.

Den Dikken, Marcel. 2007. *The blackwell companion to syntax, volume i* chap. Specificational Copular Sentences and Pseudoclefts. Blackwell Publishing Ltd.

Den Dikken, Marcel. 2013. Predication and specification in the syntax of cleft sentences. In *Cleft structures*, vol. 8 3, 285–301. John Benjamins Publishing Company.

Dey, Srabasti. 2023. Aspect-marked copulas in bangla. In *Proceedings of 31st console*, .

Heycock, Caroline. 2012. Specification, equation, and agreement in copular sentences. *The Canadian Journal of Linguistics / La revue canadienne de linguistique* 57(2). 209–240.

Heycock, Caroline & Anthony Kroch. 1999. Pseudocleft Connectedness : Implications for the LF Interface Level. *Linguistic Inquiry* 30(3). 365–397.

Higgins, Francis Roger. 1973. *The pseudo-cleft construction in english*: Massachusetts Institute of Technology dissertation.

Mikkelsen, Line. 2005. *Copular clauses : specification, predication and equation*. John Benjamins Publishing Company.

Moro, Andrea. 1997. *The raising of predicates*. Cambridge University Press.

Roy, Isabelle & Ur Shlonsky. 2019. *Copulas across languages* chap. Aspects of the syntax of 'ce' in French copular sentences Aspects of the syntax of ce in French copular sentences. Oxford University Press.

Selvanathan, Nagarajan. 2016. *Inversion in copular clauses and its consequences*: Rutgers, The State University of New Jersey dissertation.

Srivastav, Veneeta. 1991. The syntax and semantics of correlatives. *Natural Language and Linguistic Theory* .

Thompson, Hanne-Ruth. 2003. *Toward a definitve grammar of bengali: A practical study and critique of research on selected grammatical structures*: School of Oriental and African Studies dissertation.

Tamil =$\bar{e}$ as a focus-sensitive scalar operator

GOPALAKRISHNAN RAMAMURTHY, *Independent*

ABSTRACT

The Tamil enclitic =$\bar{e}$, described as an "emphatic" focus particle in literature, gives rise to a wide range of interpretations depending on its syntactic position and the discourse context. When =$\bar{e}$ is cliticised sentence-medially, it can license scalar additive, exclusive, intensifying and precisifying inferences. When it attaches sentence-finally, =$\bar{e}$ serves to signal that the prejacent proposition is already in the common ground. This paper offers an analysis of =$\bar{e}$. I argue that the sentence-medial and sentence-final uses of =$\bar{e}$ correspond to two distinct lexical items, =$\bar{e}_1$ and =$\bar{e}_2$ respectively, motivated by the fact that only =$\bar{e}_1$ is obligatorily sensitive to focus. In the rest of this paper, I examine =$\bar{e}_1$. I propose that =$\bar{e}_1$ associates its prejacent with the highest-ranked proposition among a contextually specified set of alternatives. The various discourse effects associated with =$\bar{e}_1$ are shown to arise due to its compatibility with multiple types of alternative propositions and strength rankings, and other contextual factors.

1 Introduction

Tamil has an extremely multi-functional discourse enclitic =$\bar{e}$. When =$\bar{e}$ cliticises sentence-medially within a finite clause, it can give rise to a range of interpretations depending on the context, including scalar additive (1a), scalar exclusive (1b), exhaustive (1c), intensifying (1d) and precisifying (1e) inferences.[1,2] When =$\bar{e}$ attaches at the end of a finite clause, it serves discourse functions which include signalling the speaker's belief that the prejacent proposition is already in the common ground of discourse (1f).[3]

(1) a. CONTEXT: Ram won a bronze medal in a Decathlon, and:
ʃyām veḷḷi meḍal=ē jeyi-cc-ān
Shyam silver medal=$\bar{e}$ win-PST-3.SG.M
'Shyam won even a *silver medal*.' [⇝ *This is very unexpected/noteworthy.*]

b. CONTEXT: Ram won a gold medal in a Decathlon, and:
ʃyām veḷḷi meḍal=ē jeyi-cc-ān
Shyam silver medal=$\bar{e}$ win-PST-3.SG.M
'Shyam only won a *silver medal*.' [⇝ *Nothing better than a silver medal.*]

[1]Tamil is transcribed as follows: *t* /t̪/, *d* /d̪/, *c* /tʃ/, *j* /dʒ/, *y* /j/, *ŭ* [i̯]. Macrons mark long vowels: *ā* /aː/, etc.

[2]Glosses follow the Leipzig rules, with the following additional abbreviations: ADD = additive particle, ASP = aspect, HON = honorific, HUM = human, NHUM = non-human

[3]Unless noted otherwise, all the data is based on the author's intuitions, confirmed by three other speakers.

c. CONTEXT: Ram had wanted to go to several cities in Europe, but in the end:
rām barlin-kk=ē pō-n-ān
Ram Berlin-DAT=*ē* go-PST-3.SG.M
'Shyam went *only* to Berlin.' [⇝ *Ram went nowhere else.*]

d. rām osaram-ā irŭ-pp-ān, ʃyām osaram-ā.v=ē irŭ-pp-ān
Ram height-ADV be-FUT-3SG.M, Shyam height-ADV=*ē* be-FUT-3.SG.M
'Ram is tall, (and/but) Shyam is *really* tall.' [⇝ *Shyam is taller than Ram.*]

e. ellār=um=ē va-nd-ānga
everyone-ADD=*ē* come-PST-3.PL.HUM
'*Absolutely* everyone came.' [⇝ *Every single person came.*]

f. CONTEXT: A had told B earlier that A's son lives in Delhi. A few weeks later, B asks A, "Where does your son live now?" A replies:
avan ɖelli-lɛ irŭ-kk-ān=ē
he Delhi-LOC be-PRS-3.SG.M=*ē*
'He's in Delhi, as you know.' [⇝ *A believes that B already knows this.*]

This paper offers an analysis of =$\bar{e}$. In §2, I argue that =$\bar{e}$ should be treated as two distinct lexical items, =$\bar{e}_1$ and =$\bar{e}_2$, motivated by their complementary syntactic distributions — =$\bar{e}_1$ must occur sentence-medially and =$\bar{e}_2$, sentence-finally — and the fact that only =$\bar{e}_1$ is obligatorily sensitive to focus. This departs from prior authors (Arokianathan 1981; Chevillard 1997; Murugaiyan 2009), who do not appreciate this distinction. In the rest of this paper, I examine the focus-sensitive enclitic =$\bar{e}_1$, leaving =$\bar{e}_2$ for future work.

§3 describes the range of interpretations that =$\bar{e}_1$ gives rise to in different contexts. §4 presents the analysis. I propose that =$\bar{e}_1$ is uniformly a scalar operator: =$\bar{e}_1$ introduces the presupposition that its prejacent corresponds to the highest-ranked true alternative proposition in the Current Question (Beaver & Clark 2008). Its various discourse effects arise due to its compatibility with multiple types of alternative propositions and strength rankings, as well as other contextual factors. §5 concludes.

It bears noting that =$\bar{e}_1$, described in the literature as an "emphatic" enclitic (e.g., Asher 1989: 94), resembles several other "emphatic" enclitics in other South Asian languages, such as its Telugu cognate =*ē* or Hindi =*hī*. Of particular relevance is Deo's (2023) study of the Marathi variant =*ts*, which develops a unified account of =*ts*'s similarly varied discourse uses. However, while Deo's analysis greatly informs the present study, it cannot be adopted fully for Tamil =$\bar{e}_1$, as =$\bar{e}_1$ is licensed only in a strict subset of the contexts that license Marathi =*ts*, so that applying Deo's account to =$\bar{e}_1$ would overgenerate.[4] This paper restricts its focus to the Tamil data, with the aim of studying crosslinguistic variation in future research.

[4]In particular, as Deo (2023) demonstrates, Marathi =*ts* can signal that its prejacent is the most salient focus alternative among the interlocutors; in fact, she identifies this as =*ts*'s core discourse function. Notably, Tamil =$\bar{e}_1$ does not share =*ts*'s use of coordinating the mutual salience of focus alternatives.

2 =ē corresponds to two separate lexical items

I assume Hamblin semantics for questions, and a standard Roothian alternative semantics for focus (Hamblin 1973; Rooth 1992). A question denotes a set of propositions that potentially answer it, and focus introduces alternatives in an alternative-semantic dimension. An expression α has an ordinary semantic value $[\![\alpha]\!]^o$ and a focus semantic value $[\![\alpha]\!]^f$. If α is unfocused, $[\![\alpha]\!]^f = \{[\![\alpha]\!]^o\}$. If α is focused, $[\![\alpha_F]\!]^f$ is the set of all objects of the same type as $[\![\alpha]\!]^o$. Focus semantic values of complex expressions are computed by applying functional application point-wise to each element in the FSVs of their constituents.

I adopt a Question Under Discussion framework, in which discourse structure is modelled along hierarchically stacked Questions Under Discussion (QUD) (Roberts 2012, a.o.). At any point in a conversation, there may be a number of open QUDs; following Beaver & Clark (2008), I refer to the most salient QUD, at the top of the QUD stack, as the Current Question (CQ). I also adopt Beaver & Clark's (2008) proposal that obligatorily focus-sensitive expressions conventionally associate with the CQ. Congruence between focus and the CQ is regulated by the Focus Principle: "Some part of a declarative utterance should evoke a set of alternatives containing all the Rooth–Hamblin alternatives of the CQ" (Beaver & Clark 2008: 37).

With this background, I now examine =ē's interactions with focus. When =ē cliticises sentence-medially to a constituent contained within a finite clause, =ē's host must contain the focused element.[5] Given the question (2a), (2b) is a felicitous answer, as =ē is cliticised to *ʃyāmɛ* 'Shyam.ACC', and this placement of focus makes (2b) congruent with (2a). In contrast, (2c), in which =ē is cliticised to *rām* 'Ram', is an infelicitous answer, as focus on *rām* renders (2c) incongruent with (2a).

(2) a. QUESTION: Whom did Ram see?
b. rām ʃyām-ɛy=ē pātt-ān
Ram [Shyam-ACC]$_F$=ē see.PST-3.SG.M
'Ram only saw *Shyam*.' [⇝ *He saw nobody else.*]
c. #rām=ē ʃyām-ɛ pātt-ān
#[Ram]$_F$=ē Shyam-ACC see.PST-3.SG.M
#'Only *Ram* saw Shyam.' [⇝ *Nobody else saw Shyam.*]

=ē is banned from cliticising inside NPs (i.e., to participial relative clauses and adjectival modifiers). If nominal modifiers are focused, =ē must cliticise to the nominal head, and focus on the modifier is expressed by prosodic stress (glossed in bold). (3a) and (3b) thus have different interpretations solely due to different placements of focus, which strongly suggests that =ē is obligatorily sensitive to focus.

(3) a. QUESTION: Which pen did you buy?

[5]The determinant of finiteness in Tamil, and more generally in Dravidian languages, is a matter of debate (McFadden & Sundaresan 2014; and citations therein). I remain agnostic on this issue.

nān **karŭppŭ** pēnā.v-ɛy=ē vāng-in-ēn
I [black]$_F$ pen-ACC=*ē* buy-PST-1.SG
'I only bought the *black* pen.' [⇝ *I bought no other sort of pen.*]

b. QUESTION: What did you buy?
nān karŭppŭ pēnā.v-ɛ.y=ē vāng-in-ēn
I [black pen-ACC]$_F$=*ē* buy-PST-1.SG
'I only bought the *black pen*.' [⇝ *I bought nothing else.*]

However, =*ē* does not exhibit obligatory sensitivity to focus when it occurs sentence-finally, i.e., when it is cliticised at the end of a finite clause. In (4a-b), =*ē* is felicitous regardless of the position of focus. Moreover, this sentence-final =*ē* contributes meaning that is orthogonal to focus: it signals the speaker's belief that the prejacent is in the common ground.

(4) a. QUESTION: Where did Ram go?
rām ɖelli-kki pō-n-ān=ē
Ram [Delhi-DAT]$_F$ go-PST-3.SG.M=*ē*
'Ram went to *Delhi*, as you know.' [⇝ *Spkr believes that Addr knows this.*]

b. QUESTION: Who went to Delhi?
ɖelli-kki rām pō-n-ān=ē
Delhi-DAT [Ram]$_F$ go-PST-3.SG.M=*ē*
'*Ram* went to Delhi, as you know.' [⇝ *Spkr believes that Addr knows this.*]

To highlight the contrast between sentence-medial =*ē* and sentence-final =*ē*, consider the context in (5) along with a prejacent sentence "*Siva hit*$_F$ *Shyam*", with narrow focus on the verb. As (5a) shows, for the obligatorily focus-sensitive =*ē* to associate with the verb focus, it must attach to a non-finite form of the content verb "hit" (*aɖikka* 'hit.INF'), with a dummy verb *cey* 'do' bearing the finiteness inflections (tense marking and subject agreement). (5a) thus expresses that the prejacent is the most unexpected proposition among alternatives of the form "*Siva* VERB *Shyam*" (5a-i). Crucially, when =*ē* cliticises sentence-finally to the finite verb *aɖiccān* 'he hit' (5b), this interpretation is unavailable (5b-i), meaning that focus-associating =*ē* must surface inside a finite clause. Instead, just as in (4), the sentence-final =*ē* in (5b) gives rise to an inference that is orthogonal to the focus, namely that Sita already knows the prejacent (5b-ii), which makes (5b) infelicitous in this context. Note that this "*as you know*" reading is unavailable with sentence-medial =*ē*, shown in (5a-ii).

(5) CONTEXT: Sita hears that Shyam and Siva had a fight, and asks Ram about it. Ram says: "At first Shyam and Siva were just yelling at each other, then...

a. ʃivā ʃyām-ɛ aɖi-kka.v=ē cenj-ān
Siva Shyam-ACC [hit-INF]$_F$=*ē* do.PST-3.SG.M

i. 'Siva even *hit* Shyam.' [⇝ *This was a surprising escalation of the fight.*]

ii. UNAVAILABLE: 'Siva *hit* Shyam, as you know.'

b. #ʃivā ʃyām-ɛ aɖi-cc-ān=ē
#Siva Shyam-ACC [hit-PST-3.SG.M]$_F$=$\bar{e}$
i. UNAVAILABLE: 'Siva even *hit* Shyam.'
ii. #'Siva *hit* Shyam, as you know.' [⇝ *Ram believes that Sita knows this.*]

I therefore conclude that =$\bar{e}$ represents two distinct lexical items, =$\bar{e}_1$ and =$\bar{e}_2$, differentiated by their complementary syntactic positions and differing sensitivity to focus. =$\bar{e}_1$ necessarily occurs in sentence-medial positions and is obligatorily focus-sensitive, whereas =$\bar{e}_2$ cliticises sentence-finally and does not exhibit obligatory focus-sensitivity. That is, only =$\bar{e}_1$ conventionally associates with the Current Question (Beaver & Clark 2008). In the rest of this paper, I restrict my attention to =$\bar{e}_1$. Hence, I will refer to =$\bar{e}_1$ simply as '=$\bar{e}$' when there is no ambiguity. For discussion on =$\bar{e}_2$, see Murugaiyan (2009), Chevillard (1997) and Arokianathan (1981), although none of them distinguish between =$\bar{e}_1$ and =$\bar{e}_2$.[6]

3 Empirical landscape

This section describes the range of interpretations associated with =$\bar{e}_1$, and the contexts in which they arise. This discussion draws heavily from Deo's (2020, 2023) insightful accounts of Marathi =*ts*. Note that since the rest of this paper (§3-4) deals exclusively with =$\bar{e}_1$, I refer to it simply as '=$\bar{e}$' throughout.

3.1 Scalar inferences

Firstly, =$\bar{e}$ licenses inferences similar to those associated with scalar additives like English *even* when the focus alternatives are ranked by a salient contextually determined scale. Consider a context which provides a pragmatic scale. (6) gives rise to three inferences: (i) a mirative evaluative inference that the prejacent proposition is considerably unexpected or noteworthy; (ii) an additive inference that individuals lower on the hierarchy also praised Ram; and (iii) an upper-bound inference that no-one higher on the hierarchy did so. The additive inference is cancellable: (6) can easily be followed by "*But nobody else praised me*". The upper-bound inference cannot be cancelled: (6) is judged infelicitous in contexts where the Chairman also praised Ram.

(6) CONTEXT: Ram's workplace has a Chairman, President, Vice-President and Manager, in decreasing levels of hierarchy. Ram was complimented at work. He says:
enn-ɛ prɛsiɖɛnʈ=ē pārāʈʈ-in-ārŭ
I-ACC President=$\bar{e}$ praise-PST-3.SG.HON
'Even the *President* praised me.'

However, whether =$\bar{e}$ licenses an *even*-like evaluative inference depends crucially on the position of the prejacent on the contextual scale relative to salient expectations or stan-

[6] =$\bar{e}_2$ resembles the German discourse particle *ja* in several respects (cf. Schneider 2022).

dards.[7] Consider (7), where Siva and Shyam have participated in separate events in a decathlon. In each of (7a-c), Sita's utterance about Siva's award establishes an expectation regarding the award that Shyam won. In (7a), where the prejacent ranks higher than the contextual standard (winning bronze), =*ē* gives rise to a mirative "stronger-than-expectations" inference, akin to English *even*. In (7b), where the prejacent matches the standard (winning silver), =*ē* emphasises that Shyam's performance *exactly* met the benchmark set by Siva, no more and no less. In (7c), where the prejacent ranks below the contextual standard (winning gold), =*ē* contributes a scalar exclusive inference. Note, however, that (7c) does not license a mirative "weaker-than-expectations" evaluative inference, meaning that =*ē* does not give rise to expectation-lowering effects, unlike exclusives like English *only*.[8]

(7) a. CONTEXT: Sita: "Siva won a bronze medal. What about Shyam?" Ram says:
ʃyām veḷḷi meḍal=ē jeyi-cc-ān
Shyam silver medal=*ē* win-PST-3.SG.M
'Shyam won even a *silver medal*.' [⇝ *This is very unexpected/noteworthy.*]

b. CONTEXT: Sita: "Siva won a silver medal. What about Shyam?" Ram replies:
ʃyām=um veḷḷi meḍal=ē jeyi-cc-ān
Shyam=ADD silver medal=*ē* win-PST-3.SG.M
'Shyam too won *exactly a silver medal*.' [⇝ *No better or worse than a silver.*]

c. CONTEXT: Sita: "Siva won a gold medal. What about Shyam?" Ram replies:
ʃyām veḷḷi meḍal=ē jeyi-cc-ān
Shyam silver medal=*ē* win-PST-3.SG.M
'Shyam only won a *silver medal*.' [⇝ *Nothing better than a silver.*]

The common thread in (6) and (7) is that when the alternatives are ranked by a salient contextually determined scale, =*ē* indicates that no alternative stronger than the prejacent is true. Additionally, when the prejacent exceeds contextual standards, =*ē* gives rise to a mirative "greater-than-expectation" inference, as in (6) and (7a), and further licenses additive inferences when the context supports them, as is the case in (6), but not in (7a).[9]

(8) CONTEXT: Ram was not expected to run more than 5 kms in a marathon, but:
rām pattŭ kilomīṯṯar=ē ōḍ-i-ṯṯ-ān
Ram ten kilometre=*ē* run-ASP-COMPL.PST-3.SG.M
'Ram even ran *ten kilometres*.' [⇝ *This is very unexpected/noteworthy.*]

Note that =*ē* exhibits the same behaviour in contexts involving numerical scales. (8) parallels (7a), and the same sentence would have interpretations analogous to (7b) and (7c) in the appropriate context. I do not illustrate this here due to space constraints.

[7] This discussion is inspired by ideas in Deo's (2020) earlier analysis of Marathi =*ts*. Though this is superseded by Deo (2023), her earlier approach applies well to Tamil =*ē*.

[8] Note that =*ē* also differs from Marathi =*ts* in this respect, as Deo (2023) describes =*ts* as licensing both "greater-than-expectations" and "lesser-than-expectations" mirative inferences.

[9] In this respect, =*ē* contrasts with the enclitic =*kūḍa*, which conventionally expresses both scalar and additive meanings and corresponds more directly to English *even* (Lehmann 1993: 159).

3.2 Exhaustive inferences

When the context does not provide a salient ranking over the alternatives, =*ē* gives rise to exhaustive inferences. That is, =*ē* expresses that all alternatives other than the prejacent are false, as in (9a) and (9b), the latter adapted from Asher (1989: 95). These exhaustive inferences are not cancellable. For instance, continuing (9a) with "*And I also went to Paris*" is infelicitous. Note also that just like in (7c) above, exhaustive uses of =*ē* such as in (9a-b) also do not give rise to mirative expectation-lowering effects, unlike English *only* or *just*.

(9) a. CONTEXT: Ram had gone on a vacation to Europe. After he returns, Sita asks, "Where did you go during your trip?" Ram replies:
nān barlin-kk=ē pō-n-ēn
I Berlin-DAT=*ē* go-PST-1.SG
'I only went to *Berlin*.' [⇝ *I went nowhere else.*]

b. vivasāya vēlɛ.y=ē nān en vāɻnāɭŭ pūrā.v=um
agriculture.OBL work=*ē* I I.OBL lifetime entirety=ADD
cenjŭ-giʈʈ=irŭ-nd-ēn. vēra enda toɻil=um
do.ASP-PROG=be.AUX-PST-1.SG other which occupation=ADD
teri.y-ādŭ
know-FUT.NEG.3.NHUM
'I was doing just *agriculture* my entire life. I don't know any other occupation.'

In §3.1, I presented the generalisation that when the alternatives are ranked by a salient contextual scale, =*ē* indicates that the prejacent is the strongest true alternative. Following Coppock & Beaver (2014), I assume that when no such scale is salient in the context, the alternatives are ranked by entailment, such that they give rise to a boolean lattice. If we make the additional assumption that =*ē* is compatible with multiple types of strength rankings over the alternatives, the earlier generalisation can then account for the exhaustive inferences in (9) as well.[10] I will make this precise in §4.1 and §4.2.

3.3 Intensifying inferences

When =*ē* cliticises to gradable expressions[11], it licenses intensifier-like interpretations, depending on the scale associated with the expression. In (10), which involves the upper-bounded scales OPEN and CLOSED, the argument entity is evaluated as having the property to the maximal degree, translatable with English *fully* or *completely*.

(10) a. kadavŭ tara-nd=ē irŭ-kkŭ
door open-CVB=*ē* be-PRS.3.NHUM
'The door is *fully* open.' [⇝ *It is wide open, not closed even an inch.*]

[10] In this, =*ē* contrasts with the enclitic =*maʈʈum*, which is an exclusive operator compatible only with alternatives ranked by entailment scales (Lehmann 1993: 159), akin to English *exclusively*.

[11] Tamil does not have adjectives as a lexical category (see Venkatesan 2025; and citations therein). I hence use the neutral term "gradable expression", abstracting over the morphosyntactic differences among gradable expressions in Tamil.

b. kadavŭ mūɖ-i.y=ē irŭ-kkŭ
door close-CVB=*ē* be-PRS.3.NHUM
'The door is *fully* closed.' [⇝ *It is not open even an inch.*]

In the case of an upper-open scale, like TALL, =*ē* indicates that it holds true of the argument entity relative to the highest relevant contextual standard. This can have a comparative-like effect when there is a limited set of salient entities in the discourse, as in (11a). Note that this is distinct from a comparative construction, as both Ram and Shyam must be evaluated as tall in (11b), while a comparative only requires that one be taller than the other. In the absence of a salient comparandum, =*ē* conveys the evaluation that the gradable expression holds true no matter the contextual standard. In (11b), the speaker asserts that everyone who judges his roommate's cupboard would find it dirty, no matter the context.[12]

(11) a. rām osaram-ā irŭ-pp-ān, ʃyām osaram-ā.v=ē irŭ-pp-ān
Ram height-ADV be-FUT-3.SG.M Shyam height-ADV=*ē* be-FUT-3.SG.M
'Ram is tall, (and/but) Shyam is really tall.' [⇝ *Shyam is taller than Ram.*]

b. CONTEXT: A university student says of his flatmate's cupboard:
avan alamāri aɻŭkk-ā.v=ē irŭ-kkŭ
he.OBL cupboard dirt-ADV=*ē* be-PRS.3.NHUM
'His cupboard is *completely* dirty.' [⇝ *It is dirty no matter the standard of dirtiness.*]

Additionally, when =*ē* cliticises to intensifying modifiers like *romba* 'very' (12) or *naraiya* 'a lot', =*ē* amplifies their standard-raising effect, similar to English *really very*.

(12) paɖam romba.v=ē aɻag-ā irŭ-kkŭ
picture very=*ē* beauty-ADV be-PRS.3.NHUM
'The picture is really very beautiful.' [⇝ *Even more beautiful than expected*]

=*ē*, as a focus-sensitive particle that can yield both scalar and intensifying interpretations, has parallels in Hebrew *bixlal* and Mandarin *gèng*. Greenberg (2020) and Chen & Greenberg (2024) argue that *bixlal* and *gèng* are scalar operators that can associate with focus on covert syntactic elements — specifically, a companion class argument *c* for *bixlal*, and a covert comparative COMP for *gèng*. Based on this and further cross-linguistic parallels such as Russian *voobščé*, they propose that the ability to associate with focus on covert elements is a parameter of variation for focus-sensitive operators. Orenstein & Greenberg (2021) extend this proposal to Hebrew *be-sax ha-kol*, analysing it as a scalar exclusive that associates with a covert degree modifier to license approximative inferences.

I adopt this proposal for =*ē*. Translating it into Beaver & Clark's (2008) framework, I propose that =*ē* is compatible with a Current Question consisting of alternative propositions obtained by applying varying values to covert contextual parameters in the prejacent

[12] =*ē* differs from Marathi =*ts* in its interactions with gradable expressions. Deo (2023) argues that =*ts* has no standard-raising effect when cliticised to adjectives and only yields a "no matter the context" reading.

sentence. When =*ē* cliticises to gradable expressions in their "positive form", the alternatives vary only in the contextual standard of evaluation supplied by the covert modifier POS (Kennedy 2007). Intensification obtains as =*ē* selects the strongest alternative, i.e., the proposition associated with the highest relevant standard. I will make this precise in §3.4.

3.4 Precisifying inferences

When =*ē* cliticises to expressions that allow for imprecision in their interpretation, it has similar functions as slack regulators such as English *exactly* and *right* (Lasersohn 1999). Precisifying interpretations frequently arise with universal expressions (13a), in which case =*ē* cliticises outside the obligatory additive enclitic =*um*. They are also common with numerals and deictics, such as (13b) and (13c), which correspond to English "*that exact/very*" and "*exactly three*", respectively. Precisifying interpretations also arise with spatial and temporal expressions, particularly in contexts where there is epistemic uncertainty among interlocutors, as shown in (13d).[13]

(13) a. ellār=um=ē va-nd-ānga
everyone=ADD=*ē* come-PST-3.PL.HUM
'*Absolutely* everybody came.' [⇝ *Every single person in the group came.*]

b. nān mūn=ē postagam paɖi-cc-ēn
I three=*ē* book read-PST-1.SG
'I read *exactly* three books.' [⇝ *No more or less than three books.*]

c. CONTEXT: Speaker is at a mobile phone shop and points at a phone on display:
en-akkŭ ad=ē fōn vēŋum
I-DAT that=*ē* phone want
'I want that *exact* phone.' [⇝ *The same device; not another one of that model.*]

d. CONTEXT: Sita: "I think the meeting starts at around 8:30." Ram replies:
illɛ, mīʈiŋ eʈʈ-arɛ-kk=ē ārami-kkum
NEG meeting eight-half-DAT=*ē* start-FUT.3.NHUM
'The meeting will start *exactly* at eight-thirty.' [⇝ *No earlier or later.*]

Drawing on Wiegand's (2018) account of precisifying readings of English *just*, I propose that when =*ē* cliticises to imprecise expressions, the Current Question contains alternatives that vary only in size of the pragmatic halo associated with the host of =*ē* (Lasersohn 1999). Precisification obtains as =*ē* selects the strongest alternative, i.e., the proposition associated with the smallest relevant pragmatic halo. I will elaborate on this in §4.4.

3.5 Interim summary

=*ē* licenses scalar additive, scalar exclusive, and exhaustive inferences, depending on the presence of a salient contextual ranking over the focus alternatives, and the relative strength

[13] =*ē* diverges from Marathi =*ts* in this regard, as Deo (2023) reports that =*ts* does not yield slack regulator-like inferences with spatio-temporal expressions like 'at eight-thirty'.

of contextual expectations. When $=\bar{e}$ cliticises to expressions with context-sensitive interpretations, including gradable and imprecise expressions, it has intensifying and slack regulator-like effects, respectively. In §4, I present the formal analysis of this data.

4 Analysis

Following Coppock & Beaver (2014), I assume that each discourse context c provides a common ground INFO_c, modelled as the set of worlds that verify the mutual public discourse commitments of all interlocutors (the Stalnakerian context set). Each c also provides a Current Question CQ_c and a strength ranking $\geq_c$. The CQ_c is formalised as a set of propositions (set of set of worlds), and $\geq_c$ is a partial order over the CQ_c.

The lexical entry in (14) is based on the MAX operator in Coppock & Beaver (2014). $=\bar{e}$ combines with a proposition p, which is the denotation of the prejacent sentence S at context c, and introduces the presupposition that p is stronger according to $\geq_c$ than all propositions in CQ_c that are true in world w. If this presupposition is satisfied, $=\bar{e}$ is truth-conditionally vacuous and returns p. Hence, the discourse function of $=\bar{e}$ is to indicate that p is the strongest true proposition in the CQ_c.

(14) $[\![=\bar{e}]\!]^c = \lambda w_s.\ \lambda p_{\langle s,t\rangle} : \forall q \in \text{CQ}_c\ [q(w) \rightarrow p \geq_c q].\ p$

The discourse effects associated with $=\bar{e}$ arise depending on the nature of the alternatives in the CQ_c, as well as the strength ranking (Orenstein & Greenberg 2021). When the CQ_c is ordered by a salient contextual scale, the presence of $=\bar{e}$ emphasises the maximal strength of p in the CQ_c, yielding different pragmatic effects depending on p's strength relative to contextual expectations (§4.1). In the absence of a contextually salient scale, the CQ_c is ordered by entailment, in which case $=\bar{e}$ gives rise to exhaustive inferences (§4.2).

Importantly, $=\bar{e}$ is able to operate over a CQ_c that contains alternative denotations of the prejacent sentence varying only in the values of covert contextual parameters, including standards of evaluation of gradable expressions, and the degree of restriction of pragmatic halos. In the former case, $=\bar{e}$ selects the alternative relative to the highest relevant standard, giving rise to intensifying interpretations (§4.3). In the latter case, it selects the denotation with the most restricted pragmatic halo, giving rise to precisifying interpretations (§4.4).

4.1 Accounting for scalar inferences

§3.1 showed that when the CQ_c is ordered by a salient contextual scale, $=\bar{e}$ licenses an upper-bound inference that no alternative stronger than the prejacent p is true. Additionally, if p ranks higher on the scale than salient contextual expectations, $=\bar{e}$ gives rise to a mirative "greater-than-expectation" inference akin to English *even*, and further licenses a cancellable inference that propositions weaker than p are true. (6), repeated as (15), illustrates all three effects.

(15) CONTEXT: Ram's workplace has a Chairman, President, Vice-President and Manager, in decreasing levels of hierarchy. Ram was complimented at work. He says:

enn-ɛ prɛsiɖɛnʈ=ē pārāʈʈ-in-ārŭ
I-ACC President=*ē* praise-PST-3.SG.HON
'Even the *President* praised me.'

Let *X* be shorthand for the proposition "*X praised me*". In (15), the CQ_c is the set in (16a), ranked by a scale of unexpectedness (16b). According to (14), (15) asserts *President* as true, and presupposes that it is stronger than all true alternatives in the CQ_c. I argue that these together yield the upper-bound inference due to Gricean Quality. That is, if a speaker asserts the truth of p in the actual world w_0 while presupposing that p is stronger than all alternatives true in w_0, it must be the case that the speaker is not in a position to assert the truth of any q stronger than p. Moreover, since the speaker has made explicit reference to the relative strength of the alternatives, they cannot be assumed to be ignorant of the truth of any stronger q in w_0. Hence, it must be that they believe that all alternatives stronger than p (which is *Chairman* in 15) are false, as otherwise the utterance would violate Quality.[14]

(16) a. $\{Chairman, President, Vice-President, Manager\}$
b. $Chairman \geq_c President \geq_c Vice-President \geq_c Manager$

The "greater-than-expectations" inference in (15) arises as a pragmatic implication due to the interaction of this upper-bound inference with salient contextual expectations. In the context of (15), it is understood that receiving praise from someone as senior as the President is relatively uncommon. In such a case, the speaker's use of *=ē* conveys that the strongest true alternative in fact exceeds what is normally expected. This has the effect of emphasising that the prejacent is stronger (i.e., more unexpected) than expectations, giving rise to the mirative evaluative inference in (15), as well as (7a).

Finally, drawing on ideas in Rullmann (1997), the additive inference in (15) arises as a conversational implicature. When a speaker asserts a proposition while pragmatically highlighting that it exceeds expectations, addressees reason that less unexpected alternatives are probably also true, since if p is true in w and p is more unexpected than q, q is typically also true in w. Importantly, this reasoning is based on what is typical and is defeasible, which accounts for the cancellability of the additive inference in (15) and the fact that it is not licensed in contexts incompatible with it (7a).

In contexts where the prejacent is equally strong as contextual expectations (7b; repeated as 17a), in addition to the upper bound inference (which follows from the same Gricean reasoning described above), the speaker's use of *=ē* highlights the fact that the strongest true alternative equals, and does not exceed, the contextual standard (here, winning a silver). This has the pragmatic effect of emphasising that the prejacent *exactly* equals the standard — Shyam's performance in (17a) exactly matches Siva's.

(17) a. CONTEXT: Sita: "Siva won a silver medal. What about Shyam?" Ram replies:
ʃyām=um veḷḷi meḍal=ē jeyi-cc-ān
Shyam=ADD silver medal=*ē* win-PST-3.SG.M
'Shyam too won *exactly a silver medal*.' [⇝ *No better or worse than a silver.*]

[14]This argument is inspired by Deo's 2023 account of the exclusive readings of Marathi *=ts*.

b. CONTEXT: Sita: "Siva won a gold medal. What about Shyam?" Ram replies:
ʃyām veḷḷi meḍal=ē jeyi-cc-ān
Shyam silver medal=*ē* win-PST-3.SG.M
'Shyam only won a *silver medal*.' [⇝ *Nothing better than a silver.*]

I now turn to cases where the prejacent is weaker than expectations. (7c), repeated as (17b), yields an upper-bound inference, which again follows from a combination of the assertion in (17b), the presupposition of =*ē*, and interlocutors reasoning by Gricean Quality. However, recall from §3.1 that (17b) does not give rise to mirative "lesser-than-expectations" inferences. This is unexpected — (17b) expresses that the strongest true alternative in the CQ_c is weaker than the contextual standard, which is predicted to give rise to an expectation-lowering effect in a way that is analogous to the mirative inferences in (15) and (17a). It is unclear how to explain the lack of this pragmatic effect in (17b).[15]

Finally, this account straightforwardly extends to cases in which the CQ_c is ordered by numerical scales, such as (8).

4.2 Accounting for exhaustive inferences

§3.2 showed that when the context does not provide a salient order over the CQ_c, =*ē* licenses exhaustive inferences. Following Coppock & Beaver (2014), I assume that in such cases, the CQ_c is ordered by entailment. The alternatives in the CQ_c hence give rise to a boolean lattice, in which each node corresponds to an alternative, and the nodes' arrangement corresponds to the entailment ranking. Consider (9a), repeated as (18).

(18) CONTEXT: Ram had gone on a vacation to Europe. After he returns, Sita asks, "Where did you go during your trip?" Ram replies:
nān barlin-kk=ē pō-n-ēn
I Berlin-DAT=*ē* go-PST-1.SG
'I only went to *Berlin*.' [⇝ *I went nowhere else.*]

Let us assume that Ram had planned to visit Berlin, Paris and Rome, and let *X* be shorthand for "*I went to X*". The CQ_c is hence the set in (19a). Assuming that the predicate "*go to*" is distributive, such that it holds of a group just in case it holds of each entity in the group, this set is ordered such as in (19b).

(19) a. {*Berlin*, *Paris*, *Rome*, (*Berlin* ∧ *Paris*), (*Berlin* ∧ *Rome*), (*Paris* ∧ *Rome*), (*Berlin* ∧ *Paris* ∧ *Rome*)}
b. (*Berlin* ∧ *Paris* ∧ *Rome*) $\geq_c$ (*Berlin* ∧ *Paris*) $\geq_c$ *Berlin*, etc.

According to (14), (18) asserts *Berlin* as true, and presupposes that *Berlin* is stronger than all true alternatives in the CQ_c. By the same reasoning as was described in §4.1, the

[15] Interestingly, Tamil has another focus-sensitive enclitic =*dān*, which licenses exclusive and "lesser-than-expectations" inferences, but not "greater-than expectations" inferences. Descriptively, =*ē* and =*dān* appear to be in complementary distribution in their interactions with contextual expectations.

combination of the assertion and presupposition in (18) leads interlocutors to jointly reason, by Gricean Quality, that all stronger alternatives (i.e., those that entail *Berlin*) are false:

(20) *Berlin* ∧ ¬(*Berlin* ∧ *Paris*) ∧ ¬(*Berlin* ∧ *Rome*) ∧ ¬(*Berlin* ∧ *Paris* ∧ Rome)

This entails ¬*Paris* ∧ ¬*Rome*. Hence, assuming that the speaker is cooperative, it must be that they believe *Berlin* to be the only true alternative, lest Quality be violated.

Note, however, that this account relies on the assumption that the predicate is distributive. Crucially, we find that exhaustive inferences associated with =*ē* are cancellable when the predicate is non-distributive. Consider (21), with the predicate "*drive the car*" being non-distributive in this context. In such cases, =*ē* conveys what Lehmann (1993: 157) terms "self-performance". Intuitively, (21) conveys that Ram drove the car personally, without assistance, and by his own volition, which corresponds to meanings expressed by exclusive adverbial reflexive intensifiers in English (König & Gast 2006). Notably, (21) does convey an exhaustive inference, but it is cancellable: Ram can continue his utterance by saying "*I drove partway, and Siva drove the rest of the way*".

(21) CONTEXT: Ram is talking about a road trip that he and his roommates went on over the weekend. Sita knows that Ram and his friends share a single car amongst each other. Sita asks, "Who drove the car?" Ram replies:
nān=ē vaɳḍi.y-ɛ ōṯṯ-in-ēn
I=*ē* car-ACC drive-PST-1.SG
'*I* drove the car *myself*.' [⇝ *I drove it by myself, without help, of my own volition.*]

I suggest that the cancellability of the exhaustive inference in (21) arises from the predicate "*drive the car*" being non-distributive. In the context of (21), "*X and Y drove the car*" does not entail "*X drove the car*" or "*Y drove the car*" individually. Hence, given an entailment ordering over the CQ_c, the two propositions are unordered relative to each other, so that applying the presupposition of =*ē* to one does not require the other to be false. A parallel to this is found in English *it*-clefts, which interact in an analogous way with non-distributive predicates (Velleman et al. 2012). However, the source of the additional "self-performance" inferences in (21) — that subject, Ram, performed the action personally and independently — is unclear. I leave this as an open question.

4.3 Accounting for intensifying inferences

§3.3 showed that when =*ē* cliticises to gradable expressions, it has intensifier-like effects, which depend on the nature of the scale that the expression is associated with.

Following Kennedy (2007), I assume a semantic ontology that includes a type *d* for degrees, and that gradable expressions denote measure functions from entities to degrees (type $\langle e, d \rangle$).[16] I also assume that the bare "positive forms" of gradable expressions contain a covert modifier POS (22), which supplies a contextually specified standard of evaluation

[16] As noted in §3.3, since Tamil lacks a distinct category of adjectives, this abstracts over the syntactic and potentially semantic differences in the ways that Tamil realises adjective-like expressions.

$s^c(g)$. $\text{POS}(x)(g)$ holds just in case the entity x has the property g to a degree that exceeds $s^c(g)$, thus introducing contextual variation in interpretation.

(22) a. $[\![\text{POS}]\!]^c = \lambda g_{\langle e,d\rangle}.\ \lambda x_e.\ g(x) > s^c(g)$
b. $[\![\textit{tall}]\!]^c = \lambda x_e.\ \text{TALL}_{\langle e,d\rangle}(x)$
c. $[\![\text{POS}(\text{TALL})(\text{RAM})]\!]^c = \text{TALL}(\text{RAM}) > s^c(\text{TALL})$

As discussed in §3.3, building on ideas in Greenberg (2020) and Orenstein & Greenberg (2021), I propose that $=\bar{e}$ can operate over a CQ_c that contains alternative propositions obtained by applying varying values to covert contextual parameters in the prejacent sentence *S*. In cases where the $=\bar{e}$ cliticises to a gradable expression in the positive form, the Current Question is along the lines of "To what degree does *X* have the property *Y*?". Hence, the CQ_c contains alternative denotations of *S* that differ only in the contextual standard of evaluation introduced by POS. As per (14), $=\bar{e}$ indicates that the interpretation of *S* corresponds to the strongest alternative in the CQ_c, i.e., the denotation in which the gradable expression is evaluated against the highest standard that is relevant in the context.

I first examine gradable expressions that map to upper-closed scales. In (10a), repeated as (23), the CQ_c contains alternatives in which 'open' is evaluated against different relevant standards of openness. Since OPEN is an upper-closed scale that has a salient maximal endpoint, $=\bar{e}$ selects the alternative evaluated relative to this maximal standard. (23) thus asserts that the door's openness exceeds the maximal degree, i.e., it is as widely open as it could be.

(23) kadavŭ tara-nd=ē irŭ-kkŭ
door open-CVB=$\bar{e}$ be-PRS.3.NHUM
'The door is *fully* open.' [⇝ *It is wide open, not closed even an inch.*]

When the associated scale is upper-open and lacks a maximal endpoint, such as TALL, $=\bar{e}$ indicates that the gradable expression is evaluated against the highest contextually relevant standard. In (11a), repeated as (24), two salient entities are under comparison. Assume that Ram's tallness is evaluated against a standard s_1. The CQ_c then consists of two alternatives, p_1 and p_2, where Shyam's tallness is evaluated against the standards s_1 and s_2, respectively, where $s_2 > s_1$, and therefore $p_2 \geq_c p_1$. $=\bar{e}$ selects the stronger alternative p_2. Hence, (24) asserts that Shyam's degree of tallness exceeds s_2, and Ram's tallness exceeds s_1. This gives rise to a scalar implicature that Ram's tallness does not exceed the higher standard s_2, implying that Ram is less tall than Shyam.[17]

(24) rām osaram-ā irŭ-pp-ān, ʃyām osaram-ā.v=ē irŭ-pp-ān
Ram height-ADV be-FUT-3.SG.M Shyam height-ADV=$\bar{e}$ be-FUT-3.SG.M
'Ram is tall, (and/but) Shyam is really tall.' [⇝ *Shyam is taller than Ram.*]

Compare (24) with (11b) in §3.3. In (11b), there is no salient comparandum for the argument entity. The CQ_c thus contains alternatives in which the expression 'dirty' is evaluated

[17] This argument is indebted to Greenberg's (2020) account of Hebrew *bixlal*.

against different standards of dirtiness of rooms. $=\bar{e}$ selects the alternative evaluated against the highest such standard. Hence, (11b) asserts that the flatmate's room is dirty beyond all possible standards — it would be judged dirty by everyone who evaluates it.

The effect of $=\bar{e}$ on intensifiers, as in (12), can be explained similarly. Following Kennedy & McNally (2005), I assume that a modifier like *romba* 'very' denotes a function that returns a standard of evaluation calculated based on entities to which the positive form of the gradable expression already holds. As $=\bar{e}$ has the pragmatic effect of raising the standard of evaluation for the positive form, it also has a knock-on effect of increasing the intensifier's standard-raising effect. I leave it to future work to develop this argument.

4.4 Accounting for precisifying inferences

§3.4 showed that $=\bar{e}$ has a precisifying effect when it cliticises to expressions that license imprecision, including universal expressions, numeral and deictic determiners, and spatial and temporal expressions.

Following Lasersohn (1999), I model imprecision using "pragmatic halos". In Lasersohn's framework, each expression α is associated with a pragmatic halo $H_c(\alpha)$, which is a set of objects of the same type as α's literal denotation $[\![\alpha]\!]$, that differ from $[\![\alpha]\!]$ only ways that are "pragmatically ignorable" at context c. The elements of α's pragmatic halo $H_c(\alpha)$ are ordered by their relative closeness to $[\![\alpha]\!]$, and the size of $H_c(\alpha)$ may vary across contexts. A halo $H_1(\alpha)$ is smaller than halo $H_2(\alpha)$ just in case the farthest element from $[\![\alpha]\!]$ in $H_1(\alpha)$ is closer to $[\![\alpha]\!]$ than the farthest element from $[\![\alpha]\!]$ in $H_2(\alpha)$, meaning that the smaller α's halo, the more precise the interpretation of α. Pragmatic halos of complex expressions are derived compositionally from the halos of their parts through pointwise functional application — sentences thus have halos containing propositions.

Slack regulators are operators that restrict or expand pragmatic halos, thereby regulating the degree of acceptable imprecision in interpretation. In this regard, Wiegand (2018) proposes the existence of covert slack regulators. Specifically, she argues that the precisifying readings of English *just* arise from *just* quantifying over covert slack operators. I have also proposed in this paper that $=\bar{e}$ is able to operate over alternative propositions that vary only in the values of contextual parameters.

Based on these ideas, I propose that when $=\bar{e}$ cliticises to an expression α that allows imprecision in its interpretation, the CQ_c is along the lines of "How many/Where/When exactly?". The CQ_c thus contains alternative denotations of the prejacent sentence S that vary only in the degree to which a covert slack regulator restricts the pragmatic halo of α. Since halos are compositional, assuming that the halos of other expressions in S are constant, the degree of restriction of α's halo exclusively determines the size of S's halo. The CQ_c is thus ordered by the degree of restriction of S's halo, such that $p_1 \geq_c p_2$ if p_1 is associated with a more restricted halo than p_2. As per (14), $=\bar{e}$ then indicates that the interpretation of S corresponds to the strongest alternative, i.e., the one associated with the most restricted pragmatic halo.

For instance, in (13a), repeated as (25), the CQ_c contains alternatives that differ in the

precision with which the universal expression 'everybody' is interpreted. Under a relatively imprecise interpretation, (25) may be judged true even if some individuals in the group did not come. In contrast, under the most precise interpretation, corresponding to the most restricted pragmatic halo, (25) judged true just in case every single individual in the group came. =*ē* selects this alternative, so that 'everybody' receives its strictest interpretation.

(25) ellār=um=ē va-nd-ānga
everyone=ADD=*ē* come-PST-3.PL.HUM
'*Absolutely* everybody came.' [⇝ *Every single person in the group came.*]

Finally, recall from §3.4 that when =*ē* cliticises to spatial and temporal expressions, precisifying interpretations obtain especially when there is epistemic uncertainty among interlocutors about the precise answer to the CQ_c (13d). This is explained by the fact that uncertainty regarding times and locations easily lends itself to a salient CQ_c such as "When/where exactly *X*?", in which the alternatives are ordered by their precision in answering the CQ_c.

5 Conclusion

This paper has offered a formal account of the Tamil enclitic =*ē*. I argued that =*ē* should be analysed as two distinct lexical items, $=\bar{e}_1$ and $=\bar{e}_2$, corresponding to =*ē*'s sentence-medial and sentence-final uses, respectively. This is motivated by the fact that only $=\bar{e}_1$ is obligatorily sensitive to focus, i.e., only $=\bar{e}_1$ conventionally associates with the Current Question (Beaver & Clark 2008). I then outlined the empirical range of $=\bar{e}_1$ and proposed that it is a scalar operator: $=\bar{e}_1$ associates its prejacent with the highest-ranked true alternative in the Current Question. The various discourse effects associated with $=\bar{e}_1$ arise depending on different types of alternatives in the Current Question, different strength rankings, and other contextual factors. In particular, drawing on Greenberg (2020), I proposed that $=\bar{e}_1$ is compatible with alternatives varying only in the instantiations of covert parameters, including contextual standards for gradable expressions and Lasersohnian pragmatic halos.

I conclude with a brief discussion on diachrony. As Chevillard (1997) and Arokianathan (1981) note, in Classical Tamil texts, =*ē* attests a wider range of uses than it has today. In particular, Old Tamil =*ē* had certain discourse functions that in Modern Tamil are performed by the enclitic =*dān*. In fact, impressionistically, Old Tamil =*ē* appears to have borne greater resemblance to "emphatic" focus particles in other South Asian languages, such as its Telugu cognate =*ē*, or Hindi =*hī*. I speculate that the grammaticalisation of =*dān* (from the anaphor *tān*) in post-Classical Tamil may have led to =*dān* in essence "taking over" some of the functional load of Old Tamil =*ē*, as a result of which Modern Tamil =*ē* has only a subset of the discourse functions of its South Asian counterparts

Acknowledgements

I am grateful to Ramprashanth Venkatakrishnan for detailed discussions on the data during this project. Thanks to Awais Hussain for scanning and sharing S. Arokianathan's book on

Tamil clitics. I also thank three anonymous reviewers and the attendees at FASAL-15 for their helpful comments and suggestions.
Contact: gopalkrishnan11251@gmail.com

References

Arokianathan, S. 1981. *Tamil clitics*. Dravidian Linguistics Association.

Asher, Ronald E. 1989. *Tamil*. Routledge 2nd edn.

Beaver, David & Brady Z. Clark. 2008. *Sense and sensitivity: how focus determines meaning*. Blackwell. doi:10.1002/9781444304176.

Chen, Zhuang & Yael Greenberg. 2024. An Overt Operator Associating with Covert Focus: An *Even*-Like Semantics for Mandarin *Gèng*. In Daisuke Bekki, Koji Mineshima & Elin McCready (eds.), *Proceedings of LENLS 20*, 134–154. Springer. doi:10.1007/978-3-031-60878-0_8.

Chevillard, Jean-Luc. 1997. Les particules énonciatives -ee et -taan en tamoul. *Faits de langues* 5(10). 201–208. doi:10.3406/flang.1997.1186.

Coppock, Elizabeth & David Beaver. 2014. Principles of the Exclusive Muddle. *Journal of Semantics* 31(3). 371–432. doi:10.1093/jos/fft007.

Deo, Ashwini. 2020. The strongest assertable alternative: Marathi =*c* and its counterparts. Presentation at SALT 2020, Cornell. https://osf.io/u3wf6.

Deo, Ashwini. 2023. Coordinated on the context: the many uses of Marathi =*ts*. In Akshay Aitha, Matteo Fiorini & Sreekar Raghotham (eds.), *Proceedings of (Formal) Approaches to South Asian Languages 12*, 65–84. https://ojs.ub.uni-konstanz.de/jsal/index.php/fasal/issue/view/fasal12proceedings.

Greenberg, Yael. 2020. An Overt *Even* Operator over Covert-Based Focus Alternatives: The Case of Hebrew *BIXLAL*. *Journal of Semantics* 37(1). 1–42. doi:10.1093/jos/ffz010.

Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of Language* 10(1). 41–53. doi:10.1016/B978-0-12-545850-4.50014-5.

Kennedy, Christopher. 2007. Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30(1). 1–45. doi:10.1007/s10988-006-9008-0.

Kennedy, Christopher & Louise McNally. 2005. Scale Structure, Degree Modification, and the Semantics of Gradable Predicates. *Language* 81(2). 345–381. doi:10.1353/lan.2005.0071.

König, Ekkehard & Volker Gast. 2006. Focused assertion of identity: A typology of intensifiers. *Linguistic Typology* 10(2). doi:10.1515/LINGTY.2006.008.

Lasersohn, Peter. 1999. Pragmatic Halos. *Language* 75(3). 522–551. doi:10.2307/417059.

Lehmann, Thomas. 1993. *A Grammar of Modern Tamil*. Pondicherry Institute of Linguistics and Culture.

McFadden, Thomas & Sandhya Sundaresan. 2014. Finiteness in South Asian languages: an introduction. *Natural Language & Linguistic Theory* 32. 1–27. doi:10.1007/s11049-013-9215-7.

Murugaiyan, Appasamy. 2009. Focus Constructions in Modern Tamil. *International Journal of Dravidian Linguistics* 38(2). 43–69. https://shs.hal.science/halshs-00664554v1.

Orenstein, Dina & Yael Greenberg. 2021. Approximation derived from a scalar exclusive particle associating with covert focus: The case of Hebrew *be-sax ha-kol*. *Glossa: a journal of general linguistics* 6(1). doi:10.5334/gjgl.1036.

Roberts, Craige. 2012. Information Structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6). 1–69. doi:10.3765/sp.5.6.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116. doi:10.1007/BF02342617.

Rullmann, Hotze. 1997. Even, Polarity, and Scope. In Martha Gibson, Grace Wiebe & Gary Libben (eds.), *Papers in Experimental and Theoretical Linguistics. Vol. 4*, 40–64. University of Alberta. https://semanticsarchive.net/Archive/WZhOWY5N/.

Schneider, Johannes. 2022. Three German discourse particles as speech act modifiers. In Xabier Artiagoitia, Arantzazu Elordieta & Sergio Monforte (eds.), *Discourse Particles: Syntactic, semantic, pragmatic and historical aspects*, 229–254. John Benjamins. doi:10.1075/la.276.09sch.

Velleman, Dan, David Beaver, Emilie Destruel, Dylan Bumford, Edgar Onea & Elizabeth Coppock. 2012. It-clefts are IT (Inquiry Terminating) constructions. In Anca Chereches (ed.), *Proceedings of SALT 22*, 441–460. LSA. doi:10.3765/salt.v22i0.2640.

Venkatesan, Madhusmitha. 2025. Keeping the syntax of predication and modification distinct: The view from Tamil. In *Papers from the International Workshop on the Syntax of Predication and Modification 2024*, 159–175. Kobe University. doi:10.24546/0100492873.

Wiegand, Mia. 2018. Exclusive Morphosemantics: Just and Covert Quantification. In Bennett, William, Lindsay Hracs & Dennis Ryan Storoshenko (eds.), *Proceedings of WCCFL 35*, Cascadilla Proceedings Project. https://lingref.com/cpp/wccfl/35/abstract3415.html.

Prosodic Licensing of /o/ in Katki and Sambalpuri: A nonce-based study

PRATIKSHYA GURU, *Indian Institute of Technology, Delhi*
ANUSUYA NAYAK, *Indian Institute of Technology, Delhi*
PAROMA SANYAL, *Indian Institute of Technology, Delhi*

ABSTRACT
Katki and Sambalpuri, two closely related varieties of Odia, exhibit microvariation in the phonological environments in which they allow the vowel /o/ to surface. This paper explores the Licensing Conditions on this vowel in the phonological grammar of Odia by asking three questions. First, what is special about the phonological composition of the vowel /o/ within the repertoire of Odia; second, what is special about the leftmost and rightmost syllables in Katki and Sambalpuri; and finally, is this phonological phenomenon a consequence of diachronic sound change, or is it reflected in the synchronic grammar of these languages. By comparing the cognates across the two varieties, we first establish what happens to the underlying /o/ in contexts where they do not meet the licensing conditions. Then, analyzing the lexical data from these languages we propose that contrastive characteristics of phonemes, such as headedness associated with Elements, is preserved in marked positions but neutralized elsewhere. Further a nonce-word based language game reveals that the synchronic grammar of Katki Odia does not replicate the phonological output observed in the lexical words.

1 Introduction

In this paper we will discuss the central role played by *Licensing Conditions* in phonological systems. The assumption underlying the idea of Licensing Condition is that, unless there is some phonological factor blocking its expression, a consonantal or vocalic phoneme should be present in all positions in a phonological string that corresponds to consonants or vowels. In other words, if /a/ is a phoneme in a language, then it should be attested in all possible phonological contexts where a vowel can be present. A Licensing Condition thus describes what is absent in the empirical data and the factors that block its expression in that environment.

Licensing Conditions on the expression of vowels have been discussed in preceding literature on the Eastern Indo-Aryan (EIA) language Bangla, where the vowel /a/ is severely restricted in disyllables where the preceding syllable contains the vowels /i/ or /u/ (Sanyal, 2010). The Eastern dialects of Bangla that do not have this Licensing Condition continue to have the vowel sequences [i-a] and [u-a] in surface representation. But the South-western variety has devised alternative pronunciations for such words that avoid these sequences.

(1) Licensing Condition on /a/ in Bangla

Nouns				**Verbs**	
d̪ukan~d̪**o**kan	‘shop’	šit̪ar~š**e**t̪ar	‘sitar’	ɖʰuk-a-no~ɖʰ**o**k-a-no	‘enter.cause’
šut̪a~šut̪**o**	‘thread’	pʰit̪a~pʰit̪**e**	‘tape’	šik-a-no~ š**e**k-a-no	‘learn.cause’

A similar licensing condition on the vowel /a/ is seen in the inflectional paradigm of Katki Odia where two consecutive syllables with the vowel /a/ are avoided (Sanyal 2025). Although Katki Odia has some lexical items with the vowel /a/ in consecutive syllables, such as /ʈala/ ‘lock’ and /ɟama/ ‘dress’, when compared to cognates from Bangla and Hindi, other IA languages, the distributional restriction becomes apparent. This is further seen clearly in the inflectional domain of Katki verbs (Nayak 2023). These are shown in (2).

(2) Licensing Condition on /a/ in Katki Odia

Nouns			**Verbs**	
Katki Odia	Bangla/Hindi cognates		ɖak	‘call’
cʰɔʈa	cʰaʈa	‘umbrella’	ɖak-e	‘call.1p’
ɔʈa	aʈa	‘flour’	ɖɔk-a-e	‘call.cause.1p’

These examples show that Licensing Conditions that restrict the distribution of certain vowels are not uncommon in the phonology of EIA languages. Unlike other Indo-Aryan languages that typically show phonemic length distinction in the vowels /i/, /u/ and /a/, EIA languages lack phonemic length distinction altogether.

In this paper, we focus on the distribution of the vowel /o/ with respect to specific Licensing Conditions in the Sambalpuri and Katki varieties of Odia. Since the phenomenon that we seek to describe entails the absence of a phoneme x from the surface representation in a particular phonological context, methodologically, we will compare cognates across related languages and varieties to establish that the phoneme was indeed present at some level of representation before some phonological factors blocked its expression in the surface representation. Consequently, we have chosen to compare cognates across the two varieties of Odia, Sambalpuri and Katki. As the same vowel /o/ has different distributional restrictions in these two varieties, comparing the cognates helps us visualize the Licensing Conditions on /o/ in Odia. The empirical data pertaining to this is described in §1.1.

1.1 Comparison of cognates

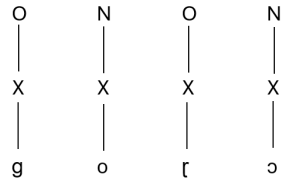
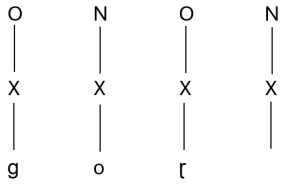
The vowel /o/ is a phoneme in both Sambalpuri and Katki forming near minimal pairs. Both varieties share the same vowel repertoire /i, e, a, ɔ, o, u/ (Pandey 2014, Sahu 2001) which includes three middle vowels: one front and two back. (3) shows that the vowel /o/ is found in contrastive distribution in both the languages.

(3) Contrastive distribution of /o/ with other vowels in both varieties

	Katki				**Sambalpuri**			
/o/-/ɔ/	corɔ	‘thief’	cɔrɔ	‘to gaze’	cor	‘thief’	cɔr	‘gaze’
/o/-/a/	goʈɔ	‘leg’	galɔ	‘cheek’	goʈ	‘leg’	gal	‘cheek’
/o/-/u/	koɭɔ	‘lap’	kuɭɔ	‘dynasty’	kol	‘lap’	kul	‘dynasty’
/o/-/i/	bolɔ	‘talk’	bilɔ	‘burrow’	bol	‘talk’	bil	‘burrow’
/o/-/e/	toɭɔ	‘to weigh’	telɔ	‘oil’	tol	‘to weigh’	tel	‘oil’

When we compare the Katki words with their cognates in Sambalpuri in (3) it becomes apparent that these words are underlyingly monosyllables. As Katki does not allow phonological strings to end in consonants, an epenthetic vowel /ɔ/ is inserted in the string final position in Katki. A visualization for this is shown in (4) using CV Theory (Lowenstamm, 1996).

(4) Visualization of monosyllabic CVC strings in Katki and Sambalpuri

Katki	**Sambalpuri**
	
No Licensing of final Empty Nucleus	Toleration of final Empty Nucleus

Comparison of the cognates for underlyingly disyllabic words between the two languages reveals the differences in the licensing of /o/. This has been shown in (5) and (7) below.

(5) Alignment to the Left Edge

Katki	**Sambalpuri**	
ɡora	ɡura	'fair.Masc'
ɟoɽe	ɟuɽe	'pair'
ɖoɭi	ɖuli	'cradle'

In Sambalpuri there are no polysyllabic words that have the vowel /o/ in the leftmost syllable. On contrary, in Katki such words are commonly found. When we compare the Katki words with /o/ in the leftmost syllable to their cognates in Sambalpuri, the /o/ seems to have been systematically replaced by the vowel /u/. Note that the dataset presented in (5) lends itself to two symmetric generalizations. First, the underlying vowel in the leftmost syllable for the words is /o/. Second, the underlying vowel in these positions is /u/. Based on the former position we can state the rule for Sambalpuri as [o→u/ σ1]. Similarly, based on the latter position we can also state the rule for Katki as [u→o/ σ1]. To select between these two descriptive positions, we need to observe whether the vowels /o/ and /u/ are attested in the σ1 position in Sambalpuri and Katki. In disyllables, while there are no Sambalpuri words with /o/ in σ1, there are plenty of Katki words with /u/ in σ1 position. This is shown in (6) below.

(6) Katki words with /u/ in σ1 position

Katki	**Sambalpuri**	
suta	suṯa	'thread'
sure	sure	'rhythmic'
muṛi	muṛi	'popped rice'

The data presented in (6) establishes the following two points. First, the vowel /u/ in Sambalpuri σ1 does not generally correspond to the vowel /o/ in cognate vocabulary from Katki. And second, most /o/ in Katki σ1 position corresponds to the vowel /u/ in Sambalpuri cognates. Just as the vowel /o/ is restricted to the leftmost syllable in Katki, /o/ is restricted to the rightmost syllable in Sambalpuri.

(7) Alignment to the Right Edge

Katki	**Sambalpuri**	
seu	seo	'apple'
hɔu	hɔo	'okay'

Comparing the rightmost syllable of lexical words across Sambalpuri and Katki is not as straightforward as comparing the leftmost syllables. As discussed in (4), contrary to Sambalpuri, which can have C-final words, Katki adds an epenthetic vowel to all C-final lexical words making that the last syllable. Such as, a disyllabic C-final word like [kɔməl] 'lotus', in Sambalpuri corresponds to the Katki cognate [kɔmə[ɔ], a trisyllable. So, for every penultimate Sambalpuri syllable, the corresponding anti-penultimate Katki syllable is to be observed. Hence, for the Sambalpuri word [kɔ(məl)], we shall observe the Katki word [kɔ(mɔ)[ɔ]. So, we are comparing the rightmost syllable in V-final words from Sambalpuri to the penultimate syllable of the corresponding Katki cognates.

Similar to the phonological pattern in (5) and (6), we find that the vowel /o/ is never realized in the rightmost syllable of polysyllabic words in Katki, but /o/ is readily attested in this position in Sambalpuri. Further, the rightmost /o/ in Sambalpuri corresponds to the vowel /u/ in Katki cognates. Hence, concluding that both varieties of Odia show the same change in vowel quality [o→u] in different environments. An interim generalization from this section is summarized in (8).

(8) Interim generalization 1

Language	Rule-based description	
Katki Odia	o→u/__(C)#	/o/ changes to /u/ in the rightmost syllable
Sambalpuri Odia	o→u/ # (C)__	/o/ changes to /u/ in the leftmost syllable

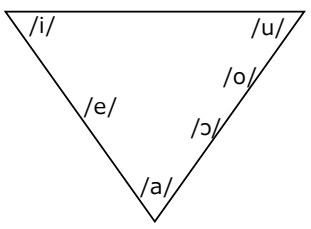
The following section discusses why these phonological changes take place in these environments.

1.2 Theoretical discussion on phonological change

There are three distinct aspects to the phonological change in vowels discussed in §1.1. First, the phonological process affects the initial and final syllables in the two varieties of Odia. What is special about these boundary positions and why do they trigger change? Second, we have deduced phonological change by comparing cognates across varieties. The phonological change does not tell us whether the process was part of a diachronic sound change or is it current in the phonological grammar of these languages. Third, among the six-vowel repertoire of Odia, only the vowel /o/ is subject to Licensing Conditions. Why does the same phonological environment not affect other vowels? Of these three aspects, we will discuss the last one in this section and come back to the first and second aspect in §2 and §4 of this paper.

Odia has six vowel repertoire: two front vowels (/i/, /e/), three back vowels (/u/, /o/ and /ɔ/) and a low vowel (/a/). If we look at the schematized placement of these vowels in the cardinal vowel space, the back appears to be more crowded in comparison to the front. Similarly, if we describe the six vocalic melodies of Odia using Element Theory (Backley, 2011), then headedness[1] appears to become crucial for phonemic identity only in case of the mid-back vowels, one of which is /o/. (9) below shows both these visualizations for the six-vowel repertoire of Odia.

(9) Visualization of the six-vowel phonemic inventory of Odia



Phoneme	Melody
/i/	|I|
/u/	|U|
/a/	|A|
/e/	|IA|
/o/	|U̲A|
/ɔ/	|UA̲|

For the first three vowel phonemes in (9), /i/, /u/ and /a/, headedness is not a crucial distinctive characteristic since the language does not contain any other phoneme with the headed |I̲|, |U̲| and |A̲| that would contrast with the non-headed |I|, |U| and |A|. Similarly, there is only one phoneme with the melodic combination |IA| in Odia. As there is no phonemic contrast between |I̲A| and |IA̲|, the headedness parameter is not distinctive for this context either.

In contrast to these cases, the headedness parameter becomes meaningful to describe the contrast between the phonemes /o/ and /ɔ/ since both of them are expressed with the melodic combination |UA|. While /o/ is expressed with a headed |U|, as in |U̲A|, /ɔ/ is expressed with a headed |A|, as in |UA̲|. Considering one of the phonemes were to be restricted to a particular position in the word, the headedness parameter would be restricted to that position. Whereas, elsewhere in the language the melodic tier would not have to be specified by headedness at all.

Of these two vowels with the melodic combination |UA|, the one with the headed |A| element forms the epenthetic vowel /ɔ/ in Katki and vocalises the empty nucleus in

[1] Headedness is marked by underlining the head element.

word-final positions (4). Being the word final epenthetic vowel in Katki, /ɔ/ cannot be restricted to a specific phonological position. Consequently, the other vowel with the |U̲A| melody, /o/ gets restricted to leftmost syllables.

Since the vowel /o/ is restricted to the leftmost syllable in Katki, one of the possibilities would be to assume that the headedness parameter is restricted to this position and does not get expressed elsewhere. However, this assumption would predict that the vowels |U̲A| and |UA̲| both get expressed the leftmost syllable but lose the headedness parameter and get expressed as |UA| elsewhere. This approach would entail that we revise the element-based configuration of the six-vowel repertoire and consider the vowel /o/ to be the only vowel that is phonologically specified for the headedness parameter. This is shown in (10).

(10) Revised description of the phonemic inventory of Odia

Phoneme	Melody
/i/	|I|
/u/	|U|
/a/	|A|
/e/	|IA|
/ɔ/	|UA|
/o/	|U̲A|

With this theoretical description, we propose that the Licensing Conditions on the vowel /o/ in Katki and Sambalpuri is in actuality a Licensing Condition on the headedness parameter. The other vowels in the phonemic repertoire do not exhibit distributional restrictions as they do not require the expression of the headedness parameter.

While the theoretical reformulation of the vowel inventory in (10) successfully explains why the vowel /o/ is the only vowel in the repertoire with such distributional restrictions, there is a remaining issue that it does not answer. If the headedness parameter can only be expressed in the specific phonological position, then elsewhere the loss of headedness should have changed the /o/ to /ɔ/. Instead, the data in §1.1 shows that elsewhere the /o/ changes to /u/. This issue will be discussed further in §4 and §5 of the paper.

2 An argument for Prosodic Licensing

In §1.1 of this paper we showed that the phoneme /o/ is restricted to the leftmost and the rightmost syllables in Katki and Sambalpuri, respectively. Eventually we revised this observation in §1.2 with the proposal that the headedness parameter is restricted to the leftmost syllable in Katki and rightmost syllable in Sambalpuri. This results in the vowel /o/, |U̲A|, to be the only vowel in the repertoire with an expression of headedness, having positional restrictions. Following this line of argumentation, in this section, we will explore why the headedness parameter is restricted to distinct positions in the two varieties of Odia.

In the Optimality Theory approach (Prince & Smolensky, 1993) to phonology, there are correspondence constraints like ANCHORING-IO (McCarthy & Prince, 1993a) that militate against any non-correspondence between the input and output segments at the designated edge. In doing so, these constraints preserve marked configurations at these locations that are otherwise lost in the string internal positions. However, the skeletal position associated with the vowel in the leftmost or rightmost nucleus in Katki and Sambalpuri is not a peripheral position in the skeletal tier and hence cannot be preserved with these formal mechanisms. Formally, one could re-calibrate ANCHORING-IO constraint to the domain of the edgemost syllable instead of segment. This adapted formal device will preserve the headedness specified melody in the Anchored position and bleach the headedness parameter elsewhere. This has been shown in (11) below.

(11) Anchoring-based formal analysis for positional restriction on /o/

Anchoring-IO : Any *segment* at the designated edge of the input has a correspondent at the same edge of the output.	Original formulation
Anchoring-IO : Any *syllable* at the designated edge of the input has a correspondent at the same edge of the output.	Revised formulation

The headedness parameter introduces one more structural device into the configuration of the melodic tier. Therefore the segment |U̲A| is more complex than the segment |UA|. |U̲A| is marked for headedness, |UA| is not. This is the reason we proposed that the headedness parameter is restricted to the edgemost syllable in §1.2. This idea is converted into the formal device of a markedness constraint in (12).

(12) Markedness constraint

*HEAD: Do not use headedness distinction in the melodic tier.

The ranking of the markedness constraint *HEAD lower than the positional faithfulness constraint ANCHORING-IO results in the preservation of headedness parameter in the nucleus of the edgemost syllable. This has been shown for the schematized disyllabic inputs [o-i] and [i-o] in (13) and (14), respectively.

(13) Input representation [o-i] with the vowels /o/ and /i/ in adjacent syllables

Katki		
|U̲A|-|I|	ANCHOR-L	*HEAD
a. ☞|U̲A|-|I|		*
b. |UA|-|I|	*!	

Sambalpuri		
|U̲A|-|I|	*HEAD	ANCHOR-L
a. |U̲A|-|I|	*!	
b. ☞|UA|-|I|		*

The two evaluation tables in (13) follow the theoretical approach of Standard Optimality Theory (Prince & Smolensky, 1993) with ranked constraints. The two output candidates are evaluated with respect to the markedness constraint *HEAD and Anchor-L that were discussed in (11) and (12). Candidate (a) violates the markedness constraint *HEAD, since it contains a headed melody |U̲A| in the output. Similarly, candidate (b)

incurs a violation of the faithfulness constraint Anchor-L as the output representation |UA| does not match the input representation |U̲A|. In Katki, Anchor-L being ranked higher than *HEAD, the headedness of the input |U̲A| is preserved in the output. In Sambalpuri, *HEAD being ranked higher than Anchor-L, the input |U̲A| corresponds to the optimal output |UA| without headedness. A similar evaluation for the input [i-o] with respect to Anchor-R and *HEAD reveals how headedness is preserved in the right edge in Sambalpuri. This is shown in (14).

(14) Input representation [i-o] with the vowels /i/ and /o/ in adjacent syllables

Katki		
|I|-|U̲A|	*HEAD	ANCHOR-R
a. |I|-|U̲A|	*!	
b. ☞|I|-|UA|		*

Sambalpuri		
|I|-|U̲A|	ANCHOR-R	*HEAD
a. ☞|I|-|U̲A|		*
b. |I|-|UA|	*!	

In (14) the input has the headed vowel /o/ in the rightmost syllable of the word. So, the constraints used for evaluation are *HEAD and Anchor-R. In Sambalpuri, candidate (a) preserves the headed input and emerges as the optimal output, ranking Anchor-R higher than *HEAD. However, in Katki, the faithfulness constraint Anchor-R ranked lower than *HEAD, hence, making candidate (b) as the optimal output.

While the evaluation tables in (13) and (14) successfully explain the preservation of the marked segment /o/ in the leftmost and rightmost syllable positions of Katki and Sambalpuri words, it does not accurately predict the vowel that the an underlying /o/ change to. By simply removing the headedness parameter from non-anchored positions, the analysis predicts that the underlying /o/ will surface as /ɔ/. However, based on the comparison of cognate vocabulary in §1.1 we have established that the underlying /o/ changes to the vowel /u/.

The change from input |U̲A| to output |U| entails loss of two melodic properties: headedness and the element |A|. Based on the element-based description of the phonemic inventory of Odia shown in (10), the element |A| is part of the melodic composition of the three other phonemes other than /o/. These are, /a/-|A|, /e/-|IA| and /ɔ/-|UA|. None of these three phonemes undergo loss of the element |A| in any phonological context in Katki or Sambalpuri. Therefore, the loss of the element |A| from the input |U̲A| is specifically linked to the presence and loss of headedness in these languages.

The formal mechanism to incorporate this dependency is to locally conjoin two markedness constraints *|A| and *HEAD. Local Conjunction (Smolensky, 1995) is that idea that a conjoint constraint A+B is violated if and only if both of its constituent constraints A and B are simultaneously violated within the specified local domain. In this case, the melodic combination *|A| and *HEAD is simultaneously violated by the output candidate |U̲A|. This formulation has been shown in (15).

(15) Local Conjunction of constraints

Constituent 1	*|A|: Do not have the element |A| in the melodic tier of a vowel segment.
Constituent 2	*Head: Do not use headedness distinction in the melodic tier of a vowel.
*|A|+*HEAD: Do not violate both constraints simultaneously in a single vowel segment	

One of the possible strategies to avoid violating this conjoint constraint is the loss of headedness. This was the case of the optimal outputs in (13) and (14). Another strategy to avoid violating this constraint is the loss of |A|. In the revised analysis presented in (16) and (17) we have introduced two additional faithfulness constraints Ident-HEAD and Ident-|A| pertaining to the two crucial melodic properties. The evaluation tableaux presented in (16) and (17) has two sub-parts each showing a possible outcome.

(16) Revised Evaluation tables for Katki, for the inputs [o-i] and [i-o]

|U̲A|-|I|	ANCHOR-L	*|A|+*HD	ID-HD	ID-|A|	*|A|	*HD	ANCHOR-R
a. ☞|U̲A|-|I|		*			*	*	
b. |UA|-|I|	*!		*		*		
c. |U̲|-|I|	*!			*		*	
d. |U|-|I|	**!		*	*			
|I|-|U̲A|	ANCHOR-L	*|A|+*HD	ID-HD	ID-|A|	*|A|	*HD	ANCHOR-R
a. |I|-|U̲A|		*!			*	*	
b. |I|-|UA|			*		*!		*
c. ☞|I|-|U̲|				*		*	*
d. |I|-|U|			*	*!			**

(17) Revised Evaluation tables for Sambalpuri, for the inputs [o-i] and [i-o]

|U̲A|-|I|	ANCHOR-R	*|A|+*HD	ID-HD	ID-|A|	*|A|	*HD	ANCHOR-L
a. |U̲A|-|I|		*!			*	*	
b. |UA|-|I|			*		*!		*
c. ☞|U̲|-|I|				*		*	*
d. |U|-|I|			*	*!			**
|I|-|U̲A|	ANCHOR-R	*|A|+*HD	ID-HD	ID-|A|	*|A|	*HD	ANCHOR-L
a. ☞|I|-|U̲A|		*			*	*	
b. |I|-|UA|	*!		*		*		
c. |I|-|U̲|	*!			*		*	
d. |I|-|U|	**!		*	*			

In the revised evaluation tables for Katki and Sambalpuri, (16) and (17), the constraints are ranked into three sets. First, the Anchoring-IO constraint dominates the conjoint

constraint and preserves the marked melodic combination |U̲A| in the designated edge. Second, the conjoint constraint outranks the faithfulness constraints triggering deletion of either headedness or |A| in candidates (b) and (c). Candidate (d) is ruled out since it violates both the faithfulness constraints simultaneously. Finally, the mutual ranking between the markedness constraints *|A| and *HD, determines (c) to be a better candidate than (b).

In this section we have proposed that the leftmost syllable of the prosodic word is preserved in Katki Odia just as the rightmost syllable is preserved in Sambalpuri Odia. While we have used the formal device of Anchoring-IO in the evaluation tables, this constraint does not clearly explain why the edgemost syllables would be preserved. One possible case could be that these are the positions for primary stress assignment in these languages. This hypothesis however needs to be independently verified through the application of other stress sensitive phonological processes in these languages. Since describing such processes would be outside the scope of this paper, we will leave this as a suggestion.

Further, note that the optimal candidate for [o→u] change in (16) and (17) is the candidate (c) rather than the candidate (d). This makes the output for sound change, |U̲|, phonologically distinct from the underlying vowel |U|. Analytically, (d) would emerge as the optimal output if and only if the markedness constraint *|A| is ranked higher than the faithfulness constraint Id-|A|. This ranking of these two constraints would trigger the deletion of the |A| melody from other vowel phonemes such as |UA| and |IA|. This is not empirically true for either of the languages. Thus, given our analytical proposal the vowel /u/ that emerges because of sound change from an underlying /o/ is predicted to be phonologically distinct from the vowel /u/ that appears lexically specified in the phonological string. In the following section, we show that this phonological distinction has empirical consequences in Sambalpuri and Katki.

3 Dissimilation in phonologically derived context

In this section we will discuss the case of blocking of [o→u] vowel raising in both Sambalpuri and Katki, when the prosodic word already has an underlying /u/ vowel. This has been shown in (18).

(18) Comparing cognates with underlying /u/

Katki		**Sambalpuri**		
Actual form	Expected form	Actual form	Expected form	
goru		gɔru	*guru	‘animal’
kʰu̲ɔ	*k(ʰ)u̲u	ku̲o		‘a type of rice’

Given the analysis in § 2, the data presented in (18) shows that an underlying |U̲A| (vowel /o/) changes to the surface form |UA| (vowel /ɔ/) instead of the expected surface form |U̲| (vowel /u/) when the other vowel in the lexical word is a non-headed |U| (vowel /u/). Thus, in Katki, the candidate output with the sequence |U|-|UA| is preferred over the candidate output with the sequence |U|-|U̲|. Similarly, in Sambalpuri the output sequence |UA|-|U| is evaluated to be more optimal than the output |U̲|-|U|. The blocking

of two segments with similar properties within the same phonological domain is referred to as Obligatory Contour Principle (OCP) (Leben 1973).

(19) Formal definition of OCP

Obligatory Contour Principle: At the melodic level, adjacent identical elements are prohibited.

(McCarthy 1986)

However, this dispreference for the sequence of two consecutive /u/ vowels is not reflected in the underived words of either Katki or Sambalpuri. This is shown in (20).

(20) Underived words with two consecutive /u/

Katki			**Sambalpuri**		
Actual form	Expected form		Actual form	Expected form	
guru	*gurɔ/*gɔru	‘teacher’	kukur	*kukɔr/ *kɔkur	‘dog’
			munus	*munɔs/*mɔnus	‘human’

This scenario where a phonological process does not apply to underived contexts but emerges a well-formedness requirement in the derived context is described as an outcome of the Strict Cycle Condition by Kiparsky (1982). The SCC states that a cyclic phonological rule only applies to contexts that have been either phonologically or morphologically derived. In (18) and (20) we see that OCP emerges as a phonological process in Katki and Sambalpuri if and only if the underlying phonological string started out with an /o/ vowel in the input. It does not apply to underived domains.

Theoretically, SCC includes both the phonologically derived and morphologically derived contexts within the ambit of the term derived environments. However, in case of Odia, both Katki and Sambalpuri, we find that OCP is not triggered by morphologically derived environments with two consecutive vowels either. This is shown in (21).

(21) Inflected verbs with two consecutive /u/

Katki	**Sambalpuri**	
c^hũ-uc^hi	c^hu-uc^he	‘touch.prog.1p’
buj^h-uc^hi	buj^h-uc^he	‘understand.prog.1p’

In (21), the CV and CVC verb roots with the lexically specified /u/ is immediately followed by the inflectional suffix for progressive aspect, also beginning with an /u/. Yet, despite the two vowels appearing in morphologically derived environments, neither the /u/ in the affix nor the /u/ in the verb root undergoes any modification in order to avoid OCP.

In summary, (18) shows that the potential for OCP violation blocks the [o→u] change in both Katki and Sambalpuri. However, the OCP violation is tolerated in both underived and morphologically derived environments, (20) and (21). Thus, the OCP in Odia treats the phonologically and morphologically derived environments differently.

This is contrary to the theoretical formulation of SCC that clubs together phonologically and morphologically derived environments as the domain for lexical rules.

If we re-analyze the dissimilation of /u/ in (18) in the context of our analysis for the [o→u] raising in §2, there is a distinction between the phonological environment of (18) and those in (20) and (21). As per the evaluation tables (16) and (17) and the two /u/ vowels in (18) are not identical. One of them is a headed |U̲| and the other a |U|, underspecified for headedness. In contrast both the /u/ vowels in (20) and (21) have the same melodic specification |U|. In summary, OCP applies when the two |U| are similar but does not apply when they are the same.

Empirical studies on the similarity effects of OCP have shown that the more similar two segments are, greater is the identity avoidance (Padgett 1991, 1992; Selkirk 1988, 1991, 1993; Yip 1989 and Pierrehumbert 1993). These studies predict that if the identity between two adjacent segments with the near identical melody |U̲| and |U| is being avoided, the language would also avoid identity between adjacent segments with completely identical melody |U|-|U|. However, this is not the case in Odia.

In (22) we show that OCP in Odia is violated by the output candidates |U̲|-|U| as well as |U|-|U|. However, this constraint being ranked lower than faithfulness constraint Id-|U|, does not have an impact on the lexically specified |U| in the input representation. OCP emerges as a crucial markedness factor in contexts where the effects of the faithfulness constraint Id-|U| is not relevant as the input segment is not an /u/.

(22) Evaluation tables for OCP effects in Katki and Sambalpuri

Katki									
|U|-|U̲A|	ANCHOR-L	*|A|+*HD	ID-HD	ID-|A|	ID-|U|	OCP	*|A|	*HD	ANCHOR-R
a. |U|-|U̲A|		*!					*	*	
b. ☞|U|-|UA|			*				*		*
c. |U|-|U̲|				*		*		*	*!
d. |U|-|U|			*	*!		*			**
e. |U|-|A|			*		*!		*		**
Sambalpuri									
|U̲A|-|U|	ANCHOR-R	*|A|+*HD	ID-HD	ID-|A|	ID-|U|	OCP	*|A|	*HD	ANCHOR-L
a. |U̲A|-|U|		*!					*	*	
b. ☞|UA|-|U|			*				*		*
c. |U̲|-|U|				*		*		*	*!
d. |U|-|U|			*	*!		*			**
e. |A|-|U|			*		*!		*		**

The evaluation tables in (22) are identical to the evaluation tables presented in (17) and (18) except for one factor. The output candidates (c) and (d) with a sequence |U̲|-|U| and |U|-|U| incur a violation of the markedness constraint OCP each. Candidate (e) is maximally distinct by avoiding the melody |U| altogether in the adjacent segments. However, this is blocked by two violations of faithfulness constraints. This the candidate (b), |UA| emerges as the optimal output.

A summary of the analysis proposed in this paper is presented in the (23).

(23) Interim generalization 2

Observation	*Analysis*
Element-based phonemic repertoire	Headedness parameter is contrastive in |UA|
Distributional restriction on /o/-|UA|	Headedness restricted to Leftmost or Rightmost
/o/→/u/ if not Leftmost/ Rightmost	ANCHOR-L/R >> *|A|+*HD >> IDENT-IO >> *|A| >>*HD
/o/→/ɔ/ if adjacent /u/	ANCHOR-L/R>> *|A|+*HD >>IDENT-IO>>OCP,*|A| >>*HD

While the data from the lexical words of Katki and Sambalpuri analysed so far paint a picture of a phonological grammar for these languages, it does not tell us whether this was a diachronic sound change or is it a synchronic process in the language. In some aspects the phonological phenomena under discussion here is similar to the case of Trisyllabic Laxing (TSL) in English. In SPE (Chomsky and Halle 1968) it was shown the long vowels in derived words like *serene* /səˈriːn/ *-serenity /səˈren.ə.ti /*, were shortened in the third syllable from the right edge but left unmodified in underived words like *ivory* and *nightingale*. Based on some of these observations theoretical architectures such as lexical phonology were proposed that involved a step-by-step interaction between morphology and phonology (Kiparsky 1982). Eventually, counter-examples such as *obesity* came to light that indicated that TSL might not have been a synchronic phonological process in English. It was a diachronic sound change that remained fossilized to certain sections of the lexicon.

Consequently, we wanted to check if the process of [o→u] raising as well as the [o→ɔ] lowering under OCP conditions is part of the synchronic grammar of Odia. In §4, we discuss the pilot study we undertook and its results.

4 Between the synchronic and diachronic processes

The native core vocabulary of a language often shows markedness restrictions that are not seen in loanwords. This observation by Ito and Mester (1995a) formed the foundation for subsequent studies on the phonology of grammaticalization processes in the lexicon. Although these lexicon stratification studies are based on meaningful lexical items, the question remains whether nonce words would be treated akin to the innermost or outermost layer of the lexicon. What we do know however is that the phonological processes associated with each stratum are supposed to be part of the synchronic phonological grammar of the language. Keeping that in mind, we designed a small nonce based elicitation task to check if the vowel /o/ raises to /u/ in non-anchored position.

4.1 Design of the study

The elicitation task had four parts. In the first part, we taught a simple syllable transposition game to the participants. They saw a disyllabic word CVCV with the

vowels /i/ and /a/ on the computer screen and were taught to read the word and then produce the game word where the two syllables were transposed. For example, [mika] should elicit [kami]. The second part was like the first but without prompts for the game word. This aimed to verify if the participants had learnt the transposition game.

In the third part, the participants were presented the actual test set of nonce words. In this set, the stimuli were categorized into two main types: disyllables, containing /o/ in both the initial or final syllables (e.g., CoCV and CVCo), and trisyllables, such as CVCVCo and CoCVCV. The design of the stimuli set is presented in (24).

(9) Stimuli patterns for the familiarity and game set

Familiarity set		Game Set			
CVCV		CVCV		CVCVCV	
CaCi	CiCa	CuCo	CoCu	CuCaCo	CoCaCu
CuCi	CiCu	CiCo	CoCi	CeCaCo	CoCaCe
CiCe	CeCi	CeCo	CoCe	CiCaCo	CoCaCi
CeCa	CaCe	CaCo	CoCa	CaCaCo	CoCaCa
CaCu	CuCa	CoCo		CɔCaCo	CoCaCɔ

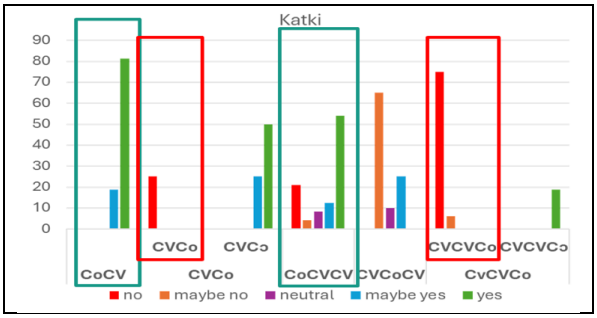
In the fourth part of the study, the participant was asked to evaluate whether the resulting transposed form could plausibly occur as a word in their respective variety of Odia. They did that using a 5-point Likert scale.

The pilot study was conducted with eight participants. Four native speakers of Katki and Sambalpuri.

4.2 Results of the study

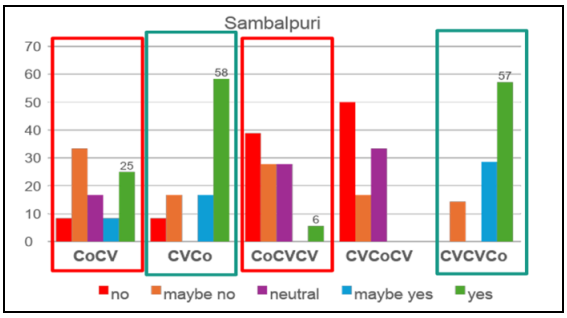
The study showed that as expected speakers of Katki showed higher acceptance rate for words with /o/ in the leftmost syllable and speakers of Sambalpuri showed a higher acceptance for rightmost /o/. The Katki speakers systematically lowered the /o/ to /ɔ/ in non-initial positions of both disyllables and trisyllables instead of the expected raising to /u/. They also remained unaware of this lowering of the vowel in their transposed utterances and gave a high rating to the transposed CVCo uttered by them as CVCɔ. The participants were less certain in trisyllables.

(10) Results for Katki speakers



In contrast to the Katki speakers, the Sambalpuri speakers accurately identified the offending vowel /o/ in non-rightmost syllable transposed contexts and reported gradient acceptability. There was a higher acceptance rate of /o/ in the rightmost syllable as compared to penult and a lower acceptance rate for /o/ in the anti-penult. Further, we observed that in the syllable transposition game, the participants altered the pattern in trisyllables to obtain a final /o/, e.g., CaCoCu is transposed as CuCaCo instead of CuCoCa.

(11) Results for Sambalpuri speakers



4.3 Analysis

The findings from the Pilot study suggest that the distributional restriction on /o/ in lexical words in both Katki and Sambalpuri is extended to nonce words. The vowel /o/ emerges as a marked vowel in both the varieties, licensed only in the anchored position. This shows that the anchoring constraint as well as the markedness of the headedness parameter are part of the synchronic phonological grammar of these languages. However, the strategy to avoid /o/ in non-anchored positions in nonce words is different from the strategy observed in lexical words.

In Katki lexical words, the vowel |U̲A| is preserved at the left-edge of the word due to the high ranking of Anchor-L. Elsewhere it changes to |U̲| in lexical words. Unlike the lexical words we see that the non-anchored /o/ emerges as |UA| in the nonce words even without an OCP context. This shows that the markedness constraint *|A|, active in the lexical domain is not active in the synchronic grammar of Katki. With *HEAD being the only active factor we have seen in (14), |UA| is expected to emerge as the optimal candidate. The evaluation table for this is repeated here in (27).

(12) Evaluation table for Katki nonce words

|I|-|U̲A|	*HEAD	ANCHOR-R
a. |I|-|U̲A|	*!	
b. ☞|I|-|UA|		*

Unlike Katki, the non-anchored /o/ in Sambalpuri, changes to either /u/ or /ɔ/ in the utterance of transposed words depending on the other vowels in the word. This shows

that the markedness constraint *|A| is active in the synchronic grammar of Sambalpuri. The relevance of Anchor-R is further underlined in the syllable transposition data from Sambalpuri participants who altered the pattern of syllable transposition in trisyllabic stimuli in order to place the syllable with /o/ in the word-final position.

5 Conclusion

This study demonstrates that the distributional restriction of the vowel /o/ in Katki and Sambalpuri varieties of Odia, is best accounted for as a case of prosodic and positional licensing, constrained by the headedness feature in the melodic tier. Comparative analysis of cognates reveals that the vowel /o/, characterized by a headed element |U|, is a marked vowel and it surfaces exclusively in edge positions: in the leftmost syllables in Katki and rightmost syllable in Sambalpuri.

This pattern is effectively explained by the interaction of the faithfulness constraints such as Anchoring-IO which preserves the marked segment in the edgemost syllables. These conclusions are further corroborated by the nonce-word experiments, confirming that the licensing conditions are active components of the synchronic phonological grammar.

The theoretical model predicts that in unmarked, non-edge positions, the headedness feature is lost, resulting in the surfacing of the non-head counterpart of the vowel /o/, *i.e*., /ɔ/ which is composed of |UA|. However, the empirical data contradicts this prediction in lexical words, the vowel /u/ characterized by the element |U| surfaces in all non-anchored positions. This represents not just a loss of headedness but also a reduction in melodic complexity. The interaction of various faithfulness and markedness constraints reveal this pattern. Notably, the non-head counterpart /ɔ/ emerges only when the lexical word already contains /u/, indicating a contextually conditioned alternation.

Furthermore, the study elucidates the role of Obligatory Contour Principle in interaction with other markedness and faithfulness constraints. It reveals how identity avoidance effect leads to dissimilation in phonologically derived environments but is systematically circumvented in underived and morphologically derived contexts wherein adjacent [u-u] sequence, can freely occur. This nuanced behaviour challenged traditional assumptions within the Strict Cycle Condition that equates both phonologically and morphologically derived environments as domains for phonological processes.

Importantly, the analysis distinguishes between the lexical vowel /u/ (|U|) and the phonetically identical /u/ (|U̲|) that has emerged from an underlying /o/. While both surface as /u/ and act as a triggering environment for OCP, they are phonologically distinct: the lexical /u/ is specified for the non-headed element |U|, the derived /u/ retains the headed element |U̲|. This distinction is maintained by a higher ranked faithfulness constraint Ident-|U| which preserves a lexical /u/ in both phonologically underived and morphologically derived context.

These findings substantiate the theoretical premise that headedness is a contrastive feature in the mid-vowel space of Odia, providing formal insight into the microvariation observed in the varieties of Katki and Sambalpuri. Furthermore, by integrating element-based analysis accompanied with the edge anchoring constraints within the theoretical framework of Optimality Theory, the research advances

understanding of how segmental markedness interacts with positional licensing to account for segmental distribution and variation across related languages.

References

Backley, Philip. 2011. An Introduction to Element Theory. Edinburgh: Edinburgh University Press.

Chomsky, Noam and Morris Halle. 1968, The Sound Pattern of English, New York: Harper & Row

Itô, Junko, and Armin Mester. 1995. *The core-periphery structure of the lexicon and constraints on reranking*. In Papers in Optimality Theory, edited by J. Beckman, S. Urbanczyk and L. Walsh. Amherst: GLSA. 181-210.

Kiparsky, Paul. 1982. *Lexical morphology and phonology.* In: The Linguistic Society of Korea (ed.), Linguistics in the morning calm, 3-91. Seoul: Hanshin

Lowenstamm, Jean. 1996. *CV as the only syllable type*. Current trends in Phonology. Models and Methods, edited by Jacques Durand & Bernard Laks, 419-441.

McCarthy, John and Alan Prince. 1993. *Generalized alignment*. In Yearbook of morphology 1993

McCarthy, John. 1986. *OCP effects: Gemination and anti-gemination.* Linguistic Inquiry, 17: 207-263

Nayak, Anusuya. 2023. *Vowel alternations in inflectional paradigms of Odia and its dialects*. Paper presented at IIT Delhi as part of PhD Research Plan Presentation.

Pandey, Pramod. 2014. Sounds and their patterns in Indic languages. New Delhi: CUP India.

Padgett, Jaye. 1997. *Perceptual distance of contrast: vowel height and nasality*. Phonology at Santa Cruz, 5, 63-78.

Pierrehumbert, Janet B. 2016. *Phonological representation: Beyond abstract versus episodic.* Annual review of linguistics, *2*(1), 33-52.

Sahu, Gobardhan. 2001. Generative Phonology of Sambalpuri, PhD thesis, Sambalpur University

Sanyal, Paroma. 2010. *Prosodic well-formedness and vowel licensing in Bangla*. PhD. Dissertation. English and Foreign Languages University, Hyderabad.

Sanyal, Paroma. 2025. *Verb Root Allomorphy in Indo-Aryan Languages*. Proceedings of FASAL 12 held in Stony Brook University, New York, USA. Pg 184-203.

Selkirk, Elizabeth. 1991. *Major place in the vowel space: vowel height features*. Ms., University of Massachusetts, Amherst.

Realization of Gemination in Malayalam Past Tense Morphology

GREESHMA JOSEPH, Indian Institute of Technology, Delhi
PAROMA SANYAL, Indian Institute of Technology, Delhi

ABSTRACT
This paper explores the emergence of geminate stops in the morpho-phonology of past-tense inflection in Malayalam verbs. Malayalam has two allomorphs that mark past tense: [-i] and [-u]. While the former attaches to stems with the phonological structure CVCC, CVC and CV without any morpho-phonological sandhi between the stem and affix, the latter allomorph only attaches to CVCC stems with two kinds of morpho-phonological assimilation patterns. Further, the past tense of the causative form of all verbs in Malayalam is uniformly marked by the [-u] allomorph.

In the light of these observations, we propose that the past tense morpheme [-u] Malayalam is underlyingly two distinct morphemes with a consonant that is not linked to the skeletal tier. So, underlyingly these are [-Cu]. The properties of the C emerge on affixation. Using Element Theory (Backey 2011) we propose that this "floating" C is of two types. It is either a dental stop |I?| or a dental nasal stop |I?L|.

1 Introduction

Although Malayalam, a South-Dravidian language, has a phonemic contrast between voiced and voiceless obstruents[1], the voiceless counterpart does not surface as a singleton consonant in an intervocalic context. Similarly, the voiced counterpart does not surface as a geminate. This distributional restriction on obstruents in Malayalam has been shown in (1).

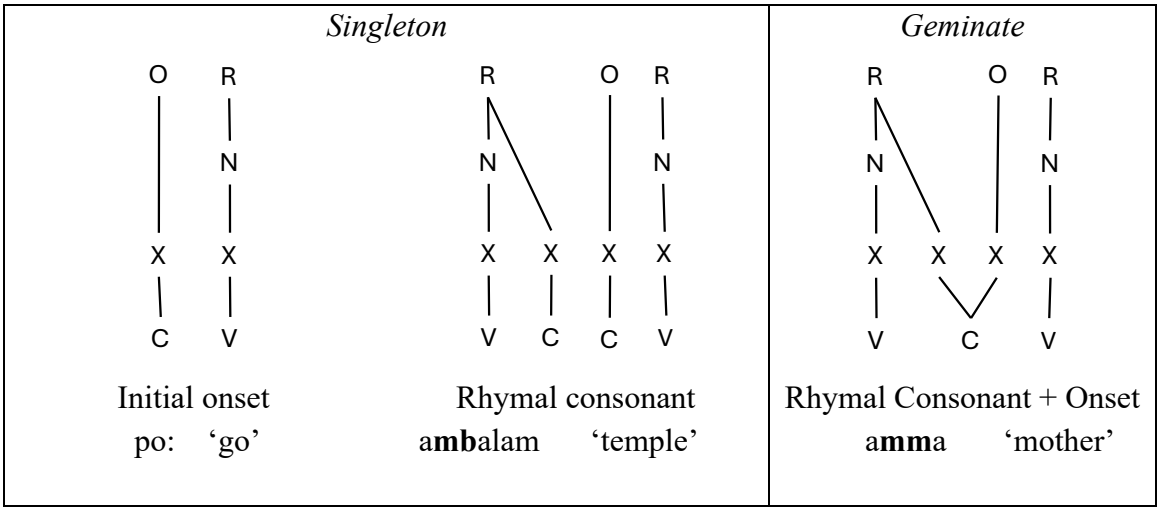
1. Distributional restriction on obstruents in Malayalam (Surface realization)

		Word initially	Intervocalic
Singleton Ç	[p, t̪, ʈ, c, k]	✓	×
Geminate ÇÇ	[pp, t̪t̪, ʈʈ, cc, kk]	×	✓
Singleton C	[b, d̪, ɖ, ɟ, g]	✓	✓
Geminate CC	[bb, d̪d̪, ɖɖ, ɟɟ, gg]	×	×

[1] Proto-Dravidian did not have a laryngeal contrast in voicing. However, many of the modern Dravidian languages have acquired this contrast historically due to the continued contact with Indo-European languages over the centuries that has resulted in lexical borrowings.

Based on these distributional restrictions, gemination and voicing are seen as phonological strategies to avoid singleton voiceless obstruents in intervocalic positions (Namboodiripad & Garellek, 2016). In Government Phonology (KLV 1990), a singleton consonant can either be linked to an onset or to a Rhymal node. While all consonants in the repertoire can be licensed in the onset position, governed by the following nucleus, a rhymal consonant needs to be governed by the following onset. For this, the rhymal consonant must not be more complex than the following onset. While a geminate can be readily analysed as a combination of a rhymal consonant and a following onset, a singleton stop in a word-initial or inter-vocalic context has to be an onset. This is shown in (2).

2. Structural representation of singleton and geminates



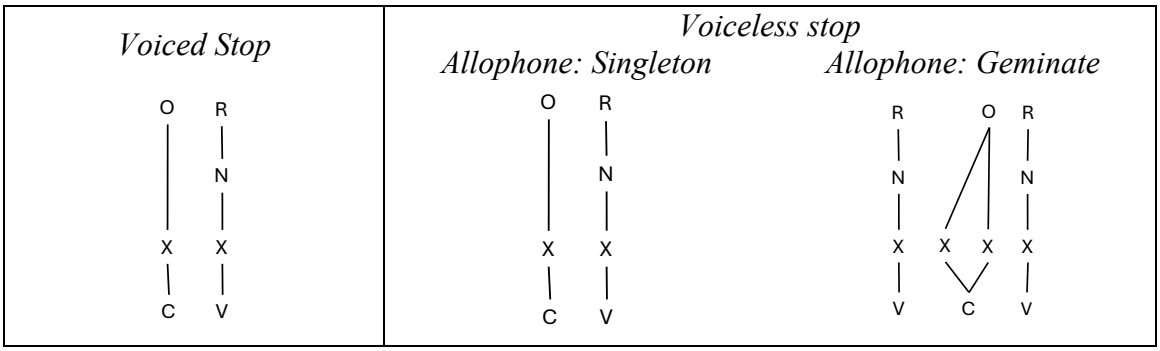
In Malayalam, the geminate stops are not restricted to intervocalic contexts. They can be preceded by laterals or rhotics in medial positions. This is shown in (3).

3. Voiceless geminates and voiced stops preceded by a C in medial position

	preceding lateral		*preceding rhotic*	
Voiced	na**lg**-uga	give.infinitive	a:**rb**a:ɖ-am	luxury
Voiceless	ka**lpp**ikk-uga	order.infinitive	a**rpp**aɳ-am	offering

This shows that the geminate in Malayalam can be formed even when the rhymal position is occupied by another consonant. Based on this we propose that in Malayalam, the geminate can be formed as an onset cluster. In fact, this medial onset cluster geminate is an allophone of the singleton voiceless onset found in word-initial positions. The revised structure of voiced and voiceless stops in Malayalam is shown in (4).

4. Structure of voiced and voiceless stops in Malayalam



The verbal domain of Malayalam shows both singleton and geminate patterns (5).

5. Gemination and voicing in Malayalam verb inflection

	Infinitive	*Past*	
Gemination	no:**kk**-uga	no:**kk**-i	look
Voicing	pa:**ɖ**-uga	pa:**ɖ**-i	sing

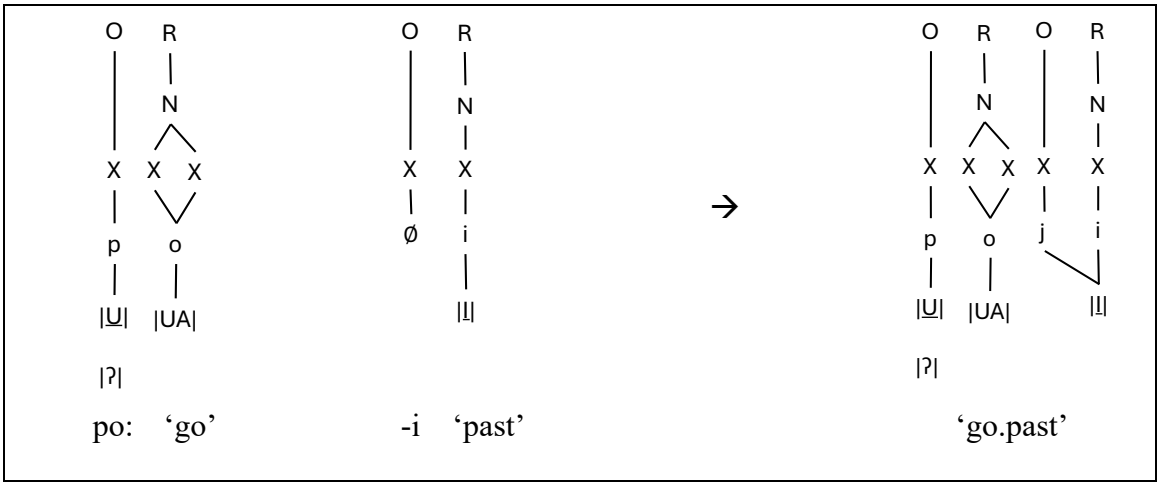
Since both voiced and voiceless stops appear in the same morphologically derived contexts, the difference in laryngeal specification must be part of the underlying phonological representation of the two verb roots in (5). The final C of the verb root in 'look' is specified as an onset cluster, while that of 'sing' is not. In this paper, we denote this structural difference with the sequence $CVCC_0$ and CVC, respectively. Further, we show that the choice of past-tense allomorph [-u] in Malayalam verbs makes a crucial reference to the underlying phonological structure of the verb root. Apart from verbs with the phonological strings $CVCC_0$ and CVC, Malayalam also has CV verbs. The phonological concatenation of these three kinds of verbal stems with suffixal morphemes that begin with V or C is expected to create the phonological strings shown in (6).

6. Phonological concatenation in Malayalam verb roots

Root	*Suffix*	*Resulting phonological string*
$CVCC_0$	-VC	$CVCC_0VC$
	-CV	$CV\mathbf{CC_0C}V$
CVC	-VC	CVCVC
	-CV	CVCCV
CV	-VC	C**VV**C
	-CV	CVCV

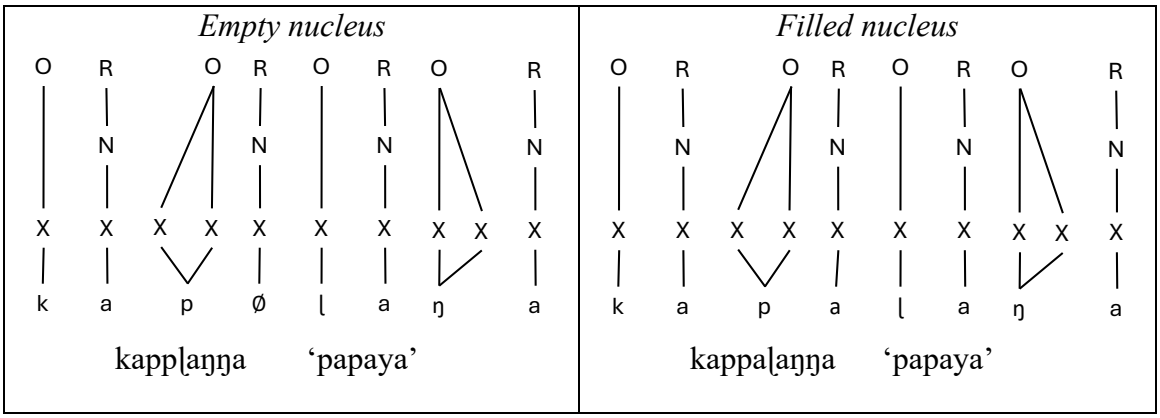
Among these concatenated phonological strings, we expect the CVVC sequence to show morpho-phonological adjustment processes in order to avoid vowel hiatus. The vowel hiatus could either be tolerated, or resolved through vowel deletion, vowel coalescence or consonant insertion. In Malayalam, we see that when a CV verb root like [po:] 'go' concatenates with the past tense suffix [-i], a homorganic glide [j] surfaces between them. This is shown in (7).

7. Realization of homorganic glide in vowel hiatus contexts



Similarly, we expect the sequence CV**CC$_0$C**V to show some phonological adjustment since the suffixal C cannot form a cluster with the geminate. Following strict locality (KLV 1990) an onset can dominate a maximum of two skeletal positions. In Malayalam, both these positions are occupied by the stop in CVCC$_0$ verb stems. If a C were to follow the geminate, it would have to do so with an intervening empty nucleus. In the non-verbal paradigm we can find examples of such CV**CC$_0$C**V sequences with optional intervening vowel. This is shown in (8).

8. Structure of CV**CC$_0$C**V sequences in Malayalam



In the inflectional paradigm of verbs, we do not see this kind of optionality, since all verbal inflections in the language appear to be V-initial suffixes (Asher & Kumari 1997). However, once we look closely at the inflectional domain of Malayalam verbs, one specific inflectional paradigm stands out as a morpho-phonological anomaly that appears to be extremely idiosyncratic. The past tense suffix exhibits a departure from the expected phonological concatenation pattern in two ways. First, the morpho-phonological processes seen at the juncture between the two morphemes in the past tense context is different from the processes seen in case of other v-initial verbal suffixes in Malayalam. This is shown in (9).

9. Contrast between Past Tense and other verbal inflections

Root type	*Past*	*Present*	*Future*	*infinitive*	
$CVCC_0$	viɭicc-u	viɭikkj-uɳɳu	viɭikkj-um	viɭikkj-uga	'call'
CVC	ʈaɳɳ-u	ʈaɾ-uɳɳu	ʈaɾ-um	ʈaɾ-uga	'give'
CV	po:j-i	po:g-uɳɳu	po:g-um	po:g-uga	'go'

All the inflectional markers attaching to the verb roots in (9) are V-initial. Yet, the stem that surfaces with the past tense suffix differs from the present, future and infinitive marked verbs. The right edge of the verbal stem in the past differs from the right edge for the same verb stem in the other morphological contexts. This indicates that the morpho-phonological readjustment at the morpheme juncture in past differs from the rest. For example, the CV root [po:] 'go' is expected to show morpho-phonological adjustment processes to avoid the sequence C**V**VC and does surface with an intervening consonant. But the consonant inserted in the past form [j] is distinct from the one inserted in elsewhere, [g].

Second, the past tense morpheme has two allomorphs, [-i] and [-u], while the other inflections do not exhibit allomorphy. Further, one of these allomorphs [-u] only attaches to C-final stems, $CVCC_0$ and CVC, producing three distinct allomorphic patterns. The data pertaining to this is shown in (10).

10. Attachment of past tense allomorph [-u]

Root Type	*Past*	*Infinitive*	
$CVCC_0$	aɖaɲɲ-u	aɖaj-uga	'close'
	vitt-u	vilkk-uga	'sell'
CVC	ʈoʈʈ-u	ʈoɖ-uga	'touch'

In the first pattern the stem-final consonant is converted to a nasal with the same place of articulation. For example, the palatal /j/ is converted to the palatal nasal /ɲ/ on attaching the past tense [-u]. In the second pattern a stem-final complex [lk] surfaces with the

voiceless alveolar geminate[2] [tt] preceding the morpheme [-u]. In the third pattern, the stem final consonant surfaces as a corresponding geminate preceding [-u] even though the verb root is a CVC. Thus, irrespective of whether the verb root ends in a single C or a CC_0, when attached to the past tense morpheme [-u] the stem ends in a geminate.

Based on the distinct morpho-phonological behavior of the past tense allomorph [-u], we propose that the surface representation for this morpheme is not identical to its lexical specification. Specifically, we argue that this past-tense morpheme is a C-initial suffix where the phonological properties of the suffix-initial C interact with the stem-final C and produce gemination patterns. While the set of verb roots that each of the past tense allomorphs, [-i] and [-Cu], concatenate with is lexically specified and hence unpredictable, the phonological adjustment processes that they undergo post concatenation is completely regular. Thus, we show that the locus of all irregularity in Malayalam past-tense allomorphy is restricted to the vocabulary item insertion rules in the morphological domain, and the domain of phonology is entirely regular.

In the following sections we will reconstruct the underlying phonological forms for the past tense allomorphs and the verb roots based on the phonological concatenation patterns. For this, we will use a combination of the CV Theory (KLV 1990) and Element Theory (Backley 2011). Relevant details of this approach to phonology have been introduced in §2.

2 CV and Element Theory

In this paper, the segmental properties of the phonological string have been analyzed using Element Theory (Backley 2011). All speech segments including consonants and vowels are described using six Elements. These six elements which are divided into three types based on whether they denote place, manner or laryngeal characteristics. These are shown in (11).

11. Three types of Elements

Place of articulation	|I|	|U|	|A|
Manner of articulation	|ʔ|	|H|	
Laryngeal property	|L|		

In this theoretical approach consonantal and vocalic place of articulation is denoted by the same elements denoting place of articulation, |I|, |U| and |A|. Voicing and nasality is denoted by |L| and stop and frication is denoted by the elements |ʔ| and |H] respectively. The element-based configuration of oral and nasal stops has been shown in (12). These will be used for the phonological analysis of sandhi patterns in §3 and §4.

[2] Voiceless alveolar geminates in Malayalam do not have a singleton counterpart. Hence it is restricted to intervocalic environments. However, the geminate is in contrastive distribution with the other voiceless geminate stops including dental and retroflex.

12. Consonantal configuration of elements

	Labial |U|	Dental |I|	Alveolar |A|	Retroflex |A|	Palatal |I|	Velar |U|
Oral stops	p |U?|	t̪ |I?|	t |A?|	ʈ |A?|	c |I?|	k |U?|
Nasal stops	m |U?L|	n̪ |I?L|	n |A?L|	ɳ |A?L|	ɲ |I?L|	ŋ |U?L|

In the table of segmental properties presented in (12), there are pairs of segments that share the same element for the place of articulation. For example, alveolar and retroflex are both marked by the element |A|. These are further distinguished by the property of headedness, marked by a horizontal line below the element. Ordinarily voicing and nasality, both considered to be laryngeal characteristics, is marked by the element |L|. A headed |L̲| marks a voiced obstruent, and without headedness it marked nasality, as in |L|.

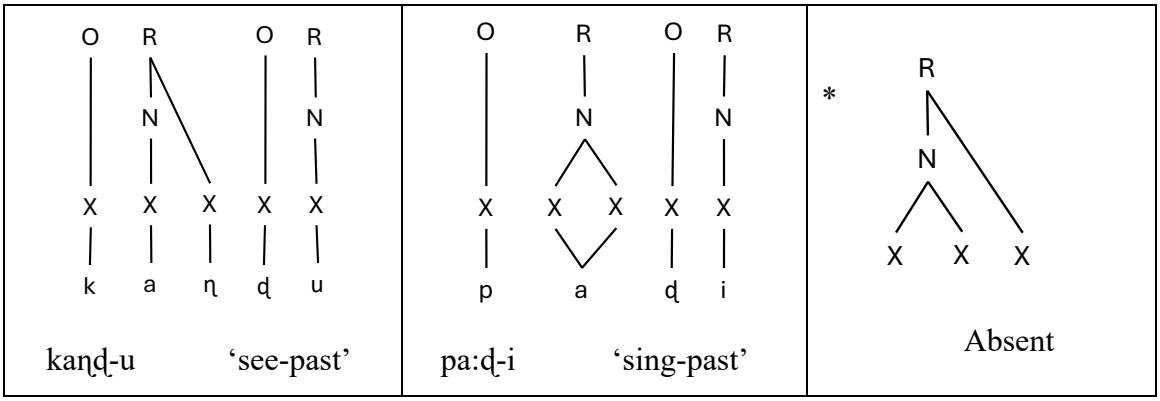
The phonological string is represented by a skeletal tier that comprises of linear sequence of x nodes. The segmental material lexically specified for a morpheme, attaches below these x-nodes. This skeletal tier is further dominated by a sequence of onset O and N nodes. Each x-node dominated by an O must be realized as a consonant and each x-node dominated by an N must surface as a vowel irrespective of the elements it dominates. This is shown in (13).

13. Representational tiers in phonology

O N	O N	O N
| |	| |	| |
x x	x x	x x
| |	| |	| |
|I| |A|	|U| |A|	|A| |A|
ja	wa	ra

Each N node in further dominated by a rhyme node denoted by R. Each of these three structural nodes O, N and R can maximally dominate two x-nodes in a representational structure. This follows from the principle of strict governance, where one of the nodes of a constituent, the governor, must locally govern the other constituent that is dominated by a common mother node (KLV 1990). Thus, a rhyme R might maximally contain either a long vowel, an N dominating two x-nodes, or a rhymal consonant in addition to a short vowel. This is shown in (14).

14. Maximal structure dominated by N and R



In the representational figures that we use to denote sandhi, we will typically show the structure below the skeletal tier unless the higher nodes are crucial to the argument being presented.

3 Morpho-Phonology in Past Tense affixation

In this paper we assume the theoretical position that all concatenative phonology takes place post-syntactically. Specifically, following Chomsky (2001), we assume that the syntactic derivation is shipped out to the phonological module in phases and at each phase the lexicalized phonological material undergoes phonological re-adjustment wherever necessary. As discussed in §1, based on the nature of the morpho-phonological sandhi patterns, we have divided the past tense morpheme into two allomorphs, a C-initial [-Cu] and a V-initial [-i]. In this section, we will discuss further details of each of these.

3.1 The [-Cu] allomorph

There are two patterns of past tense morpho-phonological sandhi on attachment of a [-Cu] suffix to the verb root. In one of these cases the sandhi produces a dental nasal geminate at the juncture of the two morphemes and in another a dental stop geminate.

15. The two patterns of sandhi with the [-Cu] allomorph

Root	*Past*	*Infinitive*	
parakk	para**n̪n̪**-u	para**kk**-uga	'fly'
koɖukk	koɖu**t̪t̪**-u	koɖu**kk**-uga	'give'

The realization of the geminate -kk- sequence with the infinitive in (15), shows that both the verb stems belong to the type $CVCC_0$. Although the phonological environment for the affixation of the past morpheme seems to be identical in both cases, a verb root ending in -kk attaching to [-Cu], the phonological output for them is not the same. In both cases the

velar stop |U?| changes to a dental stop |I?|. Additionally, in some cases, the dental stop is a nasal |I?L|. So, we find that there is a special type of [-Cu] affix, the [-C_Nu], that contains the element [L| in addition to being a velar stop.

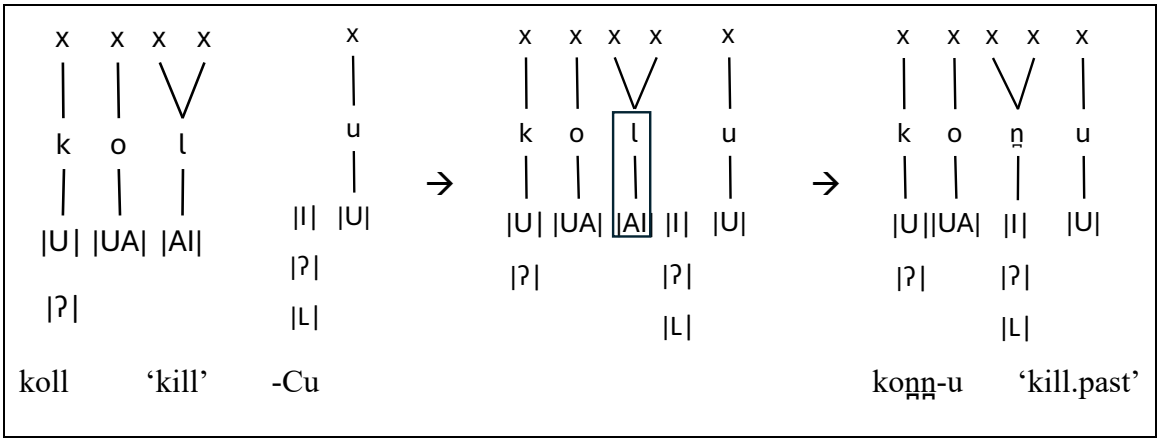
The dental nasal geminate, |I?L|, is not restricted to verb roots ending in the velar stop |U?|. In addition to velar stops it also appears verb roots that end with alveolar laterals and nasals. This is shown in (16).

16. Environment for |I?L|

Root	*Past*	*Infinitive*	
koll	ko**n̪n̪**-u	ko**ll**-uga	'kill'
t̪inn	t̪i**n̪n̪**-u	t̪i**nn**-uga	'eat'
parakk	para**n̪n̪**-u	para**kk**-uga	'fly'

All three verb roots in (16) are of the $CVCC_0$ type where the last consonant of the root is attached to two positions on the skeletal tier. As earlier, this is deduced by examining the infinitive form of the verb where the verb root attaches to the V-initial suffix [-uga]. Irrespective of the place of articulation specified in the underlying representation of the verb root, alveolar |A| or velar |U|, they surface as a dental |I| in the past form. This suggests that the element |I| must have been part of the past suffix. The skeletal positions that were earlier associated with different root specifications are delinked from their original specification and re-linked to the melodic specification of the floating C segment in the suffix initial position. This is schematically shown in (17).

17. Structural representation for dental nasal sandhi



In (17), the past tense suffixal morpheme comes with extra material that is unattached to segmental position. When this suffix concatenates with a verb root, the final melody of the verb root gets delinked in order to accommodate the floating melodic configuration from

the suffix. Just like the final root melody, this floating suffixal melody also gets linked to two adjacent segmental positions at the root edge, forming a geminate dental nasal.

The same past tense allomorph |Iʔʟ|-|U| also attaches to verb roots ending in a singleton rhotic to form geminate dental nasals. This is shown in (18).

18. Geminate |Iʔʟ| from singleton rhotics

Root	*Past*	*Infinitive*	
var	va**n̪n̪**-u	va**r**-uga	‘come’
t̪ar	t̪a**n̪n̪**-u	t̪a**r**-uga	‘give’

As per the element-based classification of consonantal segments in Backley (2011), there are three consonantal articulations that are denoted only by the place feature. These are the palatal glide /j/ corresponding to the element |I|, the labial glide /w/ denoted with |U| and the alveolar rhotic marked as |A|. This was shown in (13) of §2. In Malayalam all these consonantal phonemes that fail geminate. This shows that in order to occupy two segmental positions dominated by an onset, the melodic composition must contain the elements |ʔ| or |H| that are exclusively associated with consonants.

Unlike the voiceless stops, that cannot surface as singletons in medial positions, the nasal consonants do not have any distributional restriction. They can appear as both singleton as well as geminate in medial positions. However, just like the voiceless stops, the dental, palatal and velar nasals only appear as geminates in the word medial position. This is shown in (19).

19. Singleton-geminate contrast in medial nasals

	Singleton		*Geminate*	
|Uʔʟ|			t̪e:ŋŋa	‘coconut’
|U̲ʔʟ|	uma	'name’	umma	‘kiss’
|Iʔʟ|			paɲ̪ɲ̪i	‘pig’
|I̲ʔʟ|			maɲɲə	‘snow’
|Aʔʟ|	a:na	‘elephant’	annam	‘food’
|A̲ʔʟ|	maɳam	‘smell’	maɳɳə	‘soil’

Based on this structural restriction on nasal singletons, we propose that the verbs ending in rhotics surface as geminate due to two independent reasons. First the rhotic |A| cannot geminate, so even if we assume the stem to have the underlying structure of $CVCC_0$ the structural position of the second C will become visible only when the floating consonantal melody is realized. Second, the floating melody of dental nasal stop cannot surface as a

singleton onset, so even if we assume the stem to have the underlying structure of CVC the medial onset with the dental nasal would be realized as a geminate.

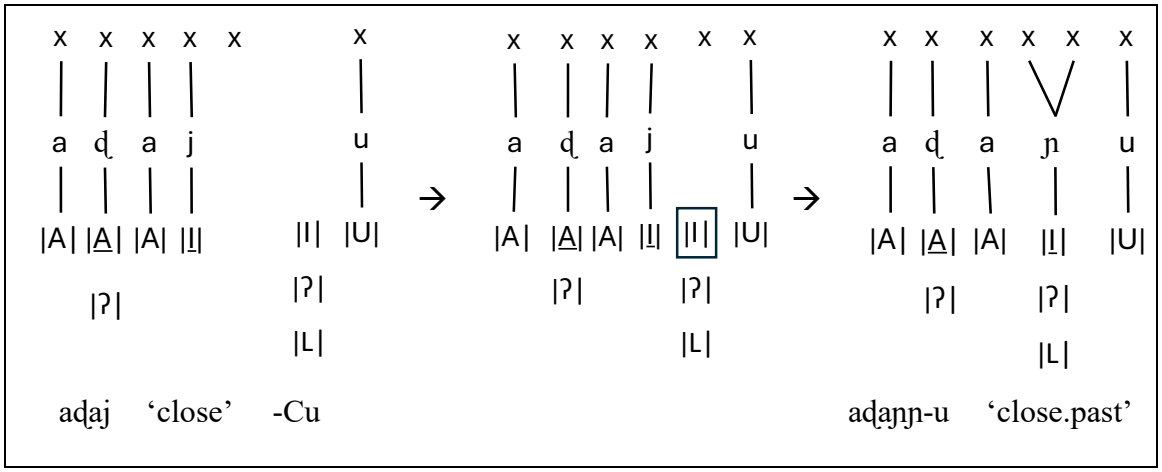
Similarly, a nasal geminate stop also surfaces when the verb root ends in a singleton palatal glide. However, unlike in the case of the rhotic, the palatality of the stem-final melody is preserved in the nasal stop that emerges with the past tense morpheme. This is shown in (20).

20. Geminate palatal nasal stop in past tense

Root	*Past*	*Infinitive*	
aḍaj	aḍa**ɲɲ**-u	aḍa**j**-uga	'close'
karaj	kara**ɲɲ**-u	kara**j**-uga	'cry'

As discussed in (12) of §2, the dental and palatal consonants share the same element |I|. While palatal has a headed |I|, the dental has a non-headed |I|. In a context where the root final consonantal melody shares the same place of articulation with the suffix initial consonantal melody, the former does not delink. Instead, the more intense realization of |I| as a headed |I| in a palatal gets coalesced with the manner and laryngeal specifications of the floating melody |I?L|. This is shown in (21).

21. Structural representation for palatal nasal sandhi



Although Malayalam has six different places of nasal articulation in the language that corresponds to the oral stops, it is crucial to note that only two of these, the dental and palatal surface with the past tense allomorph [-C_N u]. This confirms that the suffix is specified for the melody |I?L| with the element |I| that gets realized as a geminate with $CVCC_0$ type roots.

The [-Cu] allomorph with the melody |I?| shows a similar pattern where it surfaces as a geminate palatal stop when the verb root ends in a palatal consonant, and as a dental elsewhere. This is shown in (22).

22. Geminate |I?| in past tense morphology

Root	*Past*	*Infinitive*	
aḍakkj	aḍa**cc**-u	aḍa**kkj**-uga	‘close’
aḍikkj	aḍi**cc**-u	aḍi**kkj**-uga	‘beat’
koḍukk	koḍu**ṯṯ**-u	koḍu**kk**-uga	‘give’
eḍukk	eḍu**ṯṯ**-u	eḍu**kk**-uga	‘take’

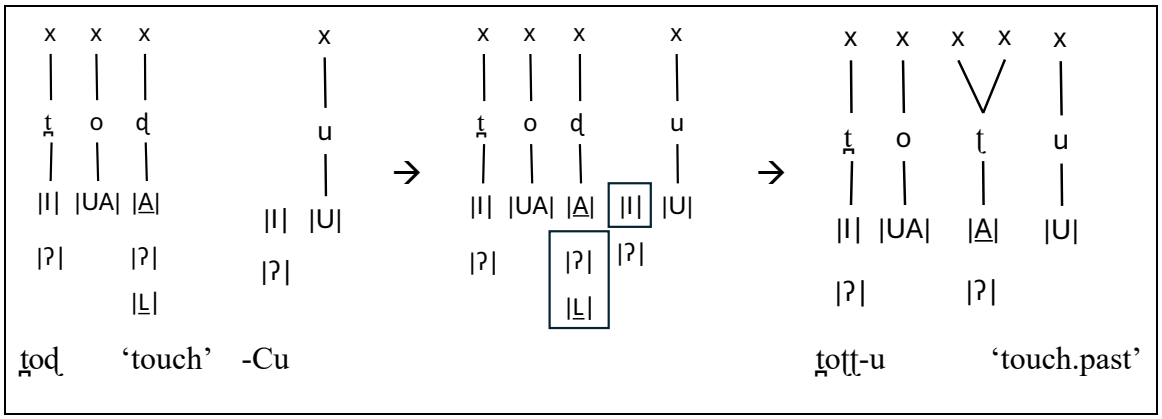
In summary, so far, we have seen two types of [-Cu] past tense suffix in Malayalam. In both cases, the suffix initial consonantal melody |I?L| and |I?| is a floating specification that attaches to the right edge of $CVCC_0$ type verb roots. In a third pattern of [-Cu] suffixation, a CVC type root ending in retroflex stops, |A?|, produces past tense forms with voiceless retroflex geminates. This is shown in (23).

23. Gemination with CVC type roots

Root	*Past*	*Infinitive*	
ṯoḍ	ṯo**ṯṯ**-u	ṯo**ḍ**-uga	‘touch’
viḍ	vi**ṯṯ**-u	vi**ḍ**-uga	‘release’

The pattern of sandhi between a stem-final retroflex, |A?L|, and the [-Cu] allomorph with the melody |I?| is similar to the pattern seen with stem-final rhotics and palatal glides. The singleton onset melody that consists a headed place |A| and voicing |L| is well-formed as an onset. Unless it loses some of its specification, it cannot accommodate any of the melodic elements from the floating C specification. In such a scenario, the stem-final C loses all specifications except place and copies the remaining melody from the floating C. This is shown in (24).

24. Structural representation of sandhi with retroflex



Thus far we have assumed that the suffix-initial C in the [-Cu] past tense allomorph is a floating consonantal specification without an anchoring in the skeletal tier. In an alternative hypothesis we can also consider the floating specifications of |ɪʔL| and |ɪʔ| to be complete segments with an x-node anchor in the skeletal tier. However, in that case we would have to make an additional stipulation for the deletion of one of the x-nodes in the skeletal tier when the [-Cu] attaches to $CVCC_0$ type verb roots. Considering that the [-Cu] allomorph attaches only to C-final verb roots, mostly $CVCC_0$ type, we prefer to analyze it as an un-anchored phonological specification that tries to parasitically express itself on the stem-final consonantal nodes. The geminate outputs are a consequence of either the original skeletal representation of the $CVCC_0$ verb roots, in cases like [t̪inn →t̪in̪n̪-u 'ate'], or a consequence of the restriction on singleton voiceless stops and certain nasal stops in medial positions. A summary of the analysis presented in this section is presented in (25).

25. Summary of [-Cu] affixation

- [-Cu] allomorphs only attach to C-final verb roots
- Based on the melodic specification, they are of two types.
    - Floating melody |ɪʔL|
    - Floating melody |ɪʔ|
- In all cases of concatenation, we see some aspect of the floating melody being realized in coalescence with the melodic specification of the stem final C.
- If the stem final C has a non-headed place feature, then the place feature of the floating melody is expressed in the ensuing geminate.
- However, if the stem final C has a headed place feature, palatal or retroflex, then the place feature of the stem final C is expressed in the ensuing geminate.

Cross-linguistically, we have seen many cases of parasitic exponence where a segment gets expressed only if the preceding or following morpheme meets a certain structural condition. A classic case is that of French liaison where the stem final nasal gets expressed only of the following word is vowel initial like [mõn ami 'my friend']. In Malayalam, the verb roots to which the [-Cu] allomorphs can attach are lexically specified, but all of these have the common characteristic of being C-final. Just like the French nasal that must attach to the empty onset of a following word to be expressed, the Malayalam floating melody must attach to a preceding skeletal position that is dominated by an onset.

3.2 The [-i] allomorph

In contrast to the [-Cu] allomorph that has a phonological dependence on the root type to which it attaches, the [-i] allomorph does not interact with the phonological shape of the verb root at all. Consequently, it attaches to all three types of verb roots. This is shown in (26).

26. Past tense marking with [-i]

Root Type	*Root*	*Past*	*Infinitive*	
$CVCC_0$	uraŋŋ	uraŋŋ-i	uraŋŋ-uga	sleep
	no:kk	no:kk-i	no:kk-uga	look
	poŋŋ	poŋŋ-i	poŋŋ-uga	levitate
	uʈakk	uʈakk-i	uʈakk-uga	stick
	karaŋŋ	karaŋŋ-i	karaŋŋ-uga	rotate
	t̪u:ŋŋ	t̪u:ŋŋ-i	t̪u:ŋŋ-uga	hang
CVC	pa:ɖ	pa:ɖ-i	pa:ɖ-uga	sing
	ka:r	ka:r-i	ka:r-uga	scream
	u:r	u:r-i	u:r-uga	remove
	mu:ɭ	mu:ɭ-i	mu:ɭ-uga	hum
CV	ku:	ku:-w-i	ku:-v-uga	howl
	po:	po:-j-i	po:-v-uga	go

The [-i] allomorph behaves exactly like all the other verbal inflections to produce regular predictable concatenation patterns.

4 Discussion

In this paper we have seen that Malayalam has certain consonants that must appear as onset clusters in the medial position. These include all six voiceless stops as well as the dental, palatal and velar nasals. The melodic configuration of each of these is shown in (27).

27. Melodic configuration of onset cluster geminates

	Labial	*Dental*	*Alveolar*	*Retroflex*	*Palatal*	*Velar*
Oral stops	p |U̲ʔ|	t̪ |Iʔ|	t |Aʔ|	ʈ |A̲ʔ|	c |I̲ʔ|	k |Uʔ|
Nasal stops		n̪ |IʔL|			ɲ |I̲ʔL|	ŋ |UʔL|

While the past tense allomorph [-Cu] has a melodic specification |I|, many of the $CVCC_0$ verb types ending in stops have a melodic specification |U|. The latter can be seen in (28).

28. Verb stems ending in [kk] or [ŋŋ]

With [-Cu] past tense allomorph			*With [-i] past tense allomorph*		
Past	*Infinitive*		*Past*	*Infinitive*	
para-**n̪n̪**-u	para-**kk**-uga	fly	no:kk-i	no:-kk-uga	look
koɖu-**t̪t̪**-u	koɖu-**kk**-uga	give	uɖa-kk-i	uɖa-kk-uga	stick
eɖu-**t̪t̪**-u	eɖu-**kk**-uga	take	ura-kk-i	ura-kk-uga	sleep [caus]
maɖu-**t̪t̪**-u	maɖu-**kk**-uga	fed up	po-kk-i	po-kk-uga	lift
t̪o:**tt**-u	t̪o:l-**kk**-uga	fail	kara-kk-i	kara-kk-uga	rotate [caus]
vi**tt**-u	vil-**kk**-uga	sell	t̪u:-kk-i	t̪u:-kk-uga	hang [caus]
aɖa-**cc**-u	aɖa-**kkj**-uga	close [tr]	ura-ŋŋ-i	ura-ŋŋ-uga	sleep
paɖi-**cc**-u	paɖi-**kkj**-uga	'study'	po-ŋŋ-i	po-ŋŋ-uga	levitate
maɖi-**cc**-u	maʈi-**kkj**-uga	'lazy'	kara-ŋŋ-i	kara-ŋŋ-uga	rotate
kaɖi-**cc**u	kaɖi-**kkj**-uga	bite	t̪u:-ŋŋ-i	t̪u:-ŋŋ-uga	hang
Systematic phonological re-adjustment			No phonological re-adjustment		

While stems ending in [kk] can occur with either of the past tense allomorphs, the velar nasal [ŋŋ] appears only in stems that attach with the [-i] allomorph. Similarly, the stems ending in labial or dental stops [pp] or [t̪t̪] invariably select the [-i] allomorph. Some examples are shown in (29).

29. Verb stems ending in [pp] or [t̪t̪]

Past	*Infinitive*	
capp-i	capp-uga	'lick'
ku:pp-i	ku:pp-uga	'fold'
kut̪t̪-i	kut̪t̪-uga	'stab'

Given our hypothesis, had the $CVCC_0$ verb stems ending in [pp] concatenated with the [-Cu] allomorph, the headed |U| of the labial would have maintained its place in the surface representation just like the headed |I| and |A| or the palatal and retroflex. Consequently, no melodic element from the floating C would have found surface realization. This situation is avoided in the language through allomorph selection.

Unlike [pp] the onset [t̪t̪] does not have a headed characteristic. However, the melodic representation of the stem final C exactly matches with the melodic configuration of the floating C. In this situation, it would be impossible to distinguish between verbs that ended

with other consonants like [kk] and those that end with [ṯṯ] in their past tense forms. Such a lexical merger is avoided through allomorph selection.

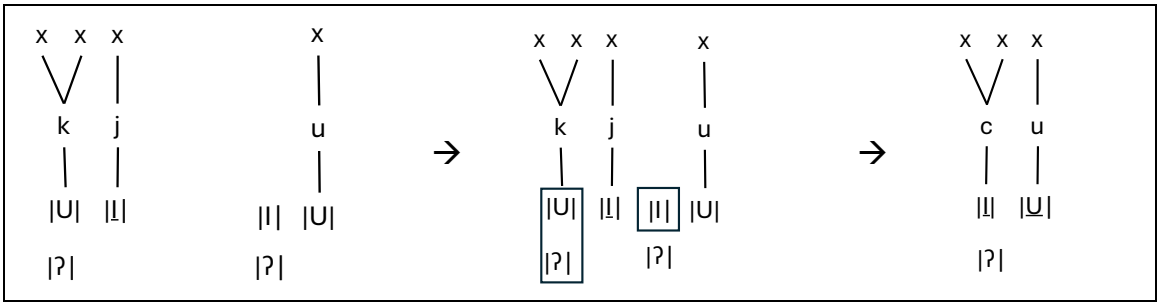
Finally, we observe that there is no lexical selection in past tense allomorphy after the addition of the causative morpheme. Irrespective of whether the non-causative verb form selected for a [-Cu] or a [-i] past tense morpheme, every causative-past is formed by attaching the [-Cu] allomorph. This is shown in (30).

30. Causative-past

	Past	*Causative Past* [-C_0 u]		
	ko-ṉṉu		koll-i-ccu	kill
	para-ṉṉu	para-ppi-ccu	para-ṯṯi / para-ṯṯi-ccu	fly
	va-ṉṉ-u		var-u-ṯṯi/ varu-ṯṯi-ccu	come
[-C_N u]	kara-ɲɲu	karaji-ppi-ccu	karaji-ccu	cry
	ali-ɲɲu	aliji-ppi-ccu	aliji-ccu	melt
	aɖa-ɲɲu		aɖa-ji-ccu	close [unacc]
	koɖu-ṯṯu	koɖu-ppi-ccu		give
	eɖu-ṯṯu	eɖu-ppi-ccu		take
	maɖu-ṯṯu	maɖu-ppi-ccu		fed up
	ṯo:-ttu	ṯo:l-ppi-ccu		fail
	vi-ttu	vil-ppi-ccu/vi-ppi-ccu		sell
[-C_0 u]	aɖa-ccu	aɖa-ppi-cc-u		close [transitive]
	paɖi-ccu	paɖi-ppi-ccu		‘study’
	maɖi-ccu	maɖi-ppi-ccu		‘lazy’
	kaɖi-ccu	kaɖi-ppi-ccu		bite
	ṯoṯ-u		ṯoɖ-i:-ccu	touch
	viṯ-u		viɖ-i:-ccu	release
	no:kk-i	no:kk-i-ppi-ccu	no:kk-i-ccu	look
	uɖakk-i	uɖakk-i-ppi-ccu	uɖakk-i-ccu	stick
	mukk-i	mukk-i-ppi-ccu	mukk-i-ccu	dip [transitive]
	ɲekk-i	ɲekk-i-ppi-ccu	ɲekk-i-ccu	press
	capp-i	capp-i-ppi-ccu	capp-i-ccu	‘lick’
[-i]	ku:pp-i		ku:pp-i-ccu	‘gesture’
	kuṯṯ-i		kuṯṯ-i-ccu	‘stab’
	pa:ɖ-i	pa:ɖ-i-ppi-ccu	pa:ɖ-i-ccu	sing
	ka:r-i	ka:r-i-ppi-ccu	ka:r-i-ccu	scream
	u:ɾ-i	u:ɾ-i-ppi-ccu	u:ɾ-i-ccu	remove
	ku:-v-i	ku:v-i-ppi-ccu	ku:-v-i-ccu	howl

The causative infinitive for all these verbs ends with [-kkj-uga]. This [-kkj-] part phonologically interacts with the [-Cu] past tense allomorph to produce [-ccu] in the causative-past morphology. This is shown in (31).

31. Formation of the causative past morpheme [-ccu]



Since the causative morpheme ends in a palatal glide |I|, the past tense allomorph with the palatal vowel [-i] is systematically avoided in these causative contexts. Malayalam also has a handful of intransitive verbs ending in geminate velar nasals, that form their causative counterpart with velar stops. Since, there is no palatal context in these cases, all these verbs both the intransitive as well as the causative form concatenate with the allomorph [-i] to form the past tense form. This is shown in (32).

32. Selection of [-i] allomorph in non-palatal contexts

Intransitive				*Causative*			
Root	Past	Infinitive		Root	Past	Infinitive	
uraŋŋ	uraŋŋ-i	uraŋŋ-uga	sleep	urakk	urakk-i	urakk-uga	sleep
poŋŋ	poŋŋ-i	poŋŋ-uga	levitate	pokk	pokk-i	pokk-uga	lift
karaŋŋ	karaŋŋ-i	karaŋŋ-uga	rotate	karakk	karakk-i	karakk-uga	rotate
t̪u:ŋŋ	t̪u:ŋŋ-i	t̪u:ŋŋ-uga	hang	t̪u:kk	t̪u:kk-i	t̪u:kk-uga	hang

5 Conclusion

In this paper we have proposed that the past tense morpheme in Malayalam has three allomorphs, [-C_N u], [-C_0 u], and [-i]. Both the [-Cu] allomorphs have a floating dental stop melody that is not anchored to a specific position on the skeletal tier. However, this floating melody needs to be expressed in some manner in the geminate onset cluster preceding the suffixal vowel. For this, these allomorphs attach only to C-final stems. While the allomorph [-C_0 u] can attach to both CVC and $CVCC_0$ stems, the [-C_N u] allomorph only concatenates with $CVCC_0$ stems. In phonological environments where the floating melody would not be

clearly discernable in the melodic properties of the geminate onset, such as stems ending in [pp] or [t̪t̪], the [-Cu] past tense marker fails to attach.

Unlike the [-Cu] allomorphs, the [-i] suffix does not phonologically interact with the verbal stem to which it attaches. Therefore, it can easily attach to all the three types of verbal stems: $CVCC_0$, CVC and CV. However, the [-i] allomorph fails to attach in the phonological environment when the stem final segment is a palatal. We show that this is the reason it fails to appear after the causative morpheme [-kkj-]. These distributional restrictions are summarized in (33) below.

33. Summary of the past-tense allomorphs in Malayalam

	[-C_N u]		[-C_0 u]		[-i]	
Root $CVCC_0$	para-kk	✓	koɖu-kk	✓	ɲekk	✓
Past	para-n̪n̪u	fly	koɖu-t̪t̪u	give	ɲekk-i	press
Root CVC	×		t̪oɖ	✓	pa:ɖ	✓
Past			t̪o-ʈu	touch	pa:ɖ-i	sing
Root CV	×		×		ku:	✓
Past					ku:-v-i	howl

References

Asher, R. E., & Kumari, T. 1997. *Malayalam*. Routledge.

Backley, P., 2011. *Introduction to element theory*. Edinburgh University Press.

Chomsky, N., 2001. Derivation by phase. *Ken Hale: A life in language/MIT*.

Kaye, J., Lowenstamm, J. and Vergnaud, J.R., 1990. Constituent structure and government in phonology. *Phonology*, *7*(1).

Mohanan, K. P. 1986. *The Theory of Lexical Phonology*. D. Reidel Publishing Company.

Namboodiripad, S. and Garellek, M., 2017. Malayalam (Namboodiri dialect). *Journal of the International Phonetic Association*, *47*(1).

Mirativity on the Table

VEDA KANAMARLAPUDI, *Stanford University*
AHMAD JABBAR, *Stanford University*

ABSTRACT

Using the Hindi-Urdu discourse particle *lo* as a case study, the present paper models mirativity within the Table model framework (Farkas & Roelofsen, 2017). This comes with an enrichment of the Table model; we propose the discourse structure to encode, in addition to discourse commitments, the time these commitments are publicized. We also incorporate a component that contains the public record of private beliefs. With these two additions, we seek to capture mirativity in the Table model, emphasizing recency as a condition for surprise and underlining the need and the conceptual difficulty of incorporating private beliefs in a public scoreboard of conversation. We show that the difficulty can be overcome to fulfill the modeling need.

1 Introduction

In this paper, we focus on *lo*, a discourse particle in Hindi-Urdu.[1] To our knowledge, *lo* hasn't been explored in the literature and hasn't even featured orthogonally in examples. We **contribute empirically** to the literature by showing that *lo* sports a rich and complex empirical profile, giving rise to a panoply of semantic puzzles, which we seek to resolve with a simple semantic generalization as in (1). We **contribute theoretically** in two ways; first, we argue that mirativity can be captured within the Table model framework (Farkas & Bruce, 2010; Farkas & Roelofsen, 2017);[2] and second, we argue that this capture requires reference to time-stamps and private belief states. Given that the components within the Table model represent the public record of the conversation, we seek to reconcile the privacy of speaker beliefs with the publicity of what the Table model seeks to capture. We introduce **Doxa**, a new Table component, and **time-stamps** to be included with each member of discourse commitments and Doxa. The reader can skip to Section 3 for the extensions.

(1) *lo* is felicitous to use at t_1, if the speaker publicly commits to p at t_1, while not having moderate to high credence towards p at t_0, where p is the content of *lo's* containing clause and $t_0 < t_1$.

We briefly foreshadow our modeling contribution for the reader's expectation of what's to be found in the coming pages in terms of theory-building. Like Gunlogson (2008), Farkas

[1]Authors contributed equally to the paper. Most special thanks to Cleo Condoravdi for her extensive feedback on this project! Thanks also to Donka Farkas, David Beaver, Miriam Butt, Rajesh Bhatt, Christine Bartels, Ashwini Deo, Seo-young Lee, Xuetong Yuan, Katherine Johnson, and audiences and participants at the Discourse Dynamics workshop at *FASAL 15* and the *Stanford Construction of Meaning* workshop! Thanks to Shrija Rawat, Azeez Hamoodur Rahman, and Mydhili Rachapudi for elaborate discussion on judgments!

[2]This point has also been made in Kraus (2019), with which we engage in Section 5.

& Bruce (2010) propose to include within the informational scoreboard (Lewis, 1979), publicized commitments of each discourse participant, in addition to what's common ground between all. Farkas & Roelofsen (2017) extend the framework in Farkas & Bruce (2010), *inter alia*, by incorporating biases in addition to full commitments, with an eye towards modeling rising declaratives and tag interrogatives.[3] One is biased towards and committed to propositions. Then, naturally, biases are represented as a set of pairs containing propositions and measures of bias, where the measure within each pair captures the extent of the bias towards the proposition in the pair. Due to wide empirical coverage, while providing a comprehensive multi-dimensional view of interpretation, Farkas & Roelofsen (2017) has emerged as a contender for a complete theory of interpretation, with an elegant division of labor between semantic composition and discourse effects.[4] To further bolster the framework's status, we argue that, with a simple enrichment, the framework can capture mirativity as well (DeLancey, 1997, 2001; Aikhenvald, 2012).[5]

What is our theoretical proposal? For each proposition p in a participant's commitment set, we propose to record the time of the participant's commitment to p. We call this time *time-stamp of commitment*, or more simply *time-stamp*. These time-stamps can slice time as thinly as one wishes; propositions might be time-stamped for the day they entered the commitment set or for the entering hour or the minute. We only make the cognitively plausible assumption that time-stamps be identified with discretized units of continuous times. We also propose the time-stamps to be recorded for biases as well. With both commitments and biases publicized at particular times, and with a public record of those times, we formalize (1) as the following: the speaker had previously not (privately) assigned moderate to high credence to the content of *lo*'s containing clause to which the speaker commits, in uttering the clause. An illustrative example will help. Suppose the speaker utters *lo S*. $[\![lo\ S]\!]$ is: the speaker had not (privately) assigned moderate to high credence towards $[\![S]\!]$ before committing to $[\![S]\!]$.[6] We also show that the optionality of privately assigning credences can be made to function coherently within the Table model, which is essentially a public record of the conversation. For this, we introduce a new component Doxa that is a public record of private beliefs. Time-stamps are extended to Doxa. Throughout the discussion, we bring Hindi-Urdu-internal and cross-linguistic evidence from English and Romance languages to motivate the extensions and understand *lo*'s contribution.

Outline of the paper. Section 2 sketches the empirical problem, while verifying (1) with each example. Section 3 proposes the extensions; sections 3.2.1 and 3.2.3 contain the time-stamps and Doxa extensions, respectively. We propose the Table-borne semantics for

[3]While Farkas & Roelofsen (2017) tie their contribution to the Inquisitive framework (Ciardelli et al., 2018), we abstract away from the Inquisitive module.

[4]The division of labor is inspired by Lauer & Condoravdi (2012) and Condoravdi & Lauer (2012) who use commitment to the content of a clause as a means for defining varying conventions of use across clause-types.

[5]Kraus (2019) attempts the same, but we seek to model mirativity more simply, while underlining the need for a reference to times and speaker private beliefs.

[6]We use $[\![\ ]\!]$ as is standard in formal semantics (Heim & Kratzer, 1998), i.e., as a function that maps expressions to their interpretations.

lo in Section 4, with some short final sections containing some reflective discussion, a comparison with Kraus (2019), and our conclusion.

2 The empirical problem

Methods. The dataset presented comprises (i) dialogues from *Dhoop Kinare* (Moin, 1987), a popular play broadcast in 1987 (with frequent reruns since) on Pakistan Television; (ii) synthetic examples with introspective judgments;[7] (iii) examples taken from the Wortschatz Leipzig Hindi Corpus of 100k sentences (Goldhahn et al., 2012; Leipzig Corpora Collection, 2019).

Design. We follow the usual strategy of constructing minimal pairs, where each pair contains examples that are identical, save for the modulation of the variable whose effect we want to measure on *lo*'s felicity. The dataset is thus structured, helping us isolate the variables that we seek to unite in an interpretable and compressed semantic generalization.

Data Distribution. Most of the attested sentences across *Dhoop Kinare* and the corpus in which *lo* occurs are declaratives. We use this statistic to motivate our focus on declaratives in sketching *lo*'s empirical profile, while also discussing interrogatives and imperatives later.

2.1 Declaratives

In (1), we presented a semantic generalization for *lo*. We motivate and verify this generalization with each context and minimal pair presented. We start with a scene from *Dhoop Kinare* (Moin, 1987, Episode 6).
DIRTY HANDS: The dad has been painting the walls distractedly, while talking to his daughter. Suddenly, the dad realizes that his hands have become all dirty from the paint.

(2) Dad: Lo, saare haath kharab ho gaye.
Dad: LO*, my hands have gotten all dirty.*

While in (2), given DIRTY HANDS , *lo* may be felicitous due to a sudden discovery, it may also just be compatible with declaratives *simpliciter*.[8] To motivate sudden discovery as a necessary condition for *lo*'s felicity in declaratives (and more broadly later), we tweak (2) a little, where the dad, instead of saying (2) upon realization, expects the hands to get dirty. This expectation is common ground before *lo* is used. Consider (3) below.

[7] We choose *Dhoop Kinare*, as it depicts the lives of young doctors in urban Karachi, with its dialogue being reflective of the Hindi/Urdu spoken across major metropolises in North India and the Punjab and Sindh provinces in Pakistan. Curiously, *Dhoop Kinare* contains many occurrences of *lo*, which many other plays we checked, even by the same playwright, couldn't match in frequency. All introspective judgments were thoroughly corroborated with native speakers of Hindi-Urdu: the authors and six other speakers from Lahore, Punjab in Pakistan and the Indian regions of Uttar Pradesh, Uttarakhand, and Hyderabad, Telangana.

[8] For instance, *lo* might be ok in declaratives by virtue of being a Force head of the sort that Bhatt & Dayal (2020) take *kya* to be. *Kya*, as per Bhatt & Dayal (2020), semantically selects for a set of propositions, but *lo*'s selection can be limited to propositions, under a framework, unlike Inquisitive Semantics (Ciardelli et al., 2018), with the type distinction between the contents of declaratives and interrogatives.

(3) Dad (before starting painting): *could you get me a towel. I always make a mess when painting*.
Daughter: *Your hands are always a sight to see*.
INTERLUDE: THE DAUGHTER GOES TO GET THE TOWEL, COMES BACK AFTER 5 MINUTES, WHILE THE DAD PAINTS DISTRACTEDLY. THE DAD LOOKS AT HIS HANDS THAT HAVE BECOME ALL DIRTY FROM THE 5 MINUTES OF PAINTING.
Dad: # Lo, saare haath kharab ho gae.
Dad: # LO, *my hands have gotten all dirty.*

Through the dialogue in (3), we seek to show that when, at time t_0, the expectation for an event is common ground, *lo*'s use, at t_1, in a speech act that proposes to make the event's occurrence common ground is not felicitous, where t_0 immediately precedes t_1. We further note that the addition of *already* makes the use of *lo* in (3) better as in (4), that may signal that the dad wasn't expecting the hands to get so dirty so quickly.

(4) Lo, abhi se hi saare haath kharab ho gae.
LO, *my hands have gotten all dirty already*.

With the examples above, we motivate a necessary condition for *lo* that the speaker commit to the asserted proposition that the speaker didn't expect to be true at the time right before. Now, we seek to assess if signaling commitment at the time of utterance is the right attitude towards the content, or if bias should suffice too.
RAIN 1: Kunal is doing his homework in the library. He decides to take a break and walk around the room. He goes to the window, and he **sees that it is raining**. Kunal says (5):

(5) Lo, bahar baarish ho rahi hai.
LO, *it's raining outside.*

To construct a minimal pair, targeting the commitment variable, contrast RAIN 1 with RAIN 2 below.
RAIN 2: Kunal is doing his homework in the library, when he decides to take a break and walk around the room. He goes to the window, and sees that it is **cloudy outside with condensation on the windowpane**.

(6) # Lo, bahar shayad baarish ho rahi hai.
LO, *it might be raining outside.*

In (2), (4), and (5), the speaker fully commits to the content of the *lo*-marked clause. In (6), there is an explicit uncertainty marker *shayad* within the *lo*-marked clause that we roughly translate in English with the epistemic modal *might*.[9] The minimal pair with (5) and (6) helps us corroborate our observation that *lo* is bad with uncertainty. In other words, *lo* requires commitment to the content of its containing clause at the time of utterance.

The above contexts, DIRTY HANDS and RAIN 1 & 2, along with their minimal pairs, suggest that commitment to p, the content of *lo*'s containing clause, at utterance time t_1, is

[9]See Jabbar (2023) for more on *shayad*, where a similar argument is built for the Hindi-Urdu discourse particle *na*.

a necessary condition for *lo*'s felicity. There's a similar question that can be raised for the immediately preceding time: at the immediately preceding time t_0, ought the speaker not be committed to p or ought the speaker not be biased towards p? Suspecting that it may be raining at t_0 is compatible with having no commitment to the proposition that it is raining at t_0. Clarifying the exact attitude whose absence *lo* requires at t_0 is thus important if we indeed see an effect of modulation of the attitude towards p at t_0 on the felicity of *lo*-p. (3), given DIRTY HANDS, is set up such that the dad has prior expectation of his hands getting dirty. This renders the use of *lo* infelicitous. Then, for the felicitous use of *lo*-p, in addition to being committed to p at t_1 (utterance time), the speaker ought not be committed to and ought not be biased towards p at t_0 (the immediately preceding time). This lack of bias and commitment towards p is exactly what makes the expression of surprise felicitous. *Lo*'s fitting this data profile can be taken as further evidence that *lo* is a mirative.

Extent of surprise. With any expression of surprise, the extent of the violation of prior expectation can be varied, which in turn can determine the felicity of the expression. For instance, in a rainy city like Manchester, if it drizzles when the forecast was for clear skies, we don't expect a life-long Mancunian to express this violation of expectation with *whoa, it is raining*. Change the scenario to a resident of the Sibi District in Pakistan; seeing rain when the forecast was for clear skies makes the use of *whoa, it is raining* quite apt.[10] Now, of course, in each case, the underlying expectation is determined through past experiences, in addition to the weather forecast. While in both cases, a prior expectation is violated, the use of *whoa* is more natural in the second instance.

What we seek to illustrate through our detour into *whoa* is that while *whoa* is an expression of surprise, the extent of the violation of prior expectation determines *whoa*'s felicity. Then, for the completeness of the semantic analysis for any mirative expression, it is important to understand if there are any conditions on the extent of surprise that warrants expression through the mirative. We set up a minimal triple below specifically to understand if there are any such conditions due to the extent of violation or the extent of surprise on *lo*'s felicity. HOMECOMING 3 is used as a control to show that in the given context, you need some expectation-violation. HOMECOMING 1 and HOMECOMING 2 seek to show that *lo* allows for a wide extent of expectation-violation.
HOMECOMING 1: A and B are roommates. A expects B to come back from work at 6pm, like usual, everyday. B comes home at 4pm today.

(7) A: Lo, tum toh 4 baje aa gaye aaj.
LO*, you're home at 4pm today*.

HOMECOMING 2: A expects B to come back from work at 4pm, but B emerges from his room at 4pm, having slept all day.

(8) A: Lo, tum toh kaam par gaye hi nahi.
LO*, you didn't even go to work today*!

[10]Sibi is considered the hotspot of Pakistan, receiving little rainfall around the year (Wikimedia Foundation, 2025).

HOMECOMING 3 (Control): A and B are roommates. A expects B to come back from work at 6pm. It is 5:55pm now. B texts A that B is nearby and will arrive in 5 minutes. B arrives after 5 minutes.

(9) A: # Lo, tum aa gae.
LO, *you're back*.

In both (7) and (8), the speaker expectation is violated. In (7), the speaker doesn't expect B to arrive early, i.e., at 4pm, and in (8), the speaker doesn't expect B to have not gone to work at all. However, arriving a few hours earlier is certainly less surprising than not having gone to work. It is felicitous to use *lo* in both situations. So, tentatively, we can say that *lo* allows for a wide range of surprise.

Now, we stop for a moment to consider two objections. The first objection goes something like this: it is not clear that the expectation-violation in HOMECOMING 1 is less severe than that in HOMECOMING 2.[11] Without a way to reliably construct contexts that vary in surprise, we cannot use the difference in felicity of *lo*, or lack thereof, across contexts to isolate a point about the range of surprise that *lo* is felicitous to use with. A different but related objection may concede that the expectation-violation in HOMECOMING 1 is less severe, but may point out that we don't exhaust the range of surprise by the two contexts, HOMECOMING 1 and HOMECOMING 2. It would then appear that the generalization that *lo* is felicitous to use in any expectation-violating context is not warranted. We consider both objections in turn.

First, to steelman the first objection, let's concede that the expectation-violation in HOMECOMING 1 is as severe as the violation in HOMECOMING 2. Even with this concession, we note that *lo* is felicitous in both the Manchester rain and the Sibi rain scenarios. This empirical fact suggests that *lo* at least accepts a wider range of expectation-violation than *whoa*. And note that here, we are not stipulating that the Manchester rain context is less surprising than Sibi rain—that would be liable to the same objection to which we are replying. Instead, we're using the variation in *whoa*'s felicity across the two contexts to take the two contexts to comprise different levels of expectation-violation. Then, we make the comparative point that *lo* allows for a wider range of surprise than *whoa*, not for its own sake, but to show that *lo* allows for a wide range of surprise. For the second objection that points out that the two contexts HOMECOMING 1 and HOMECOMING 2 don't exhaust the range of surprise, we concede that it is hard to exhaust the range of surprise through contexts. Therefore, as an approximation to its true felicity condition, we take *lo*'s felicity condition as stated below:

(10) *lo* is felicitous in contexts where the speaker is not biased towards the content of *lo*'s containing clause.

To pointedly verify (10), we construct a context where prior to committing to p, the speaker is biased towards p.

[11] Thanks to Katherine Johnson for making this point.

INFERENCE IN THE LIBRARY: Mina is doing her homework in the library, when she hears some pattering against the window.

(11) Mina: Mujhe lagta hai ki bahar baarish ho rahi hai. (looks out to check) # Lo, bahar baarish ho rahi hai.
Mina: *I think it's raining outside. (looks out to check)* # LO*, it's raining.*

As per (10), *lo* requires that the speaker not be biased towards the content of *lo*'s containing clause before utterance time. This explains (11). Now, while in INFERENCE IN THE LIBRARY, Mina was biased towards the content, the following is also quite possible. Before committing to p by way of using a *lo*-declarative, the speaker's belief state might not be sensitive to the question $?p$, where p is the content of *lo*'s containing clause. Here, in using *question-sensitivity*, we have in mind the notion explored in the philosophy and semantics literature (Yalcin, 2018; Bledin & Rawlins, 2020), whereby a belief state, construed as a set of possible worlds, might not have the resolution that makes a proposition "visible". Less abstractly, surprise can also be a reaction that stems from a general lack of awareness or due to sudden awareness about the situation.
CHRIS AT THE PARTY: A and B are going to a party. They haven't previously entertained the proposition that Chris is at the party, but they see him there. A sees Chris and says:

(12) A: Lo, Chris bhi party par hai.
A: LO*, Chris is also at the party.*

The thing to note in (12) is that A had assigned no credence (neither high nor low) to the proposition that Chris is at the party before seeing Chris. A's belief state was just not sensitive to the question *whether Chris is at the party*. So, it would be inaccurate to say that A had low prior credence to the content of (12) before uttering it. The question had simply not occurred to A. Still, the use of *lo* in (12) is felicitous. This validates (1), our generalization for *lo*, which doesn't require that the speaker have low credence towards the content of *lo*'s containing clause; instead, we framed the condition negatively, i.e., that it not be the case that the speaker had moderate to high credence towards the content at t_0. In cases like CHRIS AT THE PARTY, this condition is fulfilled, as the speaker's belief state's insensitivity to the question *whether Chris is at the party* is indeed compatible with not having assigned moderate to high credence to the proposition that Chris is at the party.

While we have looked at a wide range of contexts variable in the extent of prior speaker bias towards the content of *lo*'s containing clause, we haven't considered the scenario where the speaker has low but nonzero credence towards a proposition p before using a *lo*-declarative which denotes p.
GRADE WORRIES: A is worried about her grade in her math class, which has been quite difficult all semester long. The final scores have just come out.

(13) A: I wonder if I'll even pass the class at all … (checks her report card) Lo, main pass ho gayi!
A: *I wonder if I'll even pass the class at all …* (checks her report card) LO*, I passed!*

The speaker expresses low but nonzero credence to passing before using *lo*. We take *wonder* and *even* to be markers of low credence. It would be inaccurate to say that A has no credence assignment towards the proposition that she will pass. The proposition embedded under *wonder* is the proposition that she will pass the class, and not that she will not pass. Therefore, the expression of bias is towards the highlighted alternative (Farkas & Roelofsen, 2017) that she will pass. And we take this bias towards the proposition to be low due to its embedding under *wonder* and the use of *even*.

An interlude on credences. We pause here to note that our way of talking about credences is different from Farkas & Roelofsen (2017)'s. Farkas & Roelofsen (2017, 256) take a speaker to assign low credence to a proposition p when the speaker considers p to be somewhat more likely than the complement of p. Numerically, we understand this as still assigning >0.5 numerical credence to p. Farkas & Roelofsen (2017, 256) write "if the speaker does not consider α more likely than $\overline{\alpha}$ at all, we say that her credence in α is zero". As far as we understand this, for instances like (13) where the speaker expresses assignment of a numerical credence that is below 0.5—say 0.3 roughly—Farkas & Roelofsen (2017) would consider that numerical credence to be floored to zero on their scale.[12] While of course, this works for Farkas & Roelofsen (2017)'s purposes of understanding rising declaratives and tag interrogatives, it certainly doesn't work for *lo*. A context where the speaker assigns 0.5 to the prejacent and another context where they assign 0.3 do vary for the felicity of *lo*. So, when we say *low credence*, we have the numerical credence in mind, where we can take *low credence* to roughly lie in a range of real values between 0 and 0.4. Then, the empirical takeaway, in light of (13) is that *lo* is felicitous to use when the speaker assigns a low numerical credence to the prejacent.

So far, we've shown that *lo* goes with discovery, is incompatible with uncertainty, and is OK in both moderately unexpected and highly unexpected contexts with no prior speaker bias to the *lo*-marked proposition. We've also shown that it is OK in contexts where the speaker's belief state is not sensitive to the question of *whether p*, where p is the content of *lo*'s containing clause. We have sketched out a large part of *lo*'s empirical profile in declarative clauses, and the generalization that we have offered fits this data.

2.2 Tricky imperatives and the occasional interrogative

For completeness of our empirical analysis, we review the non-declarative examples encountered in the Wortschatz Leipzig Hindi corpus (Leipzig Corpora Collection, 2019) and *Dhoop Kinare*. First, we discuss the potential use of *lo* in imperatives. Then, we grapple with the solitary observed interrogative with *lo*.

Imperatives. Let's begin with an example where the speaker seems to use *lo* in an imperative clause.

[12] A note on em dashes: we use em dashes throughout the paper. While em dash use is often considered a sign of AI generated text (Abebe, 2025), we didn't use AI for the writing of this paper. We used AI minimally for formatting our bib entries.

DINNER: At dinner, B complains that there are only *naan*s and no *roti*s. A makes a *roti*. While handing over the *roti* to B, A says:

(14) A: Lo, yeh kha-o!
A: LO*, eat this!*

While the clause-type of A's utterance is indeed an imperative, containing the *-o* inflection on the verb that Mishra & Archana (2022) gloss as IMP.2PL, it is not clear if the *lo* used sentence-initially in (14) is indeed the discourse particle we've talked about in the context of declaratives in Section 2.1. There are a few confounding elements to note. *Lo* also occurs as the imperative form of the verb *lena*, which means *to take*. *Lena* commands require physical exchange between speaker and addressee as in the exchange of the *roti* in DINNER. We think that *lo* in (14) is not the discourse particle, but the imperative form of the verb *lena* 'to take'. We present the context DOODLE below to set up a contrast with DINNER, to strengthen our point.

In (14), we observe the *roti* being physically exchanged between speaker and addressee. Contrast this with (15), where a command is uttered by the speaker towards the addressee, but there is no exchange of a physical object.
DOODLE: A middle-schooler is doodling on an iPad. The babysitter tells them to do their homework.

(15) Babysitter: # Lo, apna kaam kar-o!
Babysitter: # LO*, do your work!*

The infelicity of (15) further corroborates that *lena* commands require physical exchange. It seems untenable to take the *lo* in (14) to be a particle use. While we don't consider examples like (14) that require an exchange of objects for understanding the distribution of *lo*, we find the following use of *lo* in an imperative environment in the Wortschatz Leipzig corpus (Goldhahn et al., 2012; Leipzig Corpora Collection, 2019) that is more congruent with the particle uses explored in Section 2.1.

(16) Lo, dekho, baarish kaun se darvaje se mere ghar mein aayi.
LO*, look from which door the rain came into my house.*

This example proves to be particularly thorny. There is no physical exchange, which we took to be the confounding factor for examples like (14).

Given that (16) is a corpus example without surrounding context, we can situate (16) in a natural context for a better understanding. The most natural context for (16) is one where the speaker of (16) did not anticipate that the rain would seep into the house through the specific door it did. The identity of the door is remarkable. (16) is felicitous only when the speaker utters it to point out the remarkability, but degraded when the remarkability comment that follows the imperative *dekho* 'see' is absent, i.e. *Lo, dekho*. Something must follow—like a remarkability comment—for the utterance to feel complete. Consider the following example, which corroborates this further.

SHOWING: A real estate agent is walking a client through a new property. The agent wants to ensure the client sees all aspects of the entryway before moving the tour along to the inner rooms.

(17) Agent: # Lo, dekhiye darvaje ko.
Agent: # LO*, look at the door.*

Lo seems incompatible with imperative clauses that aren't followed by a declarative or a rhetorical comment as in (16); when the sequence is strictly an imperative that directs the addressee to take an action as in (17), *lo* is infelicitous.[13] We invite further inquiry into the interaction between the imperatives, the remarkable comment that follows, and the felicity of *lo*; for now, we note that the felicity of *lo* in imperatives like (16) is crucially determined by whether the speaker expresses surprise or remarkability in the comment that follows the imperative. To further give a sense for the sort of environment *lo* is felicitous in, we present some English examples like (18). The reported utterance occurs in an imperative context but has the quality of pointing out something noteworthy rather than commanding an action.

DRESS: A bride-to-be goes wedding dress shopping with her entourage. Her entourage guesses that she would like to try on fitted dresses, but she comes out of the dressing room in an extravagant ballgown.

(18) Mother: Look at that gorgeous dress on you!

RECITAL: Two friends, A and B, are attending a piano recital together, where the pianist executes complex techniques during the performance.

(19) A: Listen to that *glissando*, woah.

In both (18) and (19), despite the utterance occurring in an imperative form, it serves to remark on something notable in the context. In (18), the full entourage is seated, looking at the bride-to-be enter in the ballgown. Instructing the entourage or the bride herself, literally, to *look*, would appear redundant and awkward. (19) seems felicitous to point out a remarkable execution of the *glissando* technique. The utterance in (19) seems less appropriate if literally instructing the interlocutor to listen, when the context already yields that they are doing so. This perspective is further strengthened due to cross-linguistic evidence. 2nd person imperative forms in Romance languages also contribute a mirative meaning (Villalba, 2024). Strikingly, Villalba (2024, 67) cites the marker *goita* corresponding to the verb *guaitar* 'look' in Catalan and many other miratives across Romance languages, derived from the corresponding verb in the language for 'look'. Then *dekho*/*dekhiye* 'look' is expected to have an overlapping distribution with *lo*, as the imperative verb form in the examples above can be taken to contribute a mirative meaning as well. With this provisional perspective, there really is no need to predict *lo*'s distribution in imperatives; examples like (16) are really instances of two miratives occurring together.[14]

[13] We use *direct* to include commands, requests, pleas, etc. (Kaufmann, 2011; Condoravdi & Lauer, 2012).

[14] Of course, why *lo* and *dekho* don't have a complementary distribution due to performing the same func-

Interrogatives. *Lo* appears infelicitous with interrogatives. In the Wortschatz Leipzig Hindi corpus (Leipzig Corpora Collection, 2019), there are 0 occurrences of *lo* in interrogatives. This infelicity is further corroborated by native speaker intuitions. However, we observe a single occurrence of *lo* in *Dhoop Kinare* (Moin, 1987, Episode 2):
PARKING LOT: Four friends are walking through a parking lot and find that one car in particular has been parked completely in their car's way. One of the friends, G, says (20):

(20) G: Lo ji, yeh kis-ki gaadi hai bhai?!
G: LO JI, *whose car is this* BRO?!

It is unclear whether (20) is an information-seeking question. The context is such that none of the four friends would likely be able to provide an answer to the question posed in (20), since they did not see who parked the car and do not recognize the car on sight. The question in (20) is similar to the English *Now, whose car is this?!*, which is similarly not a canonical information-seeking question. Moreover, in (20), *lo* is followed by another discourse particle *ji*. Given *lo*'s one-off and suspect distribution within interrogatives, via our semantic account, we seek to predict that it is hard to mark an interrogative with *lo*.

With a good sense of the empirical picture, which corroborates our generalization in (1), we now proceed to formalization of (1). Leveraging the argumentation above, we argue for incorporating a private beliefs component within the Table model, along with keeping track of the time when biases and commitments are made public.

3 Extending the Table model with time-stamps and Doxa

Here, we present the groundwork for formalizing the semantic generalization in (1) to precisely capture the meaning of *lo*. Our formalization assumes the discourse components in Farkas & Roelofsen (2017), who take participants, table, commitments, and evidence—all to be explained below—to be primitives. Crucially, we propose to add Doxa, a public record of private beliefs. We propose to modify commitments, evidence, and Doxa all to include time-stamps for each of their members. We formalize (1) in this enriched version.

3.1 The Table

Primitives comprise participants, table, commitments, evidence.[15]

(21) a. participants: the set of all participants in the discourse
b. table: a stack with all proposed propositions$_{inq}$[16]

tion is an interesting question, but one we don't explore currently. Jabbar & Kanamarlapudi (2023) provide evidence that particles can be stacked in Hindi-Urdu sentence-finally with propositional anaphors, so the co-occurrence of two particles is certainly not unexpected.

[15]Please note the different typeface used for the discourse components.

[16]This formulation assumes an inquisitive perspective. However, for our purposes, neither the table nor the inquisitiveness matters much. One can assume a picture where conventional update associated with declaratives happens by default, without acceptance of a proposal, and our account for *lo* should still work. Moreover, given our desire of abstracting away from the inquisitive framework for this paper, we specify inquisitive propositions with the subscript *inq* as in *propositions*$_{inq}$ to be able to more freely use *propositions*.

c. commitments: a function that takes participants in participants to sets of propositions. Where $x \in$ participants and W is the set of worlds, commitments $(x) = \{p \in \mathscr{P}(W) : x \text{ is publicly committed to } p\}$.

d. evidence: a function that takes participants in participants to sets of pairs, where each pair includes a proposition and a credence that the participant assigns to that proposition. For instance, $\langle p, [\text{high}] \rangle$ is in the set evidence (x) if x publicizes high credence to the proposition p.

Comment on (21d). We remind the reader that towards the end of Section 2.1, right after (13), we comment that our way of talking about zero, low, moderate, and high credences is different from Farkas & Roelofsen (2017)'s. We take these labels to correspond to intervals on the numerical scale from 0 to 1, whose values are used to express degrees of belief or credences in Bayesian Epistemology (Lin, 2024).

3.2 Extensions

We propose two extensions. First, we motivate the need to keep track of the time of each publicized commitment and bias. Second, we motivate the incorporation of a private beliefs component within the Table model, without which, we argue, surprise, in the way we conceive it and the speaker expresses in using *lo*, cannot be captured. We extend the time-keeping to the private belief component as well.

3.2.1 Time-stamps

(21c) and (21d) contain speaker commitments and biases that are public. As we foreshadowed above, from (21c) and (21d), it is clear that the sets to which both commitments and evidence map a participant x encode no information for *when* x publicized their commitments and biases for particular propositions. We propose to modify commitments and evidence.

(22) commitments: a function that takes participants in participants to sets of proposition-time pairs. Where $x \in$ participants and $p \in \mathscr{P}(W)$, commitments (x) $= \{(p,t) : x \text{ is publicly committed to } p, \text{ and } t \text{ is when } x \text{ publicly committed to } p\}$

(23) evidence: a function that takes participants in participants to sets of pairs of proposition-credence pairs and times. Where $x \in$ participants and $p \in \mathscr{P}(W)$, evidence (x) $= \{(\langle p,c \rangle, t) : x \text{ assigns credence } c \text{ to } p, \text{ and } t \text{ is when } \langle p,c \rangle \text{ was publicized}\}$

Explaining (22) and (23) with some illustrative examples. We co-opt an example presented by Farkas & Roelofsen (2017, 240). Consider APPLICATIONS below, and keep in mind the temporal course of the dialogue.

APPLICATIONS: Belinda and Chris are evaluating applications, when Chris expresses his exasperation by uttering (24). In an effort to keep Chris's disenchantment at bay, Belinda goes over an application and hands it over to Chris, who, thinking that the application is good, says (25). Let the dialogue be such that (24) is said at t_0 and (25) at t_1, where $t_0 < t_1$.

(24) C: None of the applications so far have impressed me.

(25) C: This is a good one ↑ ?

We can consider Chris to be publicly committed to the proposition that none of the applications have impressed him due to his utterance of (24). This can be represented as (26), where $p_1 = [\![\textit{none of the applications have impressed Chris}]\!]$. Further, we follow Farkas & Roelofsen (2017) in taking rising declaratives to signal zero to low credence.[17] Then, (25) can be represented as (27), where p_2 is the proposition that the application that Belinda has asked Chris to consider is a good one.

(26) $(p_1, t_0) \in \mathsf{commitments}(\text{Chris})$.

(27) $(\langle p_2, [\text{zero, low}]\rangle, t_1) \in \mathsf{evidence}(\text{Chris})$

The above examples illustrate how commitments and biases are to be completely represented in a system with time-stamps recording the publicizing time of the commitments and biases. Assuming a Kaplan-Lewis style two-dimensional semantics (Kaplan, 1979; Lewis, 1981), with the context c providing all the discourse components, (21a), (21b), (22), (23), and the speaker sp, the semantics for *lo* can be stated as in (28). No new typed domain is introduced for times, as we can use members of the same domain that supplies parameters to evaluate temporal expressions. However, of course, ts that are used to track the time of the publicizing of commitments and biases aren't performing the function of evaluation of propositions, but instead, that of time-stamping, as stated in (22) and (23).

(28) $[\![lo]\!]^{c,\langle w,t\rangle} = \lambda p : \underbrace{(\langle p, [\text{moderate, high}]\rangle, t_{-1}) \notin \mathsf{evidence}(sp)}_{\text{Presupposition: no moderate to high credence for } p}. \quad \underbrace{(p,t) \in \mathsf{commitments}(sp)}_{\text{Discourse effect: speaker commits to } p}$

We show that **(28) is incorrect**. We show that its failing lies in making reference to a publicized component in *lo*'s presupposition. This motivates our second extension, i.e., incorporation of a private beliefs component within the Table model.

3.2.2 The need for privacy

An important feature of both (21c) and (21d)—and (22) and (23) by extension—is that they record commitments and biases that are public. However, we think that the capture of surprise and *lo*'s distribution require reference to the private beliefs of the speaker. We argue for this from an example below. Consider the context RAIN 3 below.

RAIN 3: After a while of sitting close to the window in the living room, Gabriel decides to move to his windowless room. Michael, who is Gabriel's friend, is in Gabriel's windowless room. Gabriel knows that it is raining, something he observed when sitting by the window. However, Michael does not know this, as Gabriel hasn't publicized his commitments or

[17] As this discussion is for illustrative purposes, we stick with Farkas & Roelofsen (2017)'s way of speaking about credences, which we otherwise eschew.

bias towards the proposition that it is raining. It turns out to be infelicitous for Gabriel to reply to Michael's inquiry as in (29):[18]

(29) Michael: Is it raining outside?/I wonder whether it is raining outside.
Gabriel: # Lo, bahar barish ho rahi hai.
Gabriel: # *Lo, it is raining outside*.

However, (28) wrongly predicts *lo*'s use as in Gabriel's answer to Michael's question to be felicitous: where $p = [\![\textit{it is raining outside}]\!]$ and t_0 is utterance time, with $t_{-1} < t_0$, $(\langle p, [\text{moderate, high}]\rangle, t_{-1})$ is **not** in evidence(Gabriel), and, after Gabriel's utterance in (29), (p, t_0) **is** in commitments(Gabriel). Thus, (29) fulfills the presupposition of (28) and brings about the specified discourse effect too. But Gabriel's use of *lo* is infelicitous. The example also shows what goes wrong in general if we try to model surprise with public biases and the condition that the speaker had not previously publicized bias towards the proposition that the speaker ends up committing to by way of assertion of the proposition. To concretely make the general point, consider how it is similarly strange for Gabriel to say any of the following, in response to Michael's question, given RAIN 3.

(30) a. Wow, it is raining.
b. Whoa, it is raining.

What does the above mean for the extension to the Table model? The function evidence as laid out in (23) maps each participant to their public biases. There is no private belief component in (21)-(23). Without a reference to the private beliefs of the speaker, we can't really model speaker surprise, which should be inexpressible in cases like (29), where the speaker privately believes the proposition yet to be publicly committed to, by way asserting *lo*+p, where p = the content of *lo*'s containing clause. However, if we introduce a private belief component, we have to reconcile it with the fact that the Table model seeks to represent the public record of the conversation. As a result, we introduce Doxa.

3.2.3 Doxa

Doxa is the public record of the private beliefs of the speaker. Let's explain this further. First, note that the public record of the private beliefs of the speaker need not represent the private beliefs accurately. We propose to include Doxa as a component alongside those listed in (21).

(31) Doxa: a function that takes participants in participants to sets of pairs, where each pair includes a proposition-credence pair—where the credence encodes the participant's confidence in the proposition's truth—and a time-stamp.
If $(\langle p, [\text{moderate, high}]\rangle, t) \in \text{Doxa}(x)$, then it is public information that x is privately biased towards p at t.

[18] In this example, we have a question *is it raining outside?*. However, we have also provided an alternative: *I wonder whether it is raining outside.* Our rationale behind including the alternative is that none of the examples in earlier sections include a *lo*-declarative used as a response to a question, so the fact that *lo* is used in response to a question may be considered as a confounding factor for *lo*'s infelicity in (29).

While the members of evidence are formally similar to members of Doxa, the conditional in (31) that we isolate for pointed comment in (32) brings out the difference between evidence and Doxa.

(32) If $(\langle p, [\text{moderate, high}]\rangle, t) \in \text{Doxa}(x)$, then it is public information that x is privately biased towards p at t.

Note that in preface to (31), we noted that the public record of the private beliefs of the speaker need not represent the private beliefs accurately. Then, we also seem to take it as a sufficient condition for private bias towards p that Doxa contain p as in $(\langle p, [\text{moderate, high}]\rangle, t) \in \text{Doxa}(x)$. On the surface, this seems like an inconsistency, as the sufficiency of p's membership of $\text{Doxa}(x)$ for x's private bias towards p seems to require accurate tracking of private beliefs for Doxa. However, the consequent in (32) can be thought of as containing a publicity operator, i.e., *it is **public** information*. Then x can privately believe that the best Shakespeare play is *Measure for Measure*, while it being public information that x privately believes that *Hamlet* is the best play.

Objection. Now, an objection that can be raised right away is that if all we achieve through introduction of Doxa is the sort of private-public duality of belief as exemplified by the Bard example above, then this can also be achieved via commitments and evidence.[19] We surely don't take commitments and evidence to reliably track private states, so the duality comes by design, and expressions in language never really require reliable tracking of private belief states. Then, the objection continues: unless we show that we can indeed keep track of private beliefs—something that commitments and evidence don't record—there isn't really a point to introducing Doxa. Even if we show that such tracking is possible, to motivate Doxa, we also need to show that linguistic expressions make reference to private states. However, we aren't arguing that private beliefs can be reliably tracked. Then, what's the point of Doxa? And can Doxa be motivated on empirical grounds?

Reply. The point of Doxa is that, at utterance time, one can reveal that they privately believed or didn't believe a proposition at an earlier time—a revelation about a retrospective private belief state where the private belief was present but wasn't publicized. More explicitly: a speaker can reveal at utterance time that at an earlier time t_{-1}, the speaker didn't believe that p. Moreover, this can be done while the speaker expresses their belief that p. In fact, this is exactly what happens in many cases of mirative use; only after the speaker's utterance of mirative+p, it becomes public that the speaker didn't expect p to be true. That the speaker didn't believe p at t_{-1} was not publicized by the speaker at t_{-1}. Commitments and evidence don't contain propositions unless they are publicized. One can further resist introduction of Doxa on the grounds that we can allow commitments and evidence to contain earlier times too. However, $(\langle p, [\text{moderate, high}]\rangle, t_{-1}) \in \text{evidence}(x)$ would simply mean that x publicized their bias towards p at t_{-1}. Commitments and evidence, by design, are such that things cannot enter them paired with t without getting

[19] There's a lot of work that points out the public-private duality of belief (Yalcin, 2012; Goldstein & Kirk-Giannini, 2022; Bary, 2025), often to ask the reader not to confuse common ground with common belief.

publicized at t. Whereas, one need not publicize that they believe that p for it to later enter Doxa that they privately believed that p at an earlier time. One publicizes the commitment to or bias towards propositions that go into commitments and evidence, but one publicizes information that reveals that one is/was privately biased towards the proposition and this proposition goes into Doxa—not only with the time of the utterance, but possibly also an earlier time. This shows conceptually that Doxa captures something that commitments and evidence don't. We bring out these differences empirically in the calculations in Section 4. Now, we move to the unique empirical coverage that Doxa brings.

Unique empirical coverage. We want to be able to say that in (29), by using *lo* as in *lo, it is raining*, Gabriel reveals to Michael, i.e., publicizes, that Gabriel didn't privately believe right before utterance time that it is raining. This is what contributes to *lo*'s infelicity, as Gabriel is in a windowless room and it is not possible for him to form the belief that it is raining, on the spot. The same point can be made with *whoa* and *wow*—at least for examples like (30)—where the speaker publicizes that they didn't privately believe the prejacent right before committing to the prejacent. Of course, in our discussion here, we have used *believe*, instead of talking about biases or credences, which is exactly what we shift to, in our explanations and calculations in Section 4.

To conclude this section, we emphasize that it is the public recording of whether the speaker possessed a certain belief privately at a time earlier than utterance time that the components in (21) cannot do. Moreover, it is quite compatible with the aims of the Table model, i.e., modeling conventional discourse effects, to make reference to private beliefs in a discourse component like Doxa if an expression can be taken to conventionally encode the information that at a prior time the speaker didn't hold a certain belief privately. This information is made public via the use of the expression.

4 The Table-borne semantics for *lo*

At the outset in Section 1, we stated the semantic generalization for *lo*. We repeat it below.

(33) *lo* is felicitous to use at t_1, if the speaker commits to p at t_1, while not having moderate to high credence towards p at t_0, where p is the content of *lo's* containing clause and $t_0 < t_1$.

Through our discussion in Section 3.2.3, we argued that it need not be public that the speaker does not have moderate to high credence towards p at t_0; instead, **this lack of bias could be private until it is revealed at utterance time**. The crucial thing to note is that with the use of *lo*, the speaker signals that they weren't biased towards p before they committed to it. This past lack of bias can either be signaled by having explicitly expressed bias towards the complement of the proposition or be signaled at utterance time by the use of *lo* itself. The public record that the speaker doesn't have the private belief that p at t entails that the speaker does not have any public bias towards p at t. We can lay this out more explicitly in (34) and (35). Where c is a Kaplanian context and t is a time s.t. t may be a time other than the one initialized by c:

(34) $\forall c \forall t : p \in \mathsf{evidence}_{c,t}(x) \rightarrow p \in \mathsf{Doxa}_{c,t}(x)$

(35) $\forall c \forall t : p \in \mathsf{commitments}_{c,t}(x) \rightarrow p \in \mathsf{Doxa}_{c,t}(x)$

With the above specifications in place, we can say that the sub-condition "while not having moderate to high credence towards p at t_0" in (33) should be modeled with Doxa.

We have verified the generalization in (33) with all the data presented in Section 2. We present a formalization of (33) now. We assume a two-dimensional semantics in the Kaplan-Lewis style (Kaplan, 1979; Lewis, 1981). We take c to denote the context of utterance, which includes, *inter alia*, the speaker of the utterance, notated as sp, and all the primitives of our version of the Table model, i.e., participants, table, commitments as in (22), evidence as in (23), and crucially Doxa as in (31).

(36) $[\![lo]\!]^{c,\langle w,t\rangle} = \lambda p : \underbrace{(\langle p, [\text{moderate, high}]\rangle, t_{-1}) \notin \mathsf{Doxa}(sp)}_{\text{Presupposition: no moderate to high credence for } p}. \quad \underbrace{(p,t) \in \mathsf{commitments}(sp)}_{\text{Discourse effect: speaker commits to } p}$

We show the workings of (36) with a few calculations. Let's first start with (29)—the problematic case that motivated the incorporation of the private beliefs component.

(37) $\lambda p : (\langle p, [\text{moderate, high}]\rangle, t_{-1}) \notin \mathsf{Doxa}(sp).(p,t) \in \mathsf{commitments}(sp)(\textit{it is raining})$
Discourse effect: $(\textit{it is raining}, t) \in \mathsf{commitments}(sp)$
Presupposition: $(\langle \textit{it is raining}, [\text{moderate, high}]\rangle, t_{-1}) \notin \mathsf{Doxa}(sp)$

(36) shows exactly why Gabriel's use of *lo* is infelicitous in (29). We specified in (29) that Gabriel moves into a windowless room where Michael asks him if it is raining. There is no obvious way—at least obvious to Michael—for Gabriel to *not* have privately believed, right before utterance time, that it is raining in order to be able to cooperatively assert that it is raining at utterance time.[20] That is what makes the accommodation of the Presupposition in (37) hard too, leading the addressee to think that Gabriel knew that it was raining before utterance time too. If, for instance, Gabriel instead prefaces his answer with *let me check the weather app* and finds out that it is raining, upon formation of this belief, Gabriel can indeed use *lo*, and in that instance, Gabriel's utterance as in (29) would be felicitous.

Consider another example. In (11), the speaker, Mina, publicizes her bias towards the proposition that it is raining by saying *I think it's raining outside*. The proposition that it is raining, paired with high credence, enters evidence(Mina), and by the entailment outlined in (34), Doxa(Mina) as well. As this expression of bias occurs before the utterence of *lo*'s containing clause, the presupposition $(\langle \textit{it is raining}, [\text{moderate, high}]\rangle, t_{-1}) \notin \mathsf{Doxa}(\textit{Mina})$ is not met, predicting infelicity for the use of *lo* in a following utterance as in *Lo, it is raining outside*. And indeed this utterance in (11) is infelicitous.

We also use (36) to show why *lo* doesn't go well with interrogatives. Take a canonical information-seeking question such as *is it raining?*. (36)'s discourse effect is that the speaker commits to the prejacent, which we can take to be the proposition denoted by the

[20] By *cooperative* as in *cooperatively assert*, we specify assertion of propositions for which the speaker has adequate evidence (Grice, 1975).

sentence radical *is it raining?*.[21] Then, in requiring commitment to the sentence radical, *lo*'s use in an interrogative, whose purpose is to seek information, turns out to be infelicitous. The examples considered above show the role of Doxa, how it interacts with evidence as specified in (34), and how in making reference to Doxa in the semantic clause for *lo* as in (36), we capture data like (29), which motivated the need for incorporation of a private beliefs component. As our contribution seeks to model mirativity within the Table framework, we compare our account with Kraus (2019), who seeks to do the same.

5 Comparison with Kraus (2019)

Kraus (2019)'s contribution. Kraus considers the cases of English miratives *oh* and *huh*. Kraus argues that these miratives require speaker's expectation-violation. Where p is the content of the mirative's containing clause and $\mathsf{Exp}(p)$ is supposed to capture speaker's expectation for p, Kraus proposes the mirative to contribute the following change to the state of the discourse: $\mathsf{Exp}(p) < \mathsf{Exp}(\neg p)$ is added to the speaker's discourse commitments, i.e., the changed discourse contains $[\mathsf{Exp}(p) < \mathsf{Exp}(\neg p)] \in \mathsf{commitments}(sp)$ as the new information. Further, Kraus thoroughly considers a range of prosodic data, while also nicely capturing a subtle difference between the contribution of *oh* and *huh*. For this, Kraus relies on the divergence in examples such as (38)—taken from Kraus (2019, 26)—to construct a slightly modified effect for the two, where in felicitous uses of *huh*—but not of *oh*—, the speaker must have low, but nonzero credence towards the surprising proposition, i.e., $\mathsf{Exp}(p) > 0$.

(38) *Speaker, oblivious, rounds the corner and is hit with confetti.*
   a. Oh! I didn't expect that!
   b. # Huh! I didn't expect that!

Problems. First, we note that Commitments for a participant x is intensionally defined as containing the propositions that x is publicly committed to.[22] And $[\mathsf{Exp}(p) < \mathsf{Exp}(\neg p)]$ is not a proposition. There's another worry here. Kraus's modification to the Table includes a stack of pending at-issue propositions yet to be accepted or rejected. Further, Kraus requires that $[\mathsf{Exp}(p) < \mathsf{Exp}(\neg p)]$ continue to remain in $\mathsf{commitments}(sp)$ even after p is resolved, i.e., taken off the Table. Then, if $[\mathsf{Exp}(p) < \mathsf{Exp}(\neg p)]$ seeks to model that the speaker doesn't expect p to be true, then $[\mathsf{Exp}(p) < \mathsf{Exp}(\neg p)]$'s persistence in $\mathsf{commitments}(sp)$ creates a discourse context where p is resolved and the speaker still doesn't

[21] Roelofsen & Van Gool (2010), and later Farkas (2011), Roelofsen & Farkas (2015), and Farkas & Roelofsen (2017) appeal to a level of semantic representation that they call *highlighted content*. Theiler (2021) makes a compelling case for making reference to highlighted content for explaining the semantic contribution of the discourse particle *denn* and *überhaupt* in German. Jabbar & Kanamarlapudi (2024) do the same for the Hindi-Urdu *bhala*. While we opt for the simpler option currently with sentence radicals, one can, in principle, use highlighted propositions in (36) and the calculations.

[22] We would like to acknowledge that the problems raised in these paragraphs are based on our read of Kraus (2019), which can very well be incorrect.

expect p. Not only is this context state undesirable on its own, it leads to the prediction that the speaker could in principle keep expressing surprise felicitously at repeated assertions of the proposition at which the speaker expressed surprise using *oh* at first. To see how this follows, consider (39) from Kraus (2019, 29).

(39) A: The bank is closed today. $= p$
B: Oh, I need to deposit a check. $= q$

Kraus (2019, 30) proposes that, after B's utterance of *oh*+q, commitments(B) contains: q, $\mathsf{Exp}(q) \approx 1$, and $[\mathsf{Exp}_B(p) < \mathsf{Exp}_B(\neg p)]$. Kraus specifies that the common ground is unchanged. p that was put on the table after A's assertion is popped after B's utterance of *oh*+q. Then, later in the conversation, if A were to assert p again, given that $[\mathsf{Exp}_B(p) < \mathsf{Exp}_B(\neg p)]$ is still in commitments(B), B can express surprise again with *oh*.

Our account differs from Kraus's also in that we consider miratives in environments with discovery of a fact, whose propositional content p may not be at-issue, and instead the speaker uses the mirative with p as in *lo*+p or *oh*+p. Consider (7), where the speaker says *Lo, you're home at 4pm today*, where the proposition that the addressee is home at 4pm today is surprising. All of Kraus's example include contexts where the speaker of the mirative uses the mirative to settle the issue at hand, expressing surprise at the content of the issue on the table, and not at the content of the mirative's containing clause. Our account doesn't require the restriction that the speaker of a *lo* utterance be responding only to an utterance relayed to them by an interlocutor.

Given that in (7), the speaker can say *oh, you're back home at 4pm today*, Kraus's account makes the following prediction, we think. Where $p = [\![$*you're back home at 4pm today*$]\!]$, the speaker utterance *oh, you're back home at 4pm today* not only adds $[\mathsf{Exp}(p) < \mathsf{Exp}(\neg p)]$ to speaker's commitment set, but also the content of *oh*'s containing clause, i.e., p. This leads to a commitment set with both $[\mathsf{Exp}(p) < \mathsf{Exp}(\neg p)]$ and p, meaning that the speaker doesn't expect p and that the speaker is committed to p. This is surely to be avoided.

Another issue due to $[\mathsf{Exp}(p) < \mathsf{Exp}(\neg p)]$'s inclusion in the speaker's commitment set due to using a mirative is that there's a considerable chunk of contexts like (12) that Kraus rules out as bad for mirative use, which in fact license the use of miratives. These are contexts where the speaker may not have any attitude towards the event or proposition at which the speaker expresses surprise using the English *oh* or the Hindi-Urdu *lo*. Such contexts may be taken to be instances of *sudden awareness* (Aikhenvald, 2012, 462), or more formally, as those where the speaker's belief state is just not sensitive to the relevant question (Yalcin, 2018; Bledin & Rawlins, 2020). In such contexts, the speaker can use the English *oh* and the Hindi-Urdu *lo*. The characteristic feature of examples like (12) is that the speaker had no bias either for p or $\neg p$, when the speaker ended up using *lo*+p.

Remedy. Currently ignoring that commitments require that its members have a certain type, we think that Kraus's proposal can be made to work, for the most part, if we include time-stamps.[23] More specifically, if by using a mirative, the speaker expresses $[\mathsf{Exp}(p) <$

[23]We add the qualification *for the most part*, as our suggested remedies here don't account for all of the raised objections.

$\mathsf{Exp}(\neg p)]_{t_{-1}}$, where the subscripted t_{-1} indicates that this expectation held at t_{-1}, a time before utterance, then the speaker commitments set can coherently contain both [$\mathsf{Exp}(p) < \mathsf{Exp}(\neg p)]_{t_{-1}}$ and p which was publicized at utterance time, as long as commitments records p with its publicizing time, i.e., t. We propose exactly this innovation in our paper. We think that, while Kraus doesn't make reference to time-stamps in the modeling, Kraus is presciently aware of their importance; Kraus (2019, 25) writes: "*Oh* indicates that the speaker **had** higher expectations for $\neg p$ over p" (emphasis added). The use of the past tense, even when not modeling it formally, further underscores the need for making reference to the time when each member of commitments, evidence, and Doxa entered these sets.

6 Conclusion

In addition to mapping out the empirical profile of a previously unexplored discourse particle in Hindi-Urdu, we have argued for the necessity of time-stamps for modeling mirativity within the Table model. We brought this out further in engagement with Kraus (2019), who seeks to model mirativity without reference to times. We highlight a few issues with Kraus's approach, many of which, we show to be resolvable by the simple inclusion of time-stamps. Further, with our discussion in Section 3.2.2 of examples like (29), we showed that speaker's previously unpublicized beliefs can be revealed—interestingly when the speaker no longer holds these beliefs. In this way, the speaker never publicizes these beliefs, but they come to be on public record, as constitutive of the speaker's belief state at an earlier time. Again, it is time-stamps that afford us the expressive power to be able to say the above. We hope that our contribution invites the semantics community to make further progress on modeling mirative expressions within the Table model. We also think that our inclusion of time-stamps and the public record of private beliefs via Doxa provides an opportunity to explore the meta-theory of the Table in future work.

References

Abebe, Nitsuh. 2025. With the Em dash, A.I. embraces a fading tradition. *The New York Times Magazine* `https://www.nytimes.com/2025/09/18/magazine/chatgpt-dash-hyphen-writing-communication.html`.

Aikhenvald, Alexandra Y. 2012. The essence of mirativity. *Linguistic typology* 16(3). 435–485.

Bary, Corien. 2025. Speech acts, common ground and commitments. *Linguistics and Philosophy* 48(3). 505–526.

Bhatt, Rajesh & Veneeta Dayal. 2020. Polar question particles: Hindi-urdu *kya*. *Natural Language & Linguistic Theory* 38. 1115–1144.

Bledin, Justin & Kyle Rawlins. 2020. Resistance and resolution: attentional dynamics in discourse. *Journal of Semantics* 37(1). 43–82.

Ciardelli, Ivano, Jeroen Groenendijk & Floris Roelofsen. 2018. *Inquisitive semantics*. Oxford University Press.

Condoravdi, Cleo & Sven Lauer. 2012. Imperatives: Meaning and illocutionary force. *Empirical issues in syntax and semantics* 9. 37–58.

DeLancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1. 33–52.

DeLancey, Scott. 2001. The mirative and evidentiality. *Journal of Pragmatics* 33(3). 369–382.

Farkas, Donka F. 2011. Polarity particles in English and Romanian. In *Romance linguistics 2010*, 303–328. John Benjamins Publishing Company.

Farkas, Donka F & Kim B Bruce. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27(1). 81–118.

Farkas, Donka F & Floris Roelofsen. 2017. Division of labor in the interpretation of declaratives and interrogatives. *Journal of Semantics* 34(2). 237–289.

Goldhahn, Dirk, Thomas Eckart, Uwe Quasthoff et al. 2012. Building large monolingual dictionaries at the Leipzig corpora collection: From 100 to 200 languages. In *LREC*, vol. 29, 31–43.

Goldstein, Simon & Cameron Domenico Kirk-Giannini. 2022. Contextology. *Philosophical Studies* 179(11). 3187–3209.

Grice, H. Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics, vol. 3: Speech acts*, 41–58. New York: Academic Press.

Gunlogson, Christine. 2008. A question of commitment. *Belgian Journal of Linguistics* 22(1). 101–136.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Malden, MA: Wiley Blackwell.

Jabbar, Ahmad. 2023. The Hindi-Urdu *na* and reasonable inference. *Proceedings of the 59th Chicago Linguistic Society (CLS 59)* Unpublished manuscript.

Jabbar, Ahmad & Veda Kanamarlapudi. 2023. Grounding with particles. In *Proceedings of the 27th workshop on the Semantics and Pragmatics of Dialogue (SemDial 27)*, .

Jabbar, Ahmad & Veda Kanamarlapudi. 2024. *Bhala*: Subordinating highlighted propositions. *Presentation at the 60th annual meeting of the Chicago Linguistic Society (CLS 60)* .

Kaplan, David. 1979. On the logic of demonstratives. *Journal of Philosophical Logic* 8. 81–98.

Kaufmann, Magdalena. 2011. *Interpreting imperatives*, vol. 88. Springer Science & Business Media.

Kraus, Kelsey. 2019. Intonation and expectation: English mirative contours and particles. In *Proceedings of sinn und bedeutung*, vol. 23 2, 19–36.

Lauer, Sven & Cleo Condoravdi. 2012. The basic dynamic effect of interrogative utterances. In *13th texas linguistics society conference, university of texas, austin*, .

Leipzig Corpora Collection. 2019. Hindi web text corpus (india), 2019. `https://dict.wortschatz-leipzig.de?corpusId=hin-in_web_2019`.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8. 339–359.

Lewis, Davis. 1981. Index, context, and content. *Philosophy and grammar: Papers on the occasion of the quincentennial of Uppsala University* 79–100.

Lin, Hanti. 2024. Bayesian Epistemology. In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University Summer 2024 edn.

Mishra, Anushree & S Archana. 2022. An analysis of imperatives in Hindi-Urdu. *Coyote Papers* 60–71.

Moin, Haseena. 1987. Dhoop Kinare. Television drama series. Screenplay by Haseena Moin. Broadcast on Pakistan Television (PTV).

Roelofsen, Floris & Donka F Farkas. 2015. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language* 91(2). 359–414.

Roelofsen, Floris & Sam Van Gool. 2010. Disjunctive questions, intonation, and highlighting. In *Logic, language and meaning: 17th Amsterdam Colloquium, Amsterdam, The Netherlands, December 16-18, 2009, revised selected papers*, 384–394. Springer.

Theiler, Nadine. 2021. *Denn* as a highlighting-sensitive particle. *Linguistics and Philosophy* 44(2). 323–362.

Villalba, Xavier. 2024. Romance exclamative markers at the syntax-pragmatics interface: A compositional approach to exclamativity. *Journal of Pragmatics* 226. 64–77.

Wikimedia Foundation. 2025. Sibi district. `https://en.wikipedia.org/wiki/Sibi_District`. Accessed: 2025-11-09.

Yalcin, Seth. 2012. Bayesian expressivism. *Proceedings of the Aristotelian Society* 112(2). 123–160.

Yalcin, Seth. 2018. Belief as question-sensitive. *Philosophy and Phenomenological Research* 97(1). 23–47.

Re-evaluating the Status of *rah-* within the Aspectual System of Hindi: From Auxiliary to Progressive Affix

BISHWANATH KUMAR, *Seoul National University*

ABSTRACT

This paper re-evaluates the status of *rah-* within the aspectual system of Hindi. While previous studies have often classified *rah-* as an auxiliary verb, I argue that it has grammaticalized further into a bound progressive morpheme. The analysis is based on four diagnostics: incompatibility with the habitual marker *-t-*, restrictions on negation, absence in negative imperatives, and morphological opposition with habitual *-t-*. These distributional patterns demonstrate that *rah-* diverges from the profile of canonical auxiliaries and vector verbs, which inflect independently, modify event structure, and participate in monoclausal [V1+V2] constructions.

A central contribution of this study is to refute the long-standing "perfective *rahā*" hypothesis. Because the progressive form *rahā* is formally homonymous with the perfective participial ending *-ā*, earlier accounts have analyzed progressive constructions as containing a perfective morpheme. This yields the paradoxical conclusion that the progressive encodes perfective aspect—an untenable analysis, since progressive and perfective are incompatible. By distinguishing the true perfective marker *-(y)ā* from agreement suffixes *-ā/-ī/-e*, and by recognizing *rah-* as the sole progressive exponent, the analysis resolves this inconsistency.

The findings are situated within complex predicate theory and the grammaticalization literature. The trajectory of *rah-* aligns with Lehmann's grammaticalization cline (lexical verb → auxiliary → TAM affix) and illustrates Hopper & Traugott's principles of decategorialization and persistence. Typologically, the Hindi evidence contributes to the well-documented tendency for posture or existence verbs ('stay, remain, sit, stand') to develop into progressive markers.

This study clarifies the synchronic role of *rah-* as an affixal progressive morpheme and revises our understanding of Hindi TAM morphology. It also raises broader implications for the analysis of Indo-Aryan aspect systems, for the theory of complex predicates, and for typological models of grammaticalization.

1 Introduction

This paper re-evaluates the morphological status of *rah-* in Hindi, arguing that it functions as a progressive aspect morpheme (affix) rather than an auxiliary or light verb. While *rah- h-* is often treated as an auxiliary sequence in the descriptive tradition, I contend that it is better analyzed as a constructional combination of the progressive marker *rah-* with the tense auxiliary *honā* 'be', where *rah-* itself is affixal in distribution

and behavior.[1] Diachronically, *rah-* derives from the lexical verb *rahnā* 'stay, remain', but in present-day Hindi it has shed verbal and auxiliary properties and grammaticalized toward bound TAM morphology.

Empirically, I revisit four diagnostics that separate *rah-* from bonafide auxiliaries and from vector verbs in compound predicates. (i) Habitual co-occurrence: auxiliaries (including vector and modal V2s) may appear with the habitual morpheme *-t-*, but *rah-* does not. (ii) Negation syntax: modal auxiliaries freely combine with negation, while *rah-* cannot surface in comparable negated environments. (iii) Negative imperatives: standard auxiliary placement allows a negator between V1 and Aux, but such placement is unavailable with *rah-*. (iv) Morphological opposition: *rah-* stands in a systematic morphological opposition to the habitual *-t-*, each pairing with the same V1 to realize distinct aspectual values.

On a grammaticalization account, these facts indicate that *rah-* has undergone decategorialization (loss of verbal/auxiliary properties) and tightened morphotactic bonding to the host verb (affix-like status), aligning with the classic main verb → auxiliary → TAM affix trajectory.

Structure of the paper. Section 2 reviews the auxiliary/light verb/vector landscape in Hindi and motivates a set of baseline expectations from the literature (especially Butt 2010; Hook 1973). Section 3 presents four diagnostics and shows that *rah-* fails to behave like a V2 or auxiliary while forming a morphological opposition with *-t-*. Section 4 develops a grammaticalization analysis, relating the Hindi facts to broader TAM pathways. Section 5 refutes the "Perfective *rahā*" Hypothesis. Section 6 concludes.

2 Auxiliaries and Vector Verbs in Hindi

In the descriptive tradition of Hindi grammar, auxiliaries are classified into three types. Tense auxiliaries such as *honā* 'be' realize present and past tense distinctions (1a–b). Modal auxiliaries such as *saknā* 'can', *chuknā* 'finish, be completed', and *pānā* 'manage to' contribute meanings of ability, completion, or possibility (1c).

(1) a. main roj vidyālay jā-t-ā hūn.
1SG.NOM daily school go-HAB-M.SG be.PRES.1SG
'I go to school every day.'

b. main roj vidyālay jā-t-ā thā
1SG.NOM daily school go- HAB-M.SG be.PST.M.SG
'I used to go to school every day.'

c. main Russian bol sak-t-ā hūn
1SG.NOM Russian speak can- HAB-M.SG be.PRES.1SG
'I can speak Russian.'

[1] Previous studies have attempted to classify *rah-* as an auxiliary. Kellogg described *rah-* in the periphrastic form *rah- h-* as an example of an auxiliary (cited in Olphen 1975:296), and Deo (2006:6, 176) likewise discusses *rah-* as a category of auxiliaries.

A third group, vector verbs, includes *lenā* 'take, receive', *denā* 'give', *jānā* 'go', *ānā* 'come', *baithnā* 'sit', *paṛnā* 'fall', *uṭhnā* 'rise', among others. These combine with a main verb to nuance aspectual or pragmatic interpretation, as in (2a~b).

(2) a. main-e president-ko khat likh diyā
1SG-ERG president- DAT letter write give.PFV.M.SG
'I (simply/decisively) wrote a letter to the president.' (Hook 1973:18)

b. nadya-ne khat likh li-ya.
Nadya-ERG letter write take-PFV.M.SG
'Nadya wrote a letter (completely).' (Butt 2010:4)

Hook (1973:17) defines such compound verbs as a structure where V1 is the lexical root and V2 an inflected secondary element (explicator, auxiliary, or operator). Crucially, no other grammatical component may intervene between V1 and V2 which highlights the strong bond between V1 and V2. Butt (2010:3~6) emphasizes that vector verbs in Hindi/Urdu are part of monoclausal complex predicates, reshaping event structure (completion, benefaction, suddenness, etc.), yet remain distinct from auxiliaries.

Scholars disagree on how many auxiliaries exist in Hindi. Estimates range from 11 (Guru 1977) to 47 (Nespital 1997), but there is general consensus on a core set, including *jānā* 'go', *lenā* 'take', *denā* 'give', *paṛnā* 'fall/befall', *uṭhnā* 'rise/get up', *baithnā* 'sit', and *ānā* 'come' (cited in Bhat 2001:2).[2] *Rahnā* is notably absent from these inventories, indicating its marginal status in the auxiliary system. Debasri (2008:120) likewise provides a comprehensive list of V2s that occur in [V1 + V2] constructions, yet *rahnā* does not appear there either.

In summary, the hierarchy of binding strength in the Hindi auxiliary system can be schematized as [tense auxiliaries < modal auxiliaries < vector verbs]. Vector verbs exhibit the closest integration with V1, but as the following section will show, *rah-* demonstrates an even tighter morphological bond, setting it apart from auxiliaries altogether.

3 Morpho-syntactic Properties of Progressive *rah-*

As shown in Section 2, the Hindi auxiliary system comprises tense auxiliaries (*honā* 'be'), modal auxiliaries (*saknā* 'can'*, pānā* 'manage to'*, chuknā* 'finish, be completed'), and vector verbs (*lenā* 'take'*, denā* 'give'*, jānā* 'go'*,* etc.). The progressive morpheme *rah-* has often been grouped into this system, but its morpho-syntactic behavior diverges sharply from these categories. In this section, I examine four core diagnostics: (i) compatibility with the habitual morpheme *-t-*, (ii) behavior under negation, (iii) placement in negative imperatives, and (iv) morphological opposition with habitual *-t-* and show that *rah-* behaves unlike auxiliaries or vector verbs.

[2] Other inventories include Kachru (1966), Bahl (1967), McGregor (1977), Hook (1974).

3.1 Restriction on Combination with the Habitual Morpheme

The habitual morpheme *-t-* is a central exponent of imperfective aspect in Hindi/Urdu (Masica 1991:327~328; Deo 2006:129). Auxiliaries, whether modal (*saknā* 'can') or vector (*jānā* 'go'), freely combine with *-t-* to yield habitual interpretations. *Rahnā*, however, cannot.

(3) a. Rām bhūl jā-t-ā hai
Ram.NOM forget go- HAB-M.SG be.PRES.3SG
'Ram tends to forget / forgets (habitually).'

b. *Rām bhūl rah-t-ā hai
Ram.NOM forget stay- HAB-M.SG be. PRES.3SG
(intended: 'Ram keeps forgetting')

The ungrammaticality of (3b) contrasts with the regular auxiliary in (3a). As Olphen (1975:296) notes, even if forms like *jā rahtā hai* surface, they lack genuine progressive interpretation and are semantically anomalous. This restriction signals that *rah-* no longer behaves like a regular auxiliary verb, but rather as a bound aspectual marker incompatible with *-t-*. Below are the additional examples that reinforce this point.

(4) a. Monu Korean bol sak-t-ā hai
Monu.NOM Korean speak can-HAB-M.SG be. PRES.3SG
'Monu can speak Korean.'

b. *Monu Korean bol rah-t-ā hai
Monu.NOM Korean speak stay-HAB-M.SG be.PRES.3SG
(intended: 'Monu speaks Korean habitually')

These data demonstrate that *rah-* cannot occupy the same morphosyntactic slot as auxiliaries vis-à-vis *-t-*. This emphasizes that incompatibility with habitual morphology is a diagnostic of grammaticalization toward bound TAM morphology.

3.2 Behavior under Negation

Modal auxiliaries combine freely with negation (Hook 1974:85), but *rah-* cannot appear in parallel contexts.

(5) a. main khat likh nahīn sak-ā
1SG.NOM letter write NEG can-PFV.M.SG
'I was not able to write a letter.' (Hook 1974:85)

b. *main khat likh nahīn rah-ā
1SG.NOM letter write NEG PROG-M.SG

(intended: ‘I was not writing a letter’)

In contrast, compound verbs with vector auxiliaries combine with negation only in restricted ways (Hook 1974:98~100, 201). Yet *rah-* constructions freely license sentential negation, provided the negator precedes the entire [V1 + rah- h-] complex. This indicates that *rah-* is not “negation-friendly” in the auxiliary slot, but instead patterns like an affix within the verbal complex.

(6) a. *mai-ne usko paise nahin de diye
1-ERG 3SG.DAT money NEG give give.PFV.M.PL
(intended: ‘I did not give him money.’) (Hook 1974:100)

b. main use paise nahin de rah-ā
1SG.NOM 3SG.DAT money NEG give PROG-M.SG
‘I am not giving him money.’

3.3 Negative Imperatives

Negative imperatives provide a classic diagnostic of auxiliary status. With ordinary auxiliaries, negation can intervene between V1 and V2. With *rah-*, this is impossible:

(7) a. bhūl mat/na jānā
forget NEG go-INF
‘Don’t forget!’

b. *bhūl mat/na rahnā
forget NEG stay-INF
(intended: ‘Don’t keep forgetting’)

The unacceptability of (7b) shows that *rah-* cannot occupy the auxiliary slot accessible to negation. By contrast, imperatives with vector verbs like *denā* ‘give’ or *lenā* ‘take’ accept negation, as shown in the examples (8a-b) below. The inability of *rah-* to host negation directly reflects the well-documented grammaticalization tendency whereby forms lose syntactic autonomy and become more tightly bound to the verb (Hopper & Traugott 2003:6-7; Norde 2020).

(8) a. mār mat/na denā.
kill NEG give-INF
‘Don’t kill!’

b. khā mat/na lenā.
eat NEG take-INF
‘Don’t eat!’

3.4 Morphological Opposition with the Habitual

A further piece of evidence is the paradigmatic opposition between habitual *-t-* and progressive *rah-*. Both attach to the same lexical verb, yielding distinct aspectual meanings:

Verb	Habitual (*-t-*)	Progressive (*rah-*)	Gloss
jā- 'go'	jātā hai	jā rahā hai	'goes / is going'
khel- 'play'	kheltā hai	khel rahā hai	'plays / is playing'
pī- 'drink'	pītā hai	pī rahā hai	'drinks / is drinking'

Table 1: Habitual vs. Progressive opposition in Hindi

This opposition illustrates that *rah-* has been reanalyzed as an inflectional competitor of *-t-*. The modern Hindi aspectual system is thus organized around morphological contrasts, in which aspectual values are realized by tightly bound markers rather than by free auxiliaries.

4 *Rah-* and Complex Predicate: From Light Verb to Affix

Within the South Asian complex predicate system, light verbs (often called vector verbs) play a well-defined role. According to Butt (2010:3–6), light verbs exhibit three core properties: (i) they are independently inflected verbs, (ii) they modify the event structure of the main verb (V1) by contributing meanings such as completion, benefaction, suddenness, or inception, and (iii) they form monoclausal predicates together with V1. These diagnostics distinguish light verbs from auxiliaries and from purely lexical verbs.

When tested against these diagnostics, *rah-* fails to qualify as a light verb. First, it does not inflect independently in the way that canonical light verbs do: forms such as *rah-ā* are agreement forms attached to a verbal complex rather than participial forms that retain verbal properties. Second, *rah-* does not introduce additional event-structural nuances such as completion (*-lenā* 'take'), benefaction (*-denā* 'give'), or inception (*-paṛnā* 'fall'). Instead, its sole contribution is the progressive reading. Third, *rah-* does not participate in the monoclausal [V1+V2] complex predicate structure, since its distributional behavior diverges sharply from other V2s. As shown in the diagnostics above (3.1–3.4), *rah-* cannot co-occur with the habitual marker *-t-*, it resists negation when placed directly after V1, and it fails to appear in negative imperatives where ordinary V2s do. In other words, it no longer behaves like a verb.

Taken together, these properties suggest that *rah-* has moved beyond the domain of complex predicates into the domain of bound aspectual morphology. Its role is not to form a syntactic predicate with V1, but rather to serve as a progressive aspect morpheme tightly integrated into the TAM system of Hindi. In this respect, *rah-* exemplifies what Lehmann (1985:7) characterizes as a canonical grammaticalization cline: main verb → auxiliary → TAM affix. *Rah-* originated as the lexical verb *rahnā* 'stay, remain'; it later

functioned in auxiliary-like uses as an aspectual verb; and in present-day Hindi it has grammaticalized further into a bound progressive marker.

This trajectory is also consistent with Hopper and Traugott's (2003:106–109) principles of grammaticalization. The development of *rah-* illustrates decategorialization, in that it has lost morphosyntactic features typical of verbs (e.g. ability to inflect independently, compatibility with negation and habitual markers). At the same time, it also illustrates persistence, as traces of its original lexical meaning 'stay/remain' still survive in certain restricted contexts and in the semantic intuition that progressive aspect involves a sense of "continuation" or "remaining in an action."

5 Refuting the "Perfective *rahā*" Hypothesis

One persistent source of confusion in the analysis of Hindi aspect is the status of *rahā* in progressive constructions. Because *rahā* superficially resembles the perfective participial form, some descriptions have taken the progressive marker to be "*rahā*" rather than *rah-*. This interpretation incorrectly implies that progressive forms incorporate the perfective marker, leading to the paradox that progressive constructions (9) contain a perfective morpheme. This analysis is untenable, since perfective and progressive are aspectually incompatible.

(9) Rām khānā khā rah-ā hai
Ram food eat PROG-M.SG be.PRES.3SG
'Ram is eating food'

Masica (1991:329) notes that forms of *rah-* occur only in shapes that are "formally homonymous with (but have lost the value of) the Perfective." In other words, the *-ā* ending in *rahā* is formally identical to the masculine singular ending seen in perfective participles (*khelā* 'played', *khāyā* 'ate'), but it does not function as a perfective morpheme in progressive contexts. Instead, it is simply the regular masculine singular agreement suffix. This homonymy explains the surface similarity but should not be taken as evidence that progressive *rah-* encodes perfective aspect.

Olphen (1975:287) provides further support for this distinction by identifying the true perfective marker of Hindi as *-y-*, in contrast to the habitual *-t-* and the durative *rah-*. The glide *-y-* appears when the verb stem ends in a vowel (e.g. *khā-* → *khāyā 'ate'*), while consonant-final stems take *-ā* directly (e.g. *khel-* → *khelā 'played'*). Thus, *-ā* in *rahā* is not a perfective morpheme but an agreement ending parallel to the gender/number alternations *rahī* (F.SG) and *rahe* (M.PL).

Following Deo (2006:176), the progressive construction has traditionally been described as consisting of a gerundival main verb followed by the auxiliary *rah-* in its perfective form and a tense auxiliary (*V-ger* + *rah-perf* + *Tns*). While this description correctly captures the morphological shape of the construction, the analysis presented here argues that *rah-* no longer functions as an auxiliary. The distributional diagnostics in 3.1~3.4 demonstrate that *rah-* has undergone grammaticalization into a progressive affix. Thus, forms like *khā rahā hai* 'is eating' are better analyzed not as containing a perfective

auxiliary *rahā*, but as a gerundival stem plus the affixal progressive morpheme *rah-*, followed by a tense auxiliary. The apparent perfective suffix *-ā* is only an agreement marker, not an aspectual exponent.

In sum, the progressive marker in Hindi is *rah-*, not *rahā*. Recognizing this distinction avoids conflating progressive with perfective morphology, and clarifies the role of agreement markers versus aspectual exponents in the TAM system.

6 Conclusion

This paper has re-evaluated the morphological status of *rah-* in Hindi and argued that it should no longer be treated as an auxiliary or vector verb but as a bound progressive morpheme within the TAM system. Through four diagnostics habitual co-occurrence, negation behavior, negative imperatives, and morphological opposition with habitual *-t-*, I demonstrated that *rah-* diverges significantly from the distributional profile of auxiliaries and vector verbs. Unlike canonical light verbs, *rah-* does not inflect independently, does not contribute additional event-structural meaning, and cannot freely combine with habitual or negative markers. Instead, its behavior reflects affix-like morphotactic bonding to the lexical verb.

A further contribution of this study has been to clarify the internal structure of progressive forms such as *khā rahā hai* 'is eating'. Because *rahā* is morphologically identical to the perfective participial ending, many earlier accounts have mistakenly treated progressive *rahā* as containing a perfective morpheme. This interpretation incorrectly implies that progressive constructions (e.g., *Rām khānā khā rahā hai* 'Ram is eating food') contain a perfective marker. Such an analysis is untenable, since perfective and progressive are aspectually incompatible. By distinguishing the true perfective marker *-(y)ā* from the agreement suffix *-ā*, and by identifying *rah-* itself as the progressive morpheme, this study refutes the "perfective *rahā*" hypothesis and resolves a long-standing descriptive ambiguity.

Situating these findings in the broader framework of complex predicate theory shows that *rah-* fails the central tests for light verbs and thus belongs outside the auxiliary system. Its development aligns closely with established grammaticalization pathways: from lexical verb (*rahnā* 'stay, remain'), to auxiliary-like aspectual verb, to inflectional marker of progressive aspect. The trajectory exemplifies both decategorialization (loss of verbal properties) and persistence (residual traces of the original lexical meaning 'stay/remain'), in line with Hopper and Traugott's principles of grammaticalization. From a typological perspective, the Hindi facts reinforce a well-documented universal tendency for verbs of posture or existence ('stay, remain, sit, stand') to evolve into progressive aspect markers.

The findings carry several implications. For descriptive and pedagogical grammars of Hindi, recognizing *rah-* as a progressive morpheme rather than an auxiliary may help explain distributional restrictions that often puzzle learners. For corpus linguistics and annotation, the analysis suggests that *rah-* should be tagged and parsed as a bound aspect marker, not as part of the auxiliary inventory. For comparative and typological research, the case of *rah-* adds to a growing body of evidence that progressive aspect frequently

develops from verbs of posture or existence, thereby enriching databases of grammaticalization pathways. Finally, for theoretical models of complex predicates, Hindi challenges us to refine the boundary between auxiliaries, light verbs, and affixes, showing that transitional categories may crystallize into inflectional morphology more rapidly than previously assumed.

At the same time, this paper represents only a basic study of the progressive marker *rah-*. Future research can extend these findings by examining larger corpora, exploring dialectal and regional variation within Hindi and related Indo-Aryan languages, and tracing in more detail the diachronic shift from lexical *rahnā* to progressive affix. Such work will not only deepen our understanding of Hindi aspect, but also contribute to broader typological and theoretical debates on the nature of grammaticalization and the categorization of TAM morphology.

Acknowledgements

I extend my sincere gratitude to Professor Seungho Nam for his invaluable guidance throughout the development of this study. I also thank the two anonymous reviewers and the audience at FASAL 15 for their insightful comments and discussions, which greatly contributed to the refinement of this work. I am grateful for the opportunity to present at FASAL 15 and to publish this paper.

Appendix A. Transcription Conventions

This paper adopts the Hunterian transliteration system (the official romanization of Hindi used by the Government of India) rather than IPA transcription. Consonants are represented using standard Roman letters and digraphs in accordance with Hunterian conventions (e.g., ch, chh, j, jh, th, dh, sh). Aspirated consonants are indicated orthographically with <h>, rather than with the IPA superscript ʰ. Retroflex and dental consonants are not distinguished in the transliteration, following standard Hunterian practice. Long vowels are marked with macrons (ā, ī, ū).

References

Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Mengistu Amberber, Brett Baker and Mark Harvey, eds., *Complex predicates: Cross-linguistic perspectives on event structure*, 48–78. Cambridge: Cambridge University Press.

Butt, Miriam and Jamshed Rizvi. 2009. Tense and aspect in Urdu. In Werner Abraham and Elisabeth Leiss (eds.), *Layers of aspect*, 43–66. Amsterdam: John Benjamins.

Bybee, Joan, Perkins Revere and Pagliuca, William. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago, IL: University of Chicago Press.

Chakrabarti, Debasri. 2008. *A study of the complex predicates of Hindi*. Doctoral dissertation, Indian Institute of Technology, Bombay.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
Deo, Ashwini. 2006. *Tense and aspect in Indo-Aryan languages: Variation and diachrony*. Doctoral dissertation, Stanford University.
Hook, Peter Edwin. 1973. *The compound verb in Hindi*. Ann Arbor, MI: Center for South and Southeast Asian Studies, University of Michigan.
Hopper, Paul J. and Elizabeth Closs Traugott. 2003. *Grammaticalization*. 2nd edn. Cambridge: Cambridge University Press.
Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* 20. 303–318.
Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
Van Olphen, Herman. 1975. Aspect, tense, and mood in the Hindi verb. *Indo-Iranian Journal* 16(4). 284–301.
Norde, Muriel. 2020. Grammaticalization in morphology. In Rochelle Lieber (ed.), *Oxford research encyclopedia of linguistics*. Oxford: Oxford University Press.

Vedic Sanskrit compounding as a window into *tough*-movement

Davide Mocci, *University of Cagliari & University of Genoa*

Abstract

In the present study, English *tough*-constructions are compared with their analog in Vedic Sanskrit, i.e., the earliest dialectal variety (dating back to ca. 1200-500 BC) of Sanskrit, an old Indo-Aryan language spoken in the Indian subcontinent for over three millennia. It is found out that Vedic Sanskrit *tough*-constructions are codified via a specific type of compound dubbed as 'bahuvrīhi'. Exploiting recent work on Sanskrit bahuvrīhis, according to which their derivation involves A-movement (more precisely, the movement of an overt phrase from within the bahuvrīhi to a bahuvrīhi-external position where they are case-marked), the present study extends this derivation to the bahuvrīhis that codify *tough*-constructions in Vedic Sanskrit. In this way, the conclusion is reached that *tough*-movement can be reduced to an instance of A-movement in Vedic Sanskrit, thereby contrasting with *tough*-movement in English, which has instead been argued to be a variety of A-bar movement (more precisely, the movement of a silent operator to the left-periphery of the infinitive contained in English *tough*-constructions).

1 Background: What is *tough*-movement?

Tough-movement has been a classic topic of investigation in generative grammar: not only are there several different analyses of *tough*-movement that reflect the different stages of development of the generative field, but there is also no uncontroversial definition of *tough*-movement that generative linguists may agree upon. That said, generative linguists do agree on the fact that, regardless of the analysis adopted, *tough*-movement yields a *tough*-construction, exemplified in (1) for English.

(1) These things are easy to reach.

Thus, in *tough*-constructions like (1), a *tough*-type adjective (e.g., *easy*, *difficult*, *tough*, *impossible*) is predicated of a subject (*these things*), which is nonetheless simultaneously interpreted as an internal argument of the infinitive introduced by the *tough*-adjective (e.g., *these things* is interpreted as the theme of the action denoted by *to reach*).[1]

For English, I adopt the so-called 'base-generation approach' to *tough*-movement championed by Chomsky (1977), and sanctioned by Keine & Poole (2017, 297-298) and Salzmann (2023, 240). The fundamental ingredients of this approach may be graphically illustrated as follows:

(2) [These things]$_{i}$ are [$_{AP}$ easy [$_{CP}$ *Op*$_{i}$ [PRO to reach t_{i}]]]

[1] See Hicks (2017) for a historical overview of *tough*-movement and *tough*-constructions.

According to the analysis in (2), *tough*-movement is the movement of a silent operator (*Op*) from an argument position (where it is assigned a thematic role) to the left periphery of an infinitival clause. Thus, in this analysis, the subject of the matrix clause (*these things*) does not undergo movement: only the silent operator does. More precisely, the remark made above, to the effect that the subject of the matrix clause is understood to be theta-marked by the embedded infinitive (i.e., is interpreted as an internal argument of the embedded infinitive) is nothing more than an illusion: what is theta-marked is again the silent operator, the illusion being ultimately traceable to the fact that the subject of the matrix clause and the silent operator are co-referential (as indicated by co-indexing).[2]

While a detailed explanation of the arguments underpinning the analysis in (2) would take us too far afield, in the present contribution I would like to focus on an aspect of this analysis: i.e., that *tough*-movement is a kind of A-bar movement in (2).[3] Thus, like any instance of A-bar movement, *tough*-movement can cross multiple CPs (3a), and is blocked by an intervening *wh*-element such as *why* (3b):[4]

(3) a. [A guy like John]$_i$ is [$_{AP}$ hard [$_{CP}$ *Op*$_i$ [PRO to imagine [$_{CP}$ t_i [any woman believing [$_{CP}$ t_i [she would agree [$_{CP}$ t_i [PRO to marry t_i]]]]]]]]]

b. ??[A guy like John]$_i$ is [$_{AP}$ hard [$_{CP}$ *Op*$_i$ [PRO to imagine [$_{CP}$ t_i [any woman wondering [$_{CP}$ why [she would agree [$_{CP}$ t_i [PRO to marry t_i]]]]]]]]]

In what follows I shall test to what extent this conception of *tough*-movement as A-bar movement can be extended to Vedic Sanskrit, i.e., to the earliest dialectal variety (dating back to ca. 1200-500 BC) of Sanskrit, an old Indo-Aryan language constituting the medium of communication in the Indian subcontinent for over three millennia. In this way, the ultimate goal of the present study is a crosslinguistic exercise: can a linguistic phenomenon like *tough*-constructions be given a unified treatment (or - put another way - be treated as a natural class) in languages, like English and Vedic Sanskrit, that vary greatly for date of attestation and geographical area?

2 *Tough*-movement and Vedic Sanskrit

In Vedic Sanskrit, *tough*-constructions are codified via compounding. To illustrate this point, let us consider the examples in (4)-(5).

[2]For an alternative approach to English *tough*-constructions, whereby *these things* starts out as the complement of *to reach* and subsequently undergoes long-distance movement to the position of subject of *are*, see, among others, Hicks (2009) and Hartman (2012, 84-97).

[3]A-bar movement is the movement of a linguistic expression that terminates in an A-bar position, i.e., a position that does not correspond to any grammatical function (equivalently: a position where no grammatical case is assigned). For ease of exposition, A-bar positions may be identified here with slots in the complementizer field (CP) of the clause, on which see Rizzi (1997).

[4]The examples in (3) are from Hicks (2009, 542). Cf. also Chomsky (1977, 103-104).

(4) *ubhé* ... *[su]-[gā́tave]*
both.NOM.DU.M [easy]-[reach.INF.DAT]

'Both these things are easy to reach.' (*Atharvaveda-Śaunakīya* 6.1.3c)[5]

(5) *kṛdhi* *[su]-[hánā]* ... *vṛtrā́*
make.IMP.AOR.2SG [easy]-[smash.INF.ACC.PL.N] enemy.ACC.PL.N

'Make [our] enemies easy to smash!' (*Ṛgveda* 7.25.5c)[6]

In these examples the *tough*-type element, which in languages like English is an adjective (e.g., *easy*), is indeed a bound morpheme (i.e., a morpheme that cannot occur outside compounding) of unclear grammatical category, spelled out as *su-*, which in itself conveys a wide range of meanings: 'well', 'good', 'easy'.[7] *su-* combines with an infinitive (*gā́tave*; *hánā*) to form a compound: *su-gā́tave* 'easy to reach' (4); *su-hánā* 'easy to smash' (5). The question now arises as to what empirical motivations support the classification of *su-gā́tave* and *su-hánā* as compounds. Indeed, Vedic Sanskrit makes a clear distinction between compounds and non-compound words. Let us consider how.

First, while any Vedic Sanskrit non-clitic word bears the pitch accent outside compounding (except for vocatives and verbs of matrix clauses), only one word is accented inside compounding:[8] cf. the following pair.

(6) a. *amítrāṇāṃ* *sénāḥ*
enemy.GEN.PL.M army.ACC.PL.F

'The army of [our] enemies.' (*Atharvaveda-Śaunakīya* 11.9.3c)

b. *amitra-senā́m*
enemy-army.ACC.SG.F

'The army of [our] enemies.' (*Atharvaveda-Śaunakīya* 3.1.3a)

Second, while compound-external nouns and adjectives are inflected in case, gender, and number in Vedic Sanskrit, they lack a case ending - thereby qualifying as stems from a morphological viewpoint - when they are allocated to the left-hand member of a compound (Lowe 2015b, 72, 93): cf. again (6a), where *amítra-* is marked with a genitive plural ending (*amítrāṇām*) and (6b), which features the stem form *amítra-*, with no case-marking.[9]

[5]References to the *Atharvaveda-Śaunakīya* are based on the edition by Kim (2022).

[6]References to the *Ṛgveda* are based on the edition by van Nooten & Holland (1994).

[7]Etymologically, *su-* parallels Ancient Greek *eu-*, which resurfaces in such English words as *eugenics*.

[8]For an overview of accentuation in Vedic Sanskrit, with special reference to the language of the *Ṛgveda*, see Elizarenkova (1995, 250-259).

[9]It is controversial whether the right-hand member of Sanskrit compounds is also to be considered as a stem. If the answer is in the affirmative, the case ending which shows up on the right-hand edge of Sanskrit compounds - e.g., the accusative-marker *-m* in (6b) - should be taken as appended not to the compound's right-hand member (i.e., *[amitra]-[senā́-m]*), but rather to the entire compound (i.e., *[[amitra]-[senā́]-m]*). For a discussion of the latter analysis, which goes back to the great ancient Indian grammarian Pāṇini (ca. 4th century BC), see Candotti & Pontillo (2019, 31), Mocci (2025, 110-112), and the references quoted therein.

Last, compounds count as one position, just like simple words, for the placement of second position clitics. This means that two sentence-initial nouns or adjectives may precede a second position clitic in Vedic Sanskrit when they combine to form a compound: see (7a), where *asyāḥ* (a second position clitic) follows the compound *dīrghā́yus*, made up of *dīrghá-* 'long' and *ā́yus-* 'life'. Conversely, when they do not form a compound, the clitic must intervene between them, on the understanding that the position following two case-marked nouns or adjectives counts as a third position, hence as an invalid host for the clitic: see (7b), where the clitic *naḥ* is sandwiched between the accusative-marked forms of *dīrghá-* 'long' and *ā́yus-*.[10]

(7) a. *dīrghá-āyur asyā yáḥ pátir/ jī́vāti*
long-life.NOM.SG.M she.GEN.CL wh.NOM.SG.M lord.NOM.SG.M live.SBJV.PRS.3SG
śarádaḥ śatám//
autumn.ACC.PL.F hundred.ACC.SG.N
'(to him) who as her husband will live, long-lived, through a hundred autumns.' (tr. Jamison & Brereton 2014; *Ṛgveda* 10.85.39cd)

b. *dīrghám na ā́yuḥ prati-búdhyamānā/ vayám*
long.ACC.SG.N we.GEN.CL life.ACC.SG.N back-wake.PTCP.PRS.NOM.PL.M we.NOM
túbhyaṃ bali-hṛ́taḥ syāma//
you.DAT tribute-bearing.NOM.PL.M be.OPT.PRS.1PL
'awakening to meet our long lifetime, may we be tribute-bearers to thee.' (tr. Whitney 1962; *Atharvaveda-Śaunakīya* 12.1.62cd)

Let us now apply these diagnostics to test the compoundhood of *su-gā́tave* (4) and *su-hánā* (5). The second diagnostic is unverifiable: although the left-hand member *su-* is not case-marked, the possibility cannot be excluded that *su-* is indeclinable to begin with, which makes it impossible to establish whether *su-* is actually a stem in *su-gā́tave* and *su-hánā*. The first diagnostic is trivially met: both *su-gā́tave* and *su-hánā* bear only one accent, namely on their right-hand member.[11] As for the third diagnostic, it cannot be directly verified, because *su-gā́tave* and *su-hánā* are not attested as co-occurring with second position clitics. Nonetheless, examples like the following make it possible to establish that compounds whose left-hand member is *su-* may occur in sentence-initial position and precede a second position clitic (*naḥ*).

(8) *su-bhāgā́n naḥ devāḥ kṛṇutā*
good-share.ACC.PL.M we.GEN.CL god.VOC.PL.M make.IMPV.PRS.2PL

[10]This is indeed an oversimplification. For an overview of the intricacies related to the positioning of clitics in the left-periphery of the Vedic Sanskrit clause, see Hock (1996), Lowe (2014), and Hale (2018).

[11]One may object that *su-* lacks the pitch accent because it is a clitic-like element and not because it is the left-hand member of a compound. However, clitics may cluster around other clitics in Vedic Sanskrit: see Lowe (2014, 35-38) for relevant examples and analysis. Since *su-gā́tave* and *su-hánā* cannot be interrupted by bona fide clitics, we may safely exclude that *su-* is a clitic in the examples under examination.

su-rátnān
good-treasure.ACC.PL.M
'Do you, O gods, make us have a good share, good treasures.' (*Ṛgveda* 10.78.8a)

It is therefore reasonable to maintain that *su-gā́tave* and *su-hánā*, just like *su-bhāga-* (8), satisfy the third diagnostic. Accordingly, we conclude that *su-gā́tave* and *su-hánā* are indeed genuine compounds. All in all, the takeaway from this section is that *tough*-constructions are codified via compounding in Vedic Sanskrit, in the sense that the *tough*-type adjective and the infinitive form a compound (*su-gā́tave*; *su-hánā*) which in turn combines with a compound-external word that is theta-marked by the compound-internal infinitive: *ubhé su-gā́tave* (4); *su-hánā vṛtrā́* (5).

Now, the codification of *tough*-constructions via compounding is expected to be of great importance for the question as to whether *tough*-movement is A-bar movement in Vedic Sanskrit just like in English. This is because the derivation of Vedic Sanskrit compounding has been independently argued to involve a specific type of syntactic movement. The main goal of the present paper is therefore to exploit the derivation of Vedic Sanskrit compounds as a window into the nature of *tough*-movement in this language. Thus, in the next couple of sections (§3-4) I provide a detailed account of how Vedic Sanskrit compounds - including *su-gā́tave* and *su-hánā* - are derived, which will pave the way for the following reflections on the nature of *tough*-movement as A- vs. A-bar movement (§5).

3 On Vedic Sanskrit bahuvrīhi compounds

According to Whitney (1896, §1287a), compounds like *su-gā́tave* (4) and *su-hánā* (5) are "probably possessives". In the classification of compounds developed within the Indian grammatical tradition, possessives correspond to so-called 'bahuvrīhi' compounds, i.e., exocentric compounds where the relation between the compound members is appositive or subordinate.[12] We assume that Whitney was correct, and categorize *su-gā́tave* and *su-hánā* as bahuvrīhis accordingly.

In a recent analysis of Vedic Sanskrit bahuvrīhis, Mocci (2022, 2025) argued that the derivation of this compound type involves movement. Thus, below we shall review Mocci's analysis with an eye to understanding what the instance of movement involved in the derivation of bahuvrīhis may tell us on the nature of *tough*-movement in Vedic Sanskrit.

To begin with, Lowe (2015b) argued that the members of Sanskrit compounds - including bahuvrīhis - are syntactic units of a special sort. He identified these units with an item theorized within Lexical-Functional Grammar: non-projecting words, i.e., heads which, like

[12]The labels used in the indigenous Indian classification of compounds have been adopted in some influential modern linguistic classifications of compounds, including Bauer (2017), who defines bahuvrīhis as 'compounds which canonically label a part of the whole which the compound denotes' (Bauer, 2017, 65). For an insightful overview of the Indian grammatical classification of compounds, see Radicchi (1985–1988), Candotti & Pontillo (2019). See instead Pontillo (2021), Candotti & Pontillo (2022) for a critical discussion of the differences and similarities between the indigenous Indian analysis of bahuvrīs (with special reference to Pāṇini's grammar) and modern linguistic analyses of this compound class.

clitics, fail to project a complement or specifier. Evidence in favor of the syntactic status of Sanskrit bahuvrīhi members comes from examples like the following, where the left-hand member of the bahuvīhi *kád-artha-* 'with what purpose' is an interrogative pronoun: *kád-* 'what/which'. Here *kád-* takes clausal scope, as reflected in the fact that it affects the scope of the clause headed by *jagmathuḥ* ('[...] you have come to our house') by turning it into a question ('[...] have you come to our house?').

(9) *ádha gmántā uśánā pṛchate vām*
then come.PTCP.AOR.ACC.DU.M Uśanā.NOM.SG.F ask.IND.PRS.3SG you.ACC.DU.CL
kád-arthā naḥ á́ gṛhám á́ jagmathuḥ
what-purpose.NOM.DU.M we.GEN.CL here house.ACC here come.IND.PRF.2DU
parākā́t diváḥ ca gmáḥ ca mártyam//
afar.ABL heaven.ABL.SG.M and earth.ABL.SG.F and mortal.ACC.SG.M
'Then Uśanā asks you both [=Indra and Kutsa] on your coming, "With what purpose (have you come) to our house? You have come from afar, from heaven and earth, to a mortal".' (*Ṛgveda* 10.22.6; tr. Jamison & Brereton 2014)

Since scope-taking elements smaller than words are unknown to Sanskrit, it follows that even a stem like *kád-* should indeed be considered as a (non-projecting) word when it is included inside a compound (see Mocci 2025, 158-160 for relevant discussion).[13]

In translating Lowe's (2015b) conclusion - that Sanskrit bahuvrīhi members are syntactic units of a special sort (non-projecting words) - into the generative framework, Mocci (2025, 168-173) suggested identifying bahuvrīhi members with lexical phrases, i.e., phrases lacking functional layers on top of them. This captures the syntactic behavior of bahuvrīhi members (as reflected in the scope-taking ability of *kád-* in (9)), on account of the fact that lexical phrases, like all phrases, are syntactic units. But it also captures the less-than-syntactic behavior of bahuvrīhi members (as reflected in the impossibility to interrupt the bahuvrīhi *dīrghá-āyus-* 'long-lived' with a clitic in (7a) vs. (7b)), in light of the fact the full-fledged syntactic status depends on functional structure, which is missing in lexical phrases by definition.[14] In this way, *su-gā́tave* (4) and *su-hánā* (5), now interpreted as bahuvrīhi compounds, are analyzable along the lines of (10), on the assumption that *su-* is an Adjective Phrase (AP), which also qualifies as a lexical phrase:

(10) a. [[$_{AP}$ su-] [$_{NP}$ gā́tave]]

[13]Sanskrit would not be isolated in its treatment of compound members as syntactic units. For example, a similar claim has been recently made for Modern Greek by Ntelitheos (2022).

[14]Put another way, nominal syntactic units undergo full-fledged syntactic operations such as movement when they are Determiner Phrases (DP), not when they are Noun Phrases (NP), according to standard assumptions. Similarly, verbal syntactic units undergo syntactic operations such as fronting (e.g., *fix the car, I wonder whether he will* - see Chomsky 1986, 20) when they are light verb phrases (vP), not when they are Verb Phrases (VP). Cf. however Bošković (2015) for an alternative, dynamic system which distinguishes between languages/constructions where DPs and vPs (as opposed to NPs and VPs) are full-fledged syntactic units (i.e., 'phases' in the terminology of this system) and languages/ constructions where NPs and VPs also qualify as full-fledged syntactic units.

b. [[$_{AP}$ su-] [$_{NP}$ hánā]]

Nonetheless, the discovery of an ordering constraint on Sanskrit bahuvrīhis forces us to complicate the simple structures in (10), which merely consist of the merger of an AP with NP. The ordering constraint states that, if the bahuvrīhi's 'external referent' is interpreted as a dependent of the bahuvrīhi member α, α is allocated to the right-hand slot of the bahuvrīhi (Mocci 2022, 2025). For example, (11a) contrasts with (11b) in that the qualifier precedes the qualificand only in (11a), and this contrast correlates with the fact that the bahuvrīhi's external referent (i.e., *ráthena* and *devā́ḥ*) is understood as a dependent (qua possessor) of the qualificand (*áśva-* 'horse') in (11a), but of the qualifier (*pitṛ́-*) 'father' in (11b).[15] The dependency between the external referent and the qualificand or qualifier is emphasized in bold in (11a-b).

(11) a. *vṛṣaṇ-aśvéna* ... *ráthena*
bull-horse.INS.SG.M chariot.INS.SG.M
Qualifier-**Qualificand** Possessor.of.**Qualificand**
'With the chariot whose horses are like bulls.' (*Ṛgveda* 8.20.10ab)

b. *tvát-pitāraḥ* ... *devā́ḥ*
you-father.NOM.PL.M god.NOM.PL.M
Qualificand-**Qualifier** Possessor.of.**Qualifier**
'The gods whose father you are.' (*Taittirīya Saṃhitā* 1.5.10.2)[16]

Thus, the structures in (10) must be complicated to capture the ordering constraint at stake. Following Mocci (2022, 2025), the complication required here is a direct consequence of the syntactic codification of two semantic relations involved in Sanskrit bahuvrīhis. Let us consider how, by focusing on the bahuvrīhis of (11a-b).

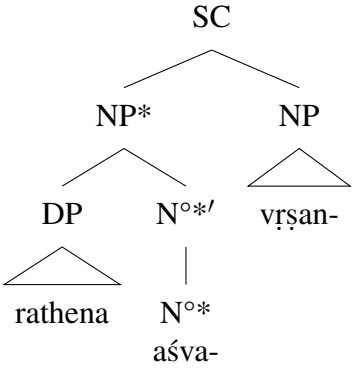
To start with, there is a qualificand-qualifier relation between the bahuvrīhi members (e.g., between *tvád-* 'you' and *pitṛ́-* 'father'), and a possessum-possessor relation between the right-hand member of the bahuvrīhi and the bahuvrīhi's external referent (e.g., between *pitṛ́-* and *devā́ḥ* 'gods'). Following den Dikken (2006, 10-12), the qualificand-qualifier relation is ultimately reducible to a subject-predicate link, which is codified in syntax via a small clause (Moro 2019), notated as SC. Moreover, in keeping with Adger (2013, 113-132), the possessum-possessor relation is codified in syntax via adjunction, which is non-distinguishable from a specifier-head configuration for our purposes (Kayne 1994, 17, Manzini 2017, 237). This yields the structures in (12)-(13) for the bahuvrīhis in (11a-b), respectively: here the possessor is the specifier of the possessum, and the predicate forms a small clause

[15]The possessum-possessor relationship holding between *ráthena* and *áśva-* in (11a), and between *devā́ḥ* and *pitṛ́-* in (11b), respectively parallels the relationship between *ráthasya* and *áśvāḥ* in *ráthasya áśvāḥ* 'the chariot's horses', and between *devā́nām* and *pitā́* in *devā́nām pitā́* 'father of the gods': in these phrases the semantic dependency between the possessor (*ráthasya*; *devā́nām*) and the possessum (*áśvāḥ*; *pitā́*) is overtly (i.e., morphosyntactically) signaled by the genitive case-marking on the possessor.

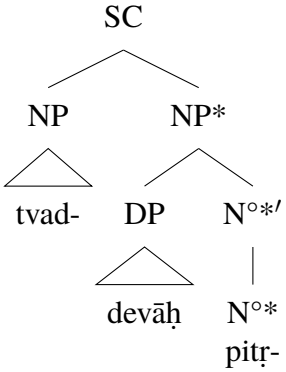
[16]References to the *Taittirīya Saṃhitā* are based on the digital edition by Fushimi (2012), following Weber (1871–1872).

with its subject.[17]

(12)



(13)

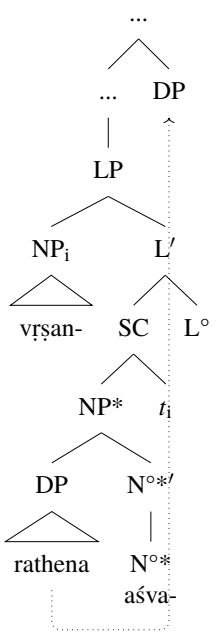


The derivation proceeds along the following lines. First, one of the two components of the small clause moves to the specifier of a functional head, dubbed as L° (i.e., linker) by Mocci (2022, 2025), in order not to violate a general syntactic principle which prevents the direct merger of two maximal projections (e.g., the NP headed by *tvád-* and the NP* headed by *pitŕ-* in (13)).[18] This yields the following phrase markers, where the small clause component raised to Spec-LP is either the predicate ((14), (17)) or the subject ((15), (16)).

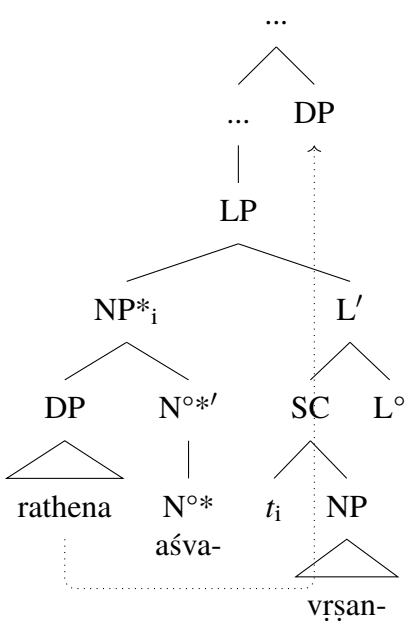
[17]Following standard practice, the predicate is allocated to the right-hand slot of the small clause in these phrase markers. It should be noted that the bahuvrīhi's external referent is represented as an inflected word (*ráthena*, GEN.SG; *devā́ḥ*, NOM.PL) and hence as a DP in these phrase markers, whereas the bahuvrīhi members are represented as stems (e.g., *áśva-*; *pitŕ-*) and hence as NPs, under the assumption that inflection belongs in the functional lexicon. In this perspective, the case marker appended to the right-hand member of bahuvrīhis (*[[vṛṣaṇ]-[aśvéna]]*, with *aśvena* 'horse.INS.SG.M'; *[[tvát]-[pitāraḥ]]*, with *pitāraḥ* 'father.NOM.PL.M') is an optical illusion: the case-marker is indeed appended to the entire bahuvrīhi (*[[vṛṣaṇ]-[aśvá-] + INS.SG.M]*; *[[tvát]-[pitṛ-] + NOM.PL.M]*), which is also to be treated as a stem, albeit a complex one, before it gets case-marked (Candotti & Pontillo 2022). On the empirical and theoretical motivation for this differential morphological and syntactic representation of the bahuvrīhi members vs. the bahuvrīhi's external referent, see Mocci (2025, 201-204).

[18]Different versions of this general principle, which was originally discovered by Moro (2000) under the rubric of 'Dynamic Antisymmetry', have been advanced in the literature. For two recent versions, dubbed as 'Labeling Algorithm' and 'Generalized Dynamic Antisymmetry', see Chomsky (2013) and Moro & Roberts (2024), respectively. For empirical and theoretical arguments in favor of the postulation of L°, see instead Mocci (2025, 194-209).

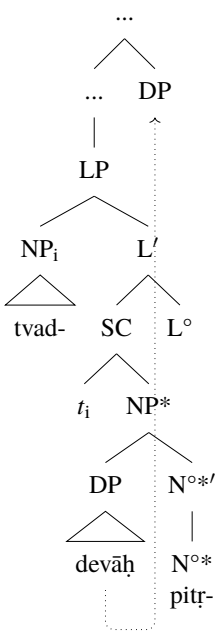
(14)



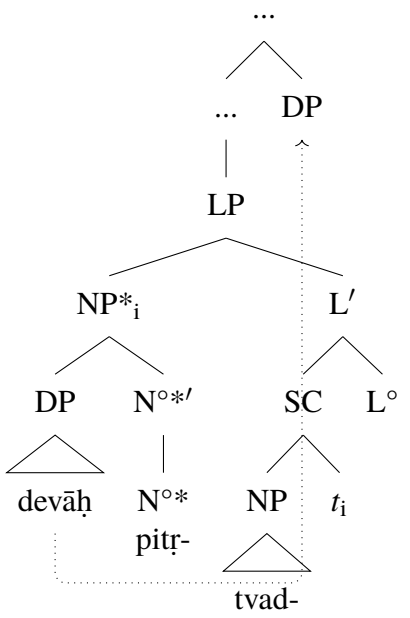
(15)



(16)



(17)



Second, the bahuvrīhi's external referent (i.e., *ráthena* and *devā́ḥ*) moves from Spec-NP* - i.e., the position where it was interpreted as the possessor of the subject, as in (14)-(15), or of the predicate, as in (16)-(17) - to an LP-external position, i.e., to a compound-external position, under the assumption that LP is the upper phrasal boundary of bahuvrīhis. This is indicated by the dotted arrows in (14)-(17).[19]

[19]It should be pointed out that, while the ordering constraint exemplified in (11) determines which bahuvrīhi member (i.e., the qualifier or the qualificand) is allocated to the right-hand slot of the bahuvrīhi, it does not say anything about the positioning of the bahuvrīhi with respect to its external referent (e.g., the positioning of

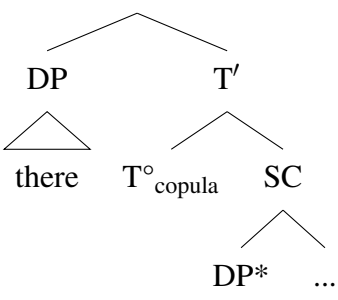
Third, the LP-internal (hence, the bahuvrīhi-internal) material maps into a linear order in keeping with the following general scheme (see Kayne 1994 for a detailed discussion of this scheme): that the specifier of a head (e.g., L°) systematically follows the complement of that head.[20] Accordingly, after the bahuvrīhi's external referent (*ráthena*) has reached an LP-external position, (14)-(15) respectively yield the linear orders *vṛṣaṇ-aśvá-* and **aśva-vṛṣan-*, the latter being unattested. Likewise, after the raising of the bahuvrīhi's external referent (*devā́ḥ*) to an LP-external position, (16)-(17) respectively yield the linear orders *tvát-pitṛ-* and **pitṛ-tvad-*, the latter being - again - unattested.

Last, the unattested orders (**aśva-vṛṣan-*, corresponding to (15), and **pitṛ-tvad-*, corresponding to (17)) are treated as violations of an independently assumed locality principle. To illustrate this point, let us consider the following pair of English copular sentences. While (18a) exemplifies the extraction of linguistic material from an in-situ (i.e., small-clause-internal) subject, (18b) exemplifies the extraction of linguistic material from an in-situ predicate nominal.

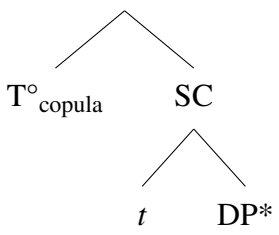
(18) a. [Which wall]$_i$ do you think there were [$_{DP*}$ two pictures of t_i]?
b. [Which riot]$_i$ do you think two pictures of the wall were [$_{DP*}$ the cause of t_i]?

Now, the relevant portions of (18a-b) are given the syntactic representations in (19)-(20), respectively, in Moro's (1997) classic work on predication. Crucially, these phrase markers show: (i) that the extraction from an in-situ subject is licensed by a local relation between a predicative head (the D° *there*) and the head (T°) which locally c-commands that subject (DP*), as depicted in (19); that the extraction from an in-situ predicate is licensed by a head (T°) locally c-commanding that predicate, as depicted in (20).

(19)



(20)



vṛṣaṇ-aśvéna with respect to *ráthena* in (11a)). In fact, a bahuvrīhi may either precede or follow its external referent. Here I shall not be concerned with the positioning of the bahuvrīhi's external referent; I will limit myself to representing its landing site to the right of the bahuvrīhi itself, as indicated in (14)-(17).

[20]This, of course, does not mean that any specifier of any head X° precedes any complement of X° in Sanskrit: syntactic movement, ultimately related to the expression of scope-discourse semantic values, may disrupt the basic Specifier-Complement order. Nonetheless, this disturbing factor may be disregarded for our purposes, inasmuch as there is no empty slot within LP that may license further instances of movement than those already represented in (14)-(17).

Regardless of the exact formulation of the locality conditions that capture the extraction from the subject and predicate nominal,[21] the following considerations can be made: (i) *vṛṣaṇ-aśvá-* is well-formed because it is the spell-out of a phrase marker (i.e., (14)) which corresponds in all relevant respects to (19), i.e., the phrase marker for the well-formed extraction from an in-situ subject; (ii) *tvát-pitṛ-* is well-formed because it corresponds in all relevant respects to (20), i.e., the phrase marker for the ill-formed extraction from an in-situ predicate. **aśva-vṛṣan-* and **pitṛ-tvad-* are instead ill-formed because they are the spell-out of phrase markers where the extraction takes place from a non-in-situ subject (15) or a non-in-situ predicate nominal (17), which is in violation of locality.[22]

All in all, the derivation of Sanskrit bahuvrīhis involves the movement of the bahuvrīhi's external referent from the phrase projected by a bahuvrīhi member, which fulfils the function of either subject or predicate. In the next section I shall extend this derivation to Vedic Sanskrit *tough*-constructions, which are codified as bahuvrīhis in this language.

4 The derivation of Vedic Sanskrit *tough*-constructions

Let us go back to the Vedic Sanskrit *tough*-constructions in (4)-(5), repeated below as (21)-(22). Since *su-gā́tave* and *su-hánā* can be categorized as bahuvrīhis for the reasons specified in the preceding section, it is reasonable to treat (21)-(22) along exactly parallel lines to (11a-b). Accordingly, I take *ubhé* and *vṛtrā́* as the external referents of *su-gā́tave* and *su-hánā*, respectively.

(21) *ubhé* ... *[su]-[gā́tave]*
both.NOM.DU.M [easy]-[reach.INF.DAT]
'Both these things are easy to reach.' (*Atharvaveda-Śaunakīya* 6.1.3c)

(22) *kṛdhi* *[su]-[hánā]* ... *vṛtrā́*
make.IMP.AOR.2SG [easy]-[smash.INF.ACC.PL.N] enemy.ACC.PL.N
'Make [our] enemies easy to smash!' (*Ṛgveda* 7.25.5c)

The parallelism with (11a-b) also requires that a subject-predicate link is assumed between the bahuvrīhi members, whereby *su-* 'easy' is the predicate (in both examples) and the right-hand members *gā́tave* 'reach.INF.DAT' and *hánā* 'smash.INF.ACC.PL.N' the corresponding subjects. This is reminiscent of Salzmann (2023, 233-237), who argues that *tough*-adjectives such as *easy* take an infinitival complement such as *to reach* as their external argument in such English *tough*-constructions as *these things are easy to reach*: considering that the relationship of adjectives (qualifiers) with their external arguments (qualificands) may be reduced to a subject-predicate link at the relevant level of representation (den Dikken

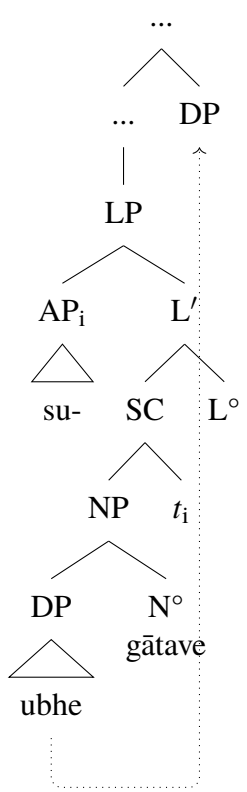
[21] See Moro (1997, 2017a,b) for a formulation based solely on selection and c-command, along the lines of the Subjacency Condition à la Cinque (1990). See instead den Dikken (2006) for a phase-based formulation. Cf. also Mocci (2025, 38-96) for a critical comparison of Moro's and den Dikken's formulations.

[22] Space constraints prevent me from discussing why locality is violated in these cases. See however Mocci (2025) for a detailed discussion of this point.

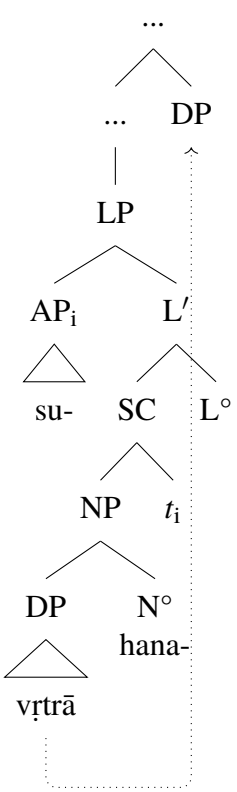
2006, 10-12), it is not unreasonable to assume that, besides being an external argument of *easy*, *to reach* also serves as its subject at the relevant level of representation.

By pursuing the parallelism with (11a-b) further, the following phrase markers are obtained for (21)-(22), where the external referents are represented as internal arguments (bearing the theme theta-role) of the infinitives *gā́tave* and *hánā*, and hence as part of the phrase projected by one of the bahuvrīhi member, in line with (16)-(19).

(23)



(24)



That is, the bahuvrīhi-internal predicate *su-* moves to Spec-LP in both (23)-(24) in order not to violate the general ban against the direct merger of two maximal projections (i.e., the SC-internal occurrences of NP and AP). Next, the bahuvrīhi's external referent (*ubhé* 'both'; *vṛtrā́* 'enemies') moves outside the bahuvrīhi, to an LP-external position, as indicated by the arrow, in both (23) and (24). Such movement does not violate locality because it is an instance of extraction from an in-situ subject, where L° (i.e., the head locally c-commanding this subject) is in a local relation with the predicate of the in-situ subject itself (i.e., with the A° *su-*). In other words, both (23)-(24) run parallel to (19), i.e., to the well-formed extraction from the in-situ subject of *there*-sentences (e.g., *which wall do you think there were two pictures of?*). In this way, the derivation of the bahuvrīhis *su-gā́tave* and *su-hánā* completes successfully, yielding the *tough*-constructions in (21)-(22).

In the next section I shall assess the consequences that the derivation of Vedic Sanskrit *tough*-constructions championed here has for the nature of *tough*-movement.

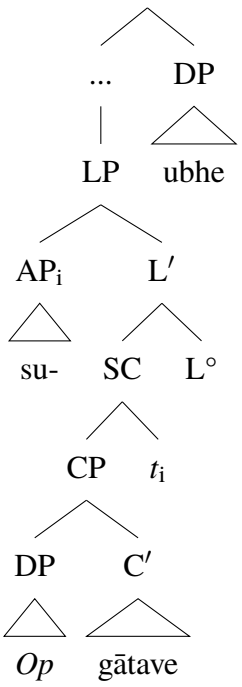
5 Consequences for the nature of *tough*-movement

There are at least two points that should be made with regard to the syntactic representations in (23)-(24). First, the infinitive (*su-gā́tave*; *su-hánā*) is represented as a noun (N°) there. This is because of a well-established ban on the right-hand member of Sanskrit compounds being a full-fledged verb, i.e., an element endowed with tense-aspect morphology (see Lowe 2015a, 269-273 on this ban). Related to this categorization of the infinitive in (23)-(24) as a noun is the second point: namely that no silent operator is housed in the left periphery of the infinitive in these phrase markers, unlike in the representation of English *tough*-constructions (2), repeated below as (25).

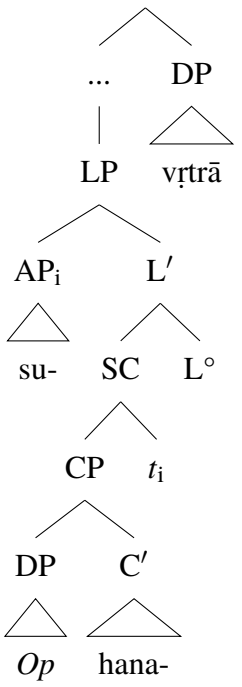
(25) [These things]$_{\text{i}}$ are [$_{\text{AP}}$ easy [$_{\text{CP}}$ *Op*$_{\text{i}}$ [PRO to reach t_{i}]]]

That is, while *tough*-movement in English is handled as the movement of a silent operator to the left-periphery of the embedded clause projected by the infinitive (see above, §1), the postulation of a null operator that moves to the left-periphery (Spec-CP) of the infinitives *gā́tave* and *hánā* would simply be an unneccesary complication. Let us consider why. The movement of the null operator to Spec-CP preempts the movement of the bahuvrīhi's external referent (*ubhé*; *vṛtrā́*) from within LP to an LP-external position: the bahuvrīhi's external referent is simply base-generated in a bahuvrīhi-external position (just as *these things* is base-generated in the matrix clause in (25)), along the lines of (26)-(27).

(26)



(27)



Now, crucially, the null operator does not cross the CP boundary projected by the infinitive in (26)-(27). This implies that no movement of linguistic material outside the phrase pro-

jected by a bahuvrīhi member (i.e., by the infinitives *gā́tave* and *hánā*) takes place in this case. However, the ordering constraint on Sanskrit bahuvrīhis (according to which if the bahuvrīhi's external referent is interpreted as a dependent of the bahuvrīhi member α, α is allocated to the right-hand slot of the bahuvrīhi) was reduced to a mere consequence of the locality conditions on the extraction of the bahuvrīhi's external referent from within the phrase projected by a bahuvrīhi member to an LP-external position (see above, §3).

Therefore, the movement of the null operator to the left-periphery of the infinitive forces us to state the linear order of the bahuvrīhis *su-gā́tave* and *su-hánā* as an axiom - not being derivable from independently assumed principles of grammar - thereby setting *su-gā́tave* and *su-hánā* aside from all other Sanskrit bahuvrīhis. Rather than accepting this treatment of *su-gā́tave* and *su-hánā* as an exceptional type of Sanskrit bahuvrīhi, I shall stick to the representations in (23)-(24), which replace the movement of a silent operator to the left-periphery of the infinitive (*gā́tave* and *hánā*) with the movement of the bahuvrīhi's external referent (*ubhé*; *vṛtrā́*) to an LP-external (i.e., bahuvrīhi-external) position. The main motivation for adopting the representations in (23)-(24) is that they make it possible to treat the linear order of virtually all Sanskrit bahuvrīhis - including Vedic Sanskrit *su-gā́tave* and *su-hánā* - not as an axiom but as a theorem, being derivable from locality (i.e., from an independently assumed principle of grammar).

All in all, the codification of *tough*-constructions via compounding in Vedic Sanskrit suggests treating *tough*-movement not as the movement of a silent operator to the left-periphery of an infinitive, but rather as the movement of an overt phrase (which I have dubbed as 'the bahuvrīhi's external referent') from within the bahuvrīhi-internal phrase projected by the infinitive to a position outside the bahuvrīhi; no left-periphery needs to be assumed for the bahuvrīhi-internal infinitive, which can be simply categorized as a noun. Since the bahuvrīhi-external position targeted by the movement of the bahuvrīhi's external referent is a position where case is assigned (or checked, depending on what theoretical approach to case marking is taken),[23] *tough*-movement in fact qualifies as a type of A-movement in Vedic Sanskrit. Accordingly, Vedic Sanskrit contrasts with English, where *tough*-movement has instead been argued to be a variety of A-bar movement, whereby a silent operator targets the left-periphery of the embedded infinitive of English *tough*-constructions. The present study is thus in line with Tayalati & Danckaert (2020), who brought data from Modern Standard Arabic in support of the conclusion that *tough*-constructions may not actually form a natural class crosslinguistically.

6 Concluding remarks

Tough-constructions are codified via bahuvrīhi compounding in Vedic Sanskrit. Since the derivation of Sanskrit bahuvrīhis has been independently established to involve the A-movement of the bahuvrīhi's external referent to a bahuvrīhi-external position, in the present contribution this derivation has been extended to the bahuvrīhis that codify *tough*-constructions

[23] See Chomsky (2000, 123-124) and Chomsky (2001, 16-17) for an approach to case-marking based on the notion of checking. See instead Chomsky (1993) for a government-based approach to case-marking.

in Vedic Sanskrit. In this way, *tough*-movement has been identified with the A-movement of the bahuvrīhi's external referent. The main takeaway from this study is therefore that *tough*-movement is A-movement in Vedic Sanskrit, which accordingly contrasts with English, where *tough*-movement has been identified with the A-bar movement of a silent operator to the left-periphery of the embedded infinitive.

References

Adger, David. 2013. *A syntax of substance*. Cambridge, MA: MIT Press.

Bauer, Laurie. 2017. *Compounds and compounding*. Cambridge: Cambridge University Press.

Bošković, Željko. 2015. From the Complex NP Constraint to everything: On deep extractions across categories. *The Linguistic Review* 32(4). 603–669.

Candotti, Maria Piera & Tiziana Pontillo. 2019. Lexical subordination in derivation and compounding. *Studi e Saggi Linguistici* LVIII(2). 155–175.

Candotti, Maria Piera & Tiziana Pontillo. 2022. Dispensing with zero in the analysis of sanskrit bahuvrīhi: Resurfacing, testing and assessing Pāṇini's model. *Journal of South Asian Linguistics* 12(1). 1–22.

Chomsky, Noam. 1977. On wh-movement. In Peter W. Culicover, Thomas Wasow & Adrian Akmajian (eds.), *Formal syntax*, 71–133. New York: Academic Press.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1993. *Lectures on Government and Binding: The Pisa lectures*. Berlin: de Gruyter Mouton 7th edn.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels, Juan Uriagereka & Samuel Jay Keyser (eds.), *Step by step: Essays on minimalist syntax in honor of howard lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. *Beyond explanatory adequacy* (MIT Occasional Papers in Linguistics 20). Cambridge, MA: Department of Linguistics and Philosophy, MIT.

Chomsky, Noam. 2013. Problems of projection. *Lingua* 130. 33–49.

Cinque, Guglielmo. 1990. *Types of ā-dependencies*. Cambridge, MA: MIT Press.

den Dikken, Marcel. 2006. *Relators and linkers: The syntax of predication, predicate inversion, and copulas*. Cambridge, MA: MIT Press.

Elizarenkova, Tatyana J. 1995. *Language and style of the Vedic Ṛṣis*. Albany, NY: State University of New York Press.

Fushimi, Makoto. 2012. Taittirīya-Saṃhitā. Electronic text input by Makoto Fushimi; TITUS version by Jost Gippert, last edited 21.04.2012. http://titus.uni-frankfurt.de/texte/etcs/ind/aind/ved/yvs/ts/ts.htm.

Hale, Mark. 2018. The syntax of indo-iranian. In Jared Klein, Brian Joseph & Matthias Fritz (eds.), *Handbook of Comparative and Historical Indo-European Linguistics*, 1924–1942. Berlin: de Gruyter Mouton.

Hartman, Jeremy. 2012. *Varieties of clausal complementation*: Massachussets Institute of Technology dissertation.

Hicks, Glyn. 2009. *Tough*-constructions and their derivation. *Linguistic Inquiry* 40(4). 535–566.

Hicks, Glyn. 2017. *Tough*-movement. In Martin Everaert & Henk van Riemsdijk (eds.), *The wiley blackwell companion to syntax, second edition*, 1–27. John Wiley & Sons, Ltd. doi:https://doi.org/10.1002/9781118358733.wbsyncom035.

Hock, Hans Henrich. 1996. Who's on First? toward a Prosodic Account of P2 Clitics. In Aaron L. Halpern & Arnold M. Zwicky (eds.), *Approaching Second: Second Position Clitics and Related Phenomena*, 199–270. Stanford, CA: CSLI Publications.

Jamison, Stephanie W. & Joel P. Brereton. 2014. *The Rigveda: The earliest religious poetry of India. an English translation*. Oxford/New York: Oxford University Press.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Keine, Stefan & Ethan Poole. 2017. Intervention in *tough*-constructions revisited. *The Linguistic Review* 34(2). 295–329.

Kim, Jeong-Soo. 2022. Atharvavedasaṃhitā der śaunakaśākhā: Eine neue Edition unter besonderer Berücksichtigung der Parallelstellen der Paippalādasaṃhitā. https://opus.bibliothek.uni-wuerzburg.de/frontdoor/index/index/docId/27703.

Lowe, John J. 2014. Accented clitics in the *ṛgveda*. *Transactions of the Philological Society* 112(1). 5–43.

Lowe, John J. 2015a. *Participles in Rigvedic Sanskrit: The syntax and semantics of adjectival verb forms*. Oxford: Oxford University Press.

Lowe, John J. 2015b. the syntax of Sanskrit compounds. *Language* 91(3). e71–e115.

Manzini, Maria Rita. 2017. Passive, smuggling and the *by*-phrase. In Nicholas LaCara, Keir Moulton & Anne-Michelle Tessier (eds.), *A Schrift to Fest Kyle Johnson*, 233–244. Linguistics Open Access Publications.

Mocci, Davide. 2022. The possessive relation in Sanskrit bahuvrīhi compounds: Ellipsis or movement? *Journal of South Asian Linguistics* 12(2). 1–30.

Mocci, Davide. 2025. *A configurational approach to Sanskrit bahuvrihi compounds* (RGG Companion Series). Lingbuzz Press. https://lingbuzz.com/j/rgg/monographs/mocci/.

Moro, Andrea. 1997. *The raising of predicates: Predicative noun phrases and the theory of clause structure*. Cambridge: Cambridge University Press.

Moro, Andrea. 2000. *Dynamic Antisymmetry*. Cambridge, MA: MIT Press.

Moro, Andrea. 2017a. Copular sentences. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax. Second Edition*, 1210–1232. Oxford: John Wiley & Sons Ltd.

Moro, Andrea. 2017b. Existential sentences. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax. Second Edition*, 1483–1508. Oxford: John Wiley & Sons Ltd.

Moro, Andrea. 2019. The geometry of predication: A configurational derivation of the defining property of clause structure. *Philosophical Transactions of the Royal Society B* 375(1791). 1–7.

Moro, Andrea & Ian Roberts. 2024. The duality of syntax: Unstable structures, labelling and linearisation. *Natural Language and Linguistic Theory* 42. 609–631.

van Nooten, Barend A. & Gary B. Holland. 1994. *Rig Veda: A Metrically Restored Text*. Cambridge, Mass.: Harvard University Press.

Ntelitheos, Dimitrios. 2022. Compounding in greek as phrasal syntax. *Languages* 7(151). 1–22.

Pontillo, Tiziana. 2021. Did the Sanskrit model bring "true enlightenment to European scholars" when they analysed and classified the bahuvrīhi compounds? *Studi Classici e Orientali* 67. 497–514.

Radicchi, Anna. 1985–1988. *La teoria pāṇiniana dei samāsa secondo l'interpretazione delle scuole grammaticali indiane dal quinto all'ottavo secolo d.C.* Firenze: Elite. Two volumes.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar*, 281–337. Dordrecht: Kluwer.

Salzmann, Martin. 2023. Experiencer intervention in English *tough* movement: Evidence from extraction of the *tough* adjective against syntactic- and semantic-intervention accounts. *Syntax* 26(2). 223–249.

Tayalati, Fayssal & Lieven Danckaert. 2020. The syntax and semantics of Modern Standard Arabic resumptive *tough*-constructions. *Folia Linguistica* 54(1). 197–238.

Weber, Albrecht. 1871–1872. *Die Taittirîya-Saṃhitâ*. Leipzig: Brockhaus. Two volumes.

Whitney, William Dwight. 1896. *A Sanskrit grammar: Including both the Classical language, and the older dialects, of Veda and Brahmana*. Leipzig: Breitkopf & Härtel. Third edition.

Whitney, William Dwight. 1962. *Atharva-Veda Saṃhitā: Translated with a critical and exegetical commentary*. New Delhi: Motilal Banarsidass. Volume 2 (Books VIII to XIX).

Not on the same page: negotiating not-at-issue content with *bərə* (Marathi)

Shaunak Phadnis, *UMass Amherst*

Abstract

This case study of Marathi discourse particle *bərə* in medial position makes two contributions: 1. *bərə* is used to *question* (how come p?) or *reject* (as if p) the prejacent p. Importantly, p is necessarily part of not-at-issue content of the preceding context update proposal 2. Hence, not-at-issue updates should be represented as proposals too. In advancing these points, a unified account of the two readings is offered. Formally, the proposalhood of not-at-issue updates is represented in the *waiting room* model (Biezma & Rawlins 2017) of discourse. The resulting outlook therefore shows that being at-issue $\neq$ being a proposal. Instead, at-issueness should only be relativized to the current QUD.

1 Introduction

It is a standard assumption that sentence meaning is made up of more than just what is *said*. It has components peripheral to what is said (the 'main point' of the utterances) such as presuppositions, implicatures, entailments etc. Discourse is sensitive to this in a particular way. All else being equal, normally, we object to or negotiate what is said/the main point (hence, labeled *at-issue content* in Potts 2005) and let the peripheral (not-at-issue, henceforth NAI) content slide.[1] Following current literature, I take at-issue content as content that addresses the current question under discussion (Beaver et al. 2017). Since we negotiate on content of utterances, utterances in a discourse are taken as *proposals* to update the context (Stalnaker 1978). So proposals (at-issue content) are evaluated but peripheral (NAI) content is taken as not intended for negotiation. A well-established empirical observation supporting this view regarding the at-issue/NAI distinction is that at-issue content can be directly (anaphorically) denied whereas, NAI content cannot be denied directly. In (1), the nominal appositive, taken to be NAI, cannot be felicitously challenged by direct denial as in B_2 whereas, the asserted component can be challenged in B_1.

(1) A: John, the blues guitarist, is visiting Mary tomorrow.
B_1: That's not true/No – John is not visiting Mary.
B_2: #That's not true/#No – John is not a blues guitarist.
B_3: Wait. This is peripheral to your point but: John isn't a blues guitarist.

This is not to say that NAI content cannot be challenged at all. In (1), B_3 is a response that challenges the content of the nominal appositive. But note that doing so halts the discourse. This is very clear in cases of presupposition failures where interlocutors have to backtrack and update the context non-monotonically as the discourse crashes. Observations such as

[1] This is of course simplifying given cases of presupposition failure.

these regarding anaphoric deniability[2] of at-issue content have repercussions in discourse pragmatics that models how the common ground[3] (CG) is updated. As a specific instance of this, in the influential and successful 'table model' of Farkas & Bruce (2010), proposals have a special slot (the table) where they are evaluated but NAI content lacks such a slot. In turn, literature that models context update phenomena (such as Murray 2014, Anderbois et al. 2015, Rett 2021) has thus treated NAI updates to be *forced* (or directly added) into the common ground as opposed to at-issue updates which go through negotiation stage.

The present paper brings evidence from Marathi discourse particle *bərə*[4] and shows that it exclusively targets NAI updates. This means that the strategy to prevent addition of NAI content into CG is grammaticalized. Such conventionalized means of challenging NAI updates, then, is also a monotonic process exactly like challenging at-issue content with response particles such as *No*. Therefore, as I show in this paper, NAI content is also negotiated in a discourse, and it needs to be represented as such i.e., at the proposal stage. Note that the claim here is not that the at-issue/not-at-issue distinction is irrelevant. The empirically novel fact is the existence of a dedicated lexical device whose argument is part of the NAI content of the previous proposal. The emerging outlook is that at-issueness should not be equated with proposalhood. Instead, both at-issue and NAI updates are proposals.

1.1 Setup

The particle *bərə* occurs optionally clause-medially in declaratives as well as in clause-final position across declaratives, imperatives and wh-interrogatives. Deo (2025) provides the first account of clause-final *bərə*. See (2) illustrating this occurrence of *bərə*.

(2) CONTEXT: *Anu needs Bilal to stay at home because the plumber will be coming and someone needs to be in the house to meet him.*

Anu: mi aata office-la dza-te ahe *bərə*
1SG.F now office-DAT go-IMPF.F.SG be.PRES BARA
'I am going to the office now (alright?/ok?)' (Deo, 2025, ex. 4)

In (2), Anu communicates to Bilal new information that she deems beneficial/important for Bilal to take into consideration. The presence of *bərə* in (2) conveys Anu's expectation that Bilal commits to the prejacent[5] respectively and act in accordance with it. Deo's account, relying on Condoravdi & Lauer's (2012) framework, explains interpretive effects arising out of interaction of *bərə* with the clause-types it occurs in. Under that account, presence of *bərə* has an *advisory* effect: informally, the speaker indicates a preference for the addressee to take commitment to the prejacent of *bərə* in view of some salient goal.

[2]Cf. Krifka (2013), Roelofsen & Farkas (2015), and Snider (2017) for instance on anaphoricity of response particles.

[3]What we jointly accept to be the case, Stalnaker (2002).

[4]Historically, an adjective meaning "well/good".

[5]The sentence without *bərə*.

In this paper, I provide a first account of clause-medial *bərə*. Unlike its clause-final counterpart, clause-medial *bərə* is only licensed in declaratives and it is distinct from clause-final *bərə* in two ways: (i) it shows agreement; (ii) clause-medial *bərə* contributes a *rejecting/questioning* effect. The basic data of interest is the following. Compare the paradigm in (3) and (4) which have minimally distinct contexts resulting in very different interpretations. Note that the only difference is whether the preceding update proposal is about predictions about the future (easily taken as not settled) in (3) where we get *rejection* vs. events easily taken to be settled in (4) where we get *questioning* interpretation.

(3) **Rejecting *bərə***

CONTEXT: *A is telling B about an upcoming concert in the city. This artist only plays at big arenas. B knows that Anu doesn't like going to large concert venues. A only knows that Anu likes this type of music.*

A: There is a Beyonce show happening in June. I should tell Anu.
(Assumption: Anu will be interested in going to this concert)

B: anu bər-i itk-ya moʈʰ-ya concert-la dza-il[6]
anu.F.SG BARA-F.SG DEM.much.PROX-OBL big-OBL concert-DAT go-**FUT**
'As if Anu will go for such a big concert'…*#but I think she'll go*

(4) **Questioning *bərə***

CONTEXT: *A is telling B about a recent Beyonce concert A went to. This artist only plays at big arenas. B knows that Anu doesn't like going to large concert venues. B asks A, "how was the concert?"*

A: It was great! Anu was singing along to every song!
(Defeasible entailment: Anu went to the big concert[7])

B: anu bər-i itk-ya moʈʰ-ya concert-la ge-l-i
anu.F.SG BARA-F.SG DEM.much.PROX-OBL big-OBL concert-DAT go-**PERF**-F.SG
'Anu went to such a big concert!? (what's up with that?)'..*#but I think she didn't*

In (3), speaker B is construed as dismissive and incredulously rejecting the prejacent, which I term as the *rejection* reading. Here, the events in discussion cannot be easily taken to be settled by B (its to be seen whether Anu will be excited by the concert news). Whereas in (4), the overall effect is of incredulously questioning the prejacent (*how come p?*). This is a context involving report of past events that can be easily taken to be settled, especially since A is giving a first-hand report of events. I term this the *questioning* reading. To see the two effects clearly, observe that in (3), the rejection reading, a continuation such as *but I think she'll go* is infelicitous. Whereas, in the questioning case, (4), the speaker appears as

[5]Marathi has two ways to mark future: (i) *-il* regular(FUT) (ii) *-ɳaar* prospective (PROSP). Negated futures require prospective morphology. Some speakers accept *bərə* with PROSP – I assume this is dialectal variation.

[6]Entailments are generally considered to be at-issue but note that in (4), the entailment does not address the QUD (*how was the concert?*) therefore it is NAI. The present accounts adopts this notion of at-issueness (see section 2.2). In turn, what is at-issue is predicted to be not closed under entailment (thanks to María Biezma for raising this). A detailed discussion is beyond the scope of this paper as it would take us too far from the main concerns.

accepting (but questioning) the prejacent since a continuation such as *but I think she didn't* is infelicitous.

The empirically and theoretically important observations regarding the behavior of *bərə* are as follows. When the prejacent can be easily taken to be settled, we get acceptance, although with expression of unexpectedness. When the prejacent cannot be easily taken to be settled, we get rejection. These are two contradictory moves. In both cases, the prejacent is an NAI inference obtained from the preceding proposal. The function of a *bərə* response is to shift attention to and to make at-issue what was not-at-issue previously.[8]

It might be tempting to think that the two (very distinct) readings are an instance of homophony with accidental similarities. However, I argue that they have the same underlying meaning: *bərə* conveys that the speaker finds the (NAI) prejacent highly unexpected. Interaction of this core meaning with different discourse conditions can explain the divergent resulting effects.

The paper is organized as follows. In section 2, I show that both readings of *bərə* have the same felicity requirements: (i) Counter-expectations (section 2.1) and (ii) NAIness of the prejacent (section 2.2). I argue, in section 2.1 that *bərə* responses express that the prejacent is highly unexpected. Moreover, there has to be an epistemic conflict for the speaker. In section 2.2, I show that both types of *bərə* responses can only target NAI content of the preceding context update proposal that is unexpected. This motivates pursuing a unified approach to the analysis. In section 3, I lay out the *waiting room* formalism (Biezma & Rawlins 2017) adopted here and show in section 3.2 how to represent NAI updates as proposals. In section 4, I provide the formal entry for *bərə* and derive the two readings based on interaction between speaker's epistemic stance and considerations on relative epistemic authority of interlocutors. In section 5, I discuss future prospects of the model presented here in explaining other related phenomena that have not received dynamic treatment. Finally, Section 6 concludes the paper.

2 Felicity conditions on *bərə*

Two discourse conditions are necessary for *bərə* responses to be licensed. The speaker must have prior expectations (as in what is compatible with the doxastic alternatives of the speaker) that are in conflict with the target of *bərə* (i.e., the prejacent). Next, the prejacent must be an NAI inference triggered by the preceding context update proposal. I first show in section 2.1 that *bərə* responses express that the prejacent is highly unexpected. After this, I show that contexts without an explicit epistemic conflict do not license *bərə* responses. In section 2.2, I define the relevant notion of at-issueness and show that both *bərə* responses cannot felicitously target an at-issue assertion.

[8] Additionally, while operating at the level of discourse dynamics, *bərə* is subject to agreement. This is beyond the scope of the paper, but I note that *bərə* is derived from an inflectional adjective. A number of Marathi adverbs derived from inflectional adjectives exhibit this property.

2.1 Counter-expectations

Both *rejecting* and *questioning bərə* express that the prejacent is highly unexpected for the speaker. We see this below in (5) as a continuation such as *'...but this is expected'* is infelicitous. A comparable approximation of infelicity of the continuations in (5) is the contrast in (6) with *I find it hard to believe…* in English.

(5) a. anu bər-i gov-ya-la dza-il
anu.F.SG BARA-F.SG goa-OBL-DAT go-FUT
'As if Anu will go to Goa, …#*but this is expected*' (*rejection* reading)

b. anu bər-i gov-ya-la ge-l-i
anu.F.SG BARA-F.SG goa-OBL-DAT go-PST-F.SG
'Anu went to Goa?! (what's up with that?), …#*but this was expected'* (*questioning* reading)

(6) a. I find it hard to believe that Anu will go to Goa ... #I expect that she would

b. I find it hard to believe that Anu went to Goa ... #I expected that she would

Thus, in a sense *bərə* responses are comparable to mirative expressions (albeit with a negative affect of incredulity) which also operate on speaker's expectations. Note that a *bərə* utterance does not express simple surprise. The prejacent is necessarily incompatible with the speaker's doxastic alternatives, i.e., a speaker of a *bərə* utterance is biased towards $\neg p$. This is not necessary for simple surprise, e.g., one may be surprised by the first snow of winter but this is not necessarily incompatible with one's doxastic alternatives. Thus, *bərə* responses require epistemic conflict with what the speaker learns in a discourse.

2.1.1 Rejecting *bərə* and counterexpectations

Rejecting *bərə* responses are incredulous rejections. Part of what it means to perform rejection involves being opinionated about its target. A context where the speaker is epistemically neutral towards the prejacent does not license rejecting *bərə* responses. In the context below, the speaker lacks an opinion towards the prejacent as they do not know Anu's preferences regarding large concerts. In such a case, there is no reason for an epistemic conflict to arise and thus, the speaker cannot felicitously include *bərə* to perform a rejection.

(7) **Sp neutral toward p, infers that Ad believes that p**
CONTEXT: *A is telling B about an upcoming Beyonce concert. This artist only plays at big arenas. B knows Anu fairly well but doesn't know her preferences regarding live music venues.*

A: Beyonce is playing a show here in June. I should tell Anu.
(Assumption: Anu will be interested in going to this concert)

B: #anu bər-i itk-ya mo[ʰ]-ya concert-la dza-il
anu.F.SG BARA-F.SG DEM.much.PROX-OBL big-OBL concert-DAT go-FUT
#'As if Anu will go for such a big concert'

Similarly, in a context where speaker's expectations are actually aligned with the relevant inference (in this case, the assumption), *bərə* responses are not licensed. In (8), the speaker has no reason to find the addressee's assumption unfounded, since the speaker shares the assumption, and hence there is no epistemic conflict. As expected, *bərə* is infelicitous here.

(8) **Sp expects p, infers that Ad believes that p**
CONTEXT: *A is telling B about an upcoming Beyonce concert. This artist only plays at big arenas. B knows Anu fairly well and knows that she loves Beyonce so much that if there is a concert by her she will go.*
A: Beyonce is playing a show here in June. I should tell Anu.
(Assumption: Anu will be interested in going to this concert)
B: #anu bər-i itk-ya moʈʰ-ya concert-la dza-il
anu.F.SG BARA-F.SG DEM.much.PROX-OBL big-OBL concert-DAT go-FUT
#'As if Anu will go for such a big concert'

Now compare (7) and (8) with (9) below (a minimally distinct context) where the speaker is also opinionated about the prejacent. Here, the speaker is aware that Anu does not prefer large concert venues. It can be inferred that the addressee thinks Anu might be interested in going for this concert. Here, a rejecting *bərə* response is licensed.

(9) **Sp expects $\neg p$, infers that Ad believes that p**
CONTEXT: *A is telling B about an upcoming Beyonce concert. This artist only plays at big arenas. B knows that Anu doesn't like going to large concert venues.*
A: Beyonce is playing a show here in June. I should tell Anu.
(Assumption: Anu will be interested in going to this concert)
B: anu bər-i itk-ya moʈʰ-ya concert-la dza-il
anu.F.SG BARA-F.SG DEM.much.PROX-OBL big-OBL concert-DAT go-FUT
'As if Anu will go for such a big concert'

Thus, misalignment between speaker's expectations and addressee's assumptions is necessary for licensing of rejecting *bərə* responses.

2.1.2 Questioning *bərə* and counterexpectations

bərə responses express unexpectedness of the prejacent. Questioning *bərə* responses, while accepting the prejacent, do not express simple surprises, since the prejacent is totally incompatible with speaker's beliefs. They require presence of counterexpectations just like rejecting *bərə*. Similar to rejecting *bərə*, the speaker of questioning *bərə* responses is also opinionated about the prejacent prior to the preceding context update proposal. As shown in (10), in the absence of any prior expectation, a questioning *bərə* response is infelicitous. In (11), the speaker's expectations align with what is learned hence, *bərə* is infelicitous.

(10) **Sp neutral toward p, infers that Ad believes that p**
CONTEXT: *Ram has been planning a surprise vacation for his new partner Anu. I*

don't know anything more about this plan. I am familiar with Anu but I don't know her preferences regarding vacation destinations or climate. I meet Ram and ask him "weren't you planning a surprise for Anu? What did she like the most". Ram says to me -

Ram: Yes! Anu loved Anjuna beach!
(Defeasible entailment: Anu went to Goa)
Me: #anu bər-i gov-ya-la ge-l-i
anu.F.SG BARA-F.SG goa-OBL-DAT go-PERF-F.SG
'Anu went to Goa?!(What's up with that?)'

(11) **Sp expects p, infers that Ad believes that p**
CONTEXT: *Ram has been planning a surprise vacation for his new partner Anu. I know that Anu's favorite vacation destinations involve seaside. If there is a beach, she will definitely go there. I meet Ram and ask him "weren't you planning a surprise for Anu? What did she like the most?". Ram says to me -*

Ram: Yes! Anu loved Anjuna beach!
(Defeasible entailment: Anu went to to Goa)
Me: #anu bər-i gov-ya-la ge-l-i
anu.F.SG BARA-F.SG goa-OBL-DAT go-PERF-F.SG
'Anu went to Goa?!(What's up with that?)'

Compared to examples presented above, In (12), the speaker's expectations that Anu will not go to the seaside are in conflict with the addressee's implicit commitment that Anu went to Goa – which is necessary to license the questioning interpretation.

(12) **Sp expects $\neg p$, infers that Ad believes that p**
CONTEXT: *Ram has been planning a surprise vacation for his new partner Anu. I know Anu very well and I know that she despises the seaside. I meet Ram and ask him "weren't you planning a surprise for Anu? What did she like the most?". Ram says to me -*

Ram: Yes! Anu loved Anjuna beach!
(Defeasible entailment: Anu went to Goa)
Me: anu bər-i gov-ya-la ge-l-i
anu.F.SG BARA-F.SG goa-OBL-DAT go-PERF-F.SG
'Anu went to Goa?!(What's up with that?)'

In summary, both questioning and rejecting *bərə* responses express that the speaker finds the prejacent of *bərə* to be highly unexpected. Additionally, both readings require an epistemic conflict with an inference from the preceding update proposal. In the next section, I show that *bərə* responses necessarily only target NAI component of the preceding update proposal.

2.2 NAIness of the prejacent

There are different conceptions of at-issueness in the literature. Following Roberts (1996), at-issueness is relativized to the current question under discussion (QUD) in the discourse (Beaver et al. 2017 a.o.). On the other hand, literature following Stalnaker (1978) (such as Farkas & Bruce 2010 and the consequent body of work), emphasizes the proposal component of utterances equating it with at-issueness (call it p-at-issueness). Furthermore, theories of "rhetorical" relations in discourse structure (Hunter & Asher 2016 a.o.) define at-issue proposition as corresponding to the component of freshly uttered sentence that can attach to the discourse tree via some 'coherence relation' (c-at-issueness). Koev (2018) shows that these notions of at-issueness are distinct and not equivalent. Thus, it is important to define the relevant notion of at-issueness for the purposes of the present inquiry. I follow Beaver et al. (2017) in relativizing at-issueness with the current QUD. The following definition, for the so-called Q-at-issueness, is adopted from Koev (2018):

(13) **Q-at-issueness:** A proposition p is Q-at-issue relative to a question under discussion Q and a context c iff
a. p is relevant to Q in c.
b. p is appropriately marked as such relative to Q in c.

Speakers often address and draw attention to NAI content in their language use. Consider (14)[9], in the most natural occurrence of A's question, it is taken as an indirect request. Per (13), what is at-issue in (14) is whether B minds or not but NAI is related to the request. B's response cannot be interpreted as answering the question. It is taken as affirming to A's request that they will take the trash out. Thus, the "main point" is not neccessarily at-issue.

(14) A: Do you mind taking the trash out?
B: Of course! (??I mind taking the trash out/ ✓I will take the trash out)

The empirical contribution of this paper is that *bərə* targets only the NAI inferences that is obtained from the preceding context update proposal. Such a class of particles is not well-studied. There is, however, work on particles that perform denial/disagreement that target NAI meaning along with at-issue meaning. *Mica* in Italian (Frana & Rawlins 2019) and *thorii* in Hindi-Urdu (Bhatt & Homer 2025) appear very similar to *rejecting bərə* but they also target at-issue content. English exclamatory *as if* constructions (Bledin & Srinivas 2020) as well as sarcastic *like* (Camp & Hawthorne 2008) have a profile similar to rejecting *bərə* but they target at-issue content as well.

We have seen positive evidence of *bərə* targeting NAI content so far. To see that *bərə* exclusively targets NAI content, consider (15-a) where a rejecting *bərə* response is infelicitous while targeting Ram's assertion. Whereas, a denial with regular negation or *thodii*[10] (Bhatt & Homer 2025) is felicitous as in (15-b). A *bərə* response targeting the tacit assumption, which is a precondition to getting married, is felicitous as seen in (15-c).

[9] p.c. María Biezma

[10] The Marathi counterpart does not differ in its function from Hindi.

(15) **Sp expects $\neg p$, Ad asserts p** (**Rejecting *bərə***)
CONTEXT: *Someone asks me and Ram, "what are Anu's plans after her graduation?", I know that Anu wants to focus on her career over anything else. Ram says:*
Ram: Anu will get married next year
(Assumption: Anu will look for a partner)

a. #anu bər-i ləgnə kər-el puɖʰ-ch-ya vərʃ-i
anu.F.SG BARA-F.SG marriage do-FUT next-GEN-OBL year-IN
'As if Anu will get married next year' (*#at-issue challenged*)

b. anu ləgnə nahi / thodii kər-ɳaar ahe puɖʰ-ch-ya vərʃ-i
anu.F.SG marriage NEG / THORII do-PROSP be.PRES next-GEN-OBL year-IN
'Anu will not get married next year' (*✓at-issue challenged*)

c. anu bər-i mulg-a shodh-el puɖʰ-ch-ya vərʃ-i
anu.F.SG BARA-F.SG boy-M find-FUT next-GEN-OBL year-IN
'As if Anu will look for a partner' (*✓NAI challenged by bərə*)

Similarly, in (16-a), questioning *bərə* also fails to target at-issue content from the preceding context update proposal. Whereas, an utterance without *bərə* as in (16-b) is felicitous. (16-b) also expresses surprise (uttered with an intonation involving a high rise). Alternatively, a response that directly expresses the *'how come?'* part is also felicitous.

(16) **Sp expects $\neg p$, Ad asserts p** (**Questioning *bərə***)
CONTEXT: *I know that Anu and Mina don't get along at all. Anu had sworn to me she will not speak with Mina again. Mina recently fell very sick. I go to Anu's house and I see that she isn't there. I ask her roommate Ram, "where is Anu?". Ram says:*
Ram: Anu went to see Mina.
(Implicature: Anu cares about Mina now)

a. #anu bər-i mina-la ɓʰet-ayla ge-l-i
anu.F.SG BARA-F.SG mina-DAT meet-INCP go-PERF-F.SG
'Anu went to see Mina!? (what's up with that?) (*#at-issue challenged*)

b. anu mina-la ɓʰet-ayla ge-l-i!?↑/ kəs-ə kaay?
anu.F.SG mina-DAT meet-INCP go-PERF-F.SG how-NEUT what
'Anu went to see Mina!?'/ 'how come?' (*✓at-issue challenged*)

c. anu bər-i mina-ch-i chinta kar-ayla laag-l-i ahe
anu.F.SG BARA-F.SG mina-GEN-F.SG worry.F.SG do-INCP V2-PERF-F.SG BE.PRES
≈'Anu cares about Mina now!? (what's up with that?)'
(*✓NAI challenged by bərə*)

The class of not-at-issue meanings is heterogeneous and large. I have not provided the exact characterization of the felicitous NAI target of *bərə*. Intuitively, a felicitous NAI target of *bərə* can be thought of as a pre-condition to accepting addressee's immediately preceding at-issue context-update proposal. Objecting to pre-conditions results in a 'resistance move' in the sense of Bledin & Rawlins (2016), where, for example in (16), the at-issue update proposal (i.e., that *Anu went to see Mina*) is put in 'limbo' which cannot be resolved till the

issue brought forward by a *bərə* response (that *Anu cares about Mina now*) is addressed. As a result, *bərə* responses shift the issue at hand by making at-issue what was previously not-at-issue in the context update proposal awaiting evaluation. In doing so, *bərə* responses indicate *not being on the same page* regarding the NAI proposition.

In summary, a felicitous *bərə* response must target the NAI proposition q from the preceding proposal that conflicts with speaker's own expectation that q is highly unexpected. In doing so, *bərə* responses yield a resistance move. Therefore, *bərə* is a grammaticalized means to challenge NAI updates meaning that NAI updates are also negotiated. Hence, they must be represented at the proposal stage. This is shown in the next section.

3 Formal implementation

Contrary to the conclusions of previous section, it is a standard assumption in much of the extant literature (e.g., Anderbois et al. 2015, Murray 2014, Rett 2021) that NAI updates directly enter CG bypassing the negotiation stage that at-issue updates must go through. While concerned with related but different issues, recent literature such as Simons (2025) (a.o.) presents cases where some backgrounded content is available for interlocutors while not being CG. The present paper provides further empirical support to this line of thinking and calls for representation of NAI content as proposals in our models of discourses.

Farkas & Bruce's (2010) influential 'table model' has been very successful in modeling various discourse phenomena. In the table model, discourse moves are functions from an input discourse structure K_i to another discourse structure K_o representing the modifications proposed by the utterance. When the modifications are accepted the current context becomes K_o. If rejected, then it stays the same.

(17) A discourse structure K in Farkas & Bruce (2010) (F&B) consists of:

a. **Common ground** (CG): the set of propositions all discourse participants commit to for the communicative purposes.

b. **Discourse commitments** (DC_x): for each participant x, the set of propositions x publicly commits to during the conversation.

c. **Table** (T): a push-down stack of issues that models discourse salience.

d. **Projected set** (ps): the set of propositions that are being considered for addition into the CG.

Below, I present how NAI updates have been represented in the table model as implemented in Rett (2021). Let S be a declarative sentence *I should tell Anu about the concert* such that p be the assertion and q the background assumption that *Anu will be interested in going to the concert*. Then the effect of the declarative S with at-issue content p and not-at-issue content q can be modeled as in (18) where q directly enters CG without going on the table.

(18) Declarative update by S with at-issue content p and not-at-issue content q:

i. $DC_{a,o} = DC_{a,i} \cup \{p\}$

ii. $T_o = push(\langle S; \{p\}\rangle, T_i)$ (at-issue update)

iii. $ps_o = ps_i \cup \{p\}$
iv. $\underline{CG_o = CG_i \cup \{q\}}$ (not-at-issue update)

The core mechanism of the table model involves updating the *table* which stores issues to be evaluated (in this framework, both assertions and questions are issues, differing only in their set cardinalities). Thus, at-issue content goes on the table to be negotiated. Since the table model does not differentiate between declaratives and questions in terms of how they update the context (i.e., both update the table and project addition of proposition(s) to CG), the notion of at-issueness is p-at-issueness which emphasizes proposalhood. As a consequence, being on the table = being a proposal = being at-issue. Given the conclusions from section 2.2, the treatment of NAI updates as in (18) cannot be right as it will make the wrong predictions for the data presented so far. Thus, we ought to represent NAI content at the proposal stage where it is available for negotiation instead of being forced into the common ground. To do so in the F&B model would require changing the fundamental architecture of the table model, an endeavor that I do not pursue here. Instead, I present the dynamic assumptions required to implement proposalhood of NAI content in the *waiting room* model of discourse in Biezma & Rawlins (2017) (henceforth, B&R) which preserves the spirit of the table model but defines context update process and proposalhood differently.

3.1 The dynamic assumptions of the discourse structure

For the goals of this paper, we will need a context structure that tracks the QUD and content that addresses it (or not). Following Stalnaker (1978) and Farkas & Bruce (2010), we will assume that utterances are proposals to update the context. Moreover, we will assume that the aim of conversation is a communal inquiry organized around questions that participants agree to jointly pursue (the QUD). Thus, a declarative is a proposal to update the CG or its associated context set (cs). Interrogatives, in turn, are proposals to update the inquiry to be pursued i.e., the QUD stack (Roberts 1996). In B&R's model, proposalhood is captured by defining a 'local context' distinguished from the proposed update in the context. A local context (l_c) is the way things are at a given point in the discourse i.e., what is jointly accepted to be the case (cs) and the joint inquiry to be pursued (QUD). A proposal, then, is a window into the way things would be if the proposed change to the way things are is accepted. This projected/future context is structurally the same as the local context but *with* proposed modifications. Thus, a full context consists of the local context and a potential projected context. The projected context is the proposal waiting to be evaluated – the "waiting room".

(19) A *local context* l_c of a context c is a tuple $\langle cs, \mathcal{Q} \rangle$ where:
   a. cs is a context set.
   b. $\mathcal{Q}$ is a stack of sets of propositions - QUD.

(20) A *context* c is a tuple $\langle cs, \mathcal{Q}, \mathscr{F} \rangle$ that is characterized as:
   a. $l_c = \langle cs, \mathcal{Q} \rangle$ is a local context.
   b. $\mathscr{F}_c$ is either a local context or $\emptyset$. $\mathscr{F}_c$ is called the *projected context*.

Note that the *projected context* is a copy of the local context *with* certain modifications (the local update). Assuming the standard *push, pop* and *top* operations[11] on QUD stacks, the immediate QUD in a local context l_c is always $top(\mathcal{Q}_c)$. The updates by a declarative and an interrogative utterance are defined as follows:

(21) Local updates over a local context l,

a. $l \oplus \ulcorner \varphi_{\langle s,t \rangle} \urcorner = \langle cs_l \cap [\![\varphi]\!], \mathcal{Q}_l \rangle$ (Declarative update)
Felicity constraints:
i.cs_l is compatible with $[\![\varphi]\!]$ (assertability)
ii.$[\![\varphi]\!]$ is relevant to $top(\mathcal{Q}_l)$

b. $l \oslash \ulcorner \varphi_{\{\langle s,t \rangle\}} \urcorner = \langle cs_l, push(\mathcal{Q}_l, [\![\varphi]\!]) \rangle$ (Interrogative update)
Felicity constraints:
i.cs_l is compatible with $\{w | \exists p \in ([\![\varphi]\!]) : p(w)\}$ (answerability)
ii.$[\![\varphi]\!]$ is relevant to $top(\mathcal{Q}_l)$ or $\mathcal{Q}_l = \langle\rangle$

Given the operations defined above for local contexts, B&R define context update operations for an assertion and a question. Along with acceptance and rejection, a 'maintenance' operations is defined – namely, elimination of a QUD once resolved ($\langle\rangle$ denotes the empty stack). These are given in table 1.

Update operations		**Constraints**
Assertion	$c + \ulcorner \text{Assert}(\varphi) \urcorner = \langle cs_c, \mathcal{Q}_c, l_c \oplus \ulcorner \varphi \urcorner \rangle$	(i) $\mathscr{F}_c = \emptyset$; (ii) $l_c \oplus \ulcorner \varphi \urcorner$ is felicitous.
Questioning	$c + \ulcorner \text{Question}(\varphi) \urcorner = \langle cs_c, \mathcal{Q}_c, l_c \oslash \ulcorner \varphi \urcorner \rangle$	(i) $\mathscr{F}_c = \emptyset$; (ii) $l_c \oslash \ulcorner \varphi \urcorner$ is felicitous.
Maintenance		**Constraints**
Dispel	$c + \ulcorner \text{Pop} \urcorner = \langle cs_c, pop(\mathcal{Q}_c), \mathscr{F}_c \rangle$	$\mathscr{F}_c = \emptyset$, $\mathcal{Q}_c \neq \langle\rangle$
Evaluation		
Accept	$c + \ulcorner \text{Accept}_x \urcorner = \langle cs_{\mathscr{F}}, \mathcal{Q}_{\mathscr{F}}, \emptyset \rangle$	
Reject	$c + \ulcorner \text{Clear} \urcorner = \langle cs_c, \mathcal{Q}_c, \emptyset \rangle$	

Table 1: Context update operations

Acceptance replaces the original context by the projected context and leaves the proposal slot (the projected context) empty. In section 3.2, I will illustrate the dynamics with a working example. To represent NAI updates at the proposal stage we will tweak how $\mathscr{F}$ is updated by a declarative update as in (21-a) to include the total information extracted from the move.

[11]These operations are traditionally defined as follows for a QUD stack $\mathcal{Q}$:

1. $push(e, \mathcal{Q})$ is the new stack obtained by adding an item e to the top of the stack $\mathcal{Q}$.
2. $pop(\mathcal{Q})$ is the stack obtained by popping off the top item of the stack $\mathcal{Q}$.
3. $top(\mathcal{Q})$ is the top item of the stack $\mathcal{Q}$.

3.2 NAI updates as proposals

Let us work through the context update process in the *waiting room* formalism using example (4) repeated here as (22).

(22) **Questioning *bərə***

CONTEXT: *A is telling B about a recent Beyonce concert A went to. This artist only plays at big arenas. B knows that Anu doesn't like going to large concert venues. B asks A, "how was the concert?"*

(QUD: How was the concert that A went to?)

A: It was great! Anu was singing along to every song!
(Defeasible entailment: Anu went to a big concert)

B: anu bər-i itk-ya mo[h]-ya concert-la ge-l-i
anu.F.SG BARA-F.SG DEM.much.PROX-OBL big-OBL concert-DAT GO-PERF-F.SG
'Anu went to such a big concert!? (what's up with that?)'

Suppose that before B's question to A, $CG = \{$RECENT-CONCERT-IN-TOWN, ARTIST-PLAY-BIG-CONCERTS, ANU-LIKE-ARTIST, A-WENT-TO-CONCERT$\}$. The dynamic context update process is demonstrated in Table 2. Let $\otimes$ be a placeholder for update operations $\oplus$ and $\oslash$.

$c = \langle cs,\ \mathcal{Q},\ \mathscr{F} \rangle$			
	The way things are $l_c = \langle cs_c, \mathcal{Q}_c \rangle$ (local context)		**What they may become ('waiting room') $\mathscr{F} = l_c \otimes \psi$**
	What participants mutually accept	Participants' goals	
c_0	cs	$\mathcal{Q}$	$\emptyset$
			B: How was the concert? $(\{p_1, \ldots, p_n\})$
c_1	cs_{c_0}	$\mathcal{Q}_{c_0}$	$\mathscr{F} = l_{c_0} \oslash \{p_1, \ldots, p_n\} = \langle cs_{c_0},\ push(\mathcal{Q}_{c_0}, \{p_1, \ldots, p_n\}) \rangle$
			A: It was great! $\{p_m\}$, $m \in [1, n]$; [Accept$_B(c_2)$ + Proposal that $p_m(c_3)$]
c_2	$cs_{\mathscr{F}_{c_1}} = cs_{c_0}$	$\mathcal{Q}_{\mathscr{F}_{c_1}} = push(\mathcal{Q}_{c_0}, \{p_1, \ldots, p_n\})$	$\emptyset$
c_3	cs_{c_0}	$\mathcal{Q}_{c_2}$	$\mathscr{F} = l_{c_2} \oplus \{q\} = \langle cs_{c_0} \cap p_m,\ \mathcal{Q}_{c_2} \rangle$
			[Silence/subsequent utterance by A]
c_4	$cs_{\mathscr{F}_{c_3}} = cs_{c_0} \cap p_m$	$\mathcal{Q}_{c_2}$	$\emptyset$
			Maintenance operation: IQUD resolution > dispel IQUD
c_5	$cs_{c_4} = cs_{c_0} \cap p_m$	$pop(\mathcal{Q}) = \mathcal{Q}_{c_0}$	$\emptyset$
$c_5 = \langle cs_{c_0} \cap p_m,\ \mathcal{Q}_{c_0},\ \emptyset \rangle$			

Table 2: Update mechanism illustrated

At c_0, the waiting room is empty as there is no proposal to be evaluated. After B's question, the local context is fixed as $l_{c_0} = \langle cs_{c_0}, \mathcal{Q}_{c_0} \rangle$ and a questioning update is performed which pushes B's question on top of the stack in the projected context. A's response to the question indicates that the proposed question is accepted resulting in the waiting room being emptied out. Acceptance results in the updated context c_2. Next, A's answer is the new proposal. At c_3, the local context is fixed as $l_{c_2} = \langle cs_{c_0}, \mathcal{Q}_{c_2} \rangle$ and the waiting room is populated with the proposed declarative update. Since A is not interrupted and continues with their next utterance, the preceding proposal is tacitly accepted by B and the waiting room is emptied out resulting in c_4 where the content of $\mathscr{F}_{c_3}$ are copied onto the local context l_{c_2}. Since resolved QUDs are popped off the stack, we yield the context c_5.

A lot more information is learned from a discourse move than just what is said (such as entailments, presuppositions, (non-canceled) implicatures, intentions etc.) (Gunlogson 2008). In the framework adopted here, the way things currently are is distinct from what they might become. Therefore, this secondary content must also be part of what things might become i.e., the projected context. This information is also tracked and evaluated by agents. Based on Biezma (2014), let's call such information extracted from a move M_i by an agent a, the *information gain* I_{a,M_i} and the literal content of the move M_i as $Content(M_i)$.

(23) a. $I_{a,M_i} = \{p : p \text{ is TRIGGERED by } M_i\}$ (Information gain from a move)
b. $Content(M_i) = [\![\alpha]\!]^{c,g}$, where α is the linguistic form used in M_i

So a declarative update involves projecting the total information gain from the move into the waiting room $\mathscr{F}$ i.e., the declarative update operation in (21-a) is modified as in (24).

(24) Given a context, $c = \langle cs, Q, \mathscr{F} \rangle$, M_i a discourse move to be evaluated in c, $\mathscr{F}_c$ defined above, a declarative update proposed via M_i by an agent a is defined as:
$\mathscr{F}_{a,cs_c,M_i} = \langle \{w : w \in (\bigcap I_{a,M_i}) \cap cs_c\}, Q_c \rangle$ (Total update proposed)
paraphrase: The cs_c update proposed by a via M_i is the set of worlds in the intersection of information gain from M_i that are part of cs.

Therefore, the NAI meaning is part of the total update that would take place if the projected update is accepted which allows us to represent NAI updates as part of the proposal. This is where the target *bərə* is located. I demonstrate this with the example in (22) in Table 3. Continuing from c_5 in table 2, the waiting room is empty at c_5 as there is no proposal to be evaluated. B continues by asserting that *Anu was singing along to every song*. This triggers the NAI inference that *Anu went to the concert*. In the proposed update by A's utterance, the information gain I_{A,M_3} would contain both these propositions. At c_6, the local context is fixed as $l_{c_5} = \langle cs_{c_5}, \mathcal{Q}_{c_0} \rangle$. Moreover, the waiting room is populated with the total update resulting from A's move. Observe that in c_6 there is a proposal waiting to be evaluated. This proposal contains both the asserted information (p) and the associated NAI information (q).

$c_5 = \langle cs_{c_4}, \mathcal{Q}_{c_0}, \emptyset \rangle$			
	The way things are $l_c = \langle cs_c, \mathcal{Q}_c \rangle$ **(local context)**		**What they may become ('waiting room')** $\mathscr{F} = l_c \otimes \psi$
	What participants mutually accept	Participants' goals	
c_5	cs_{c_4}	$\mathcal{Q}$	$\emptyset$
A: Anu was singing along to every song! (p) $\rightsquigarrow$ Anu went to the big concert (q); $I_{A,M_4} = \{p, q\}$			
c_6	cs_{c_5}	$\mathcal{Q}_{c_0}$	$\mathscr{F} = l_{c_5} \oplus (\bigcap I_{A,M_3}) =$ $\langle cs_{c_4} \cap (\bigcap I_{A,M_3}), \mathcal{Q}_{c_0} \rangle$

Table 3: Updating with NAI as proposals

Thus, **considering the total information gain from a move to be part of what is evaluated allows us to represent NAI content at the proposal stage** and the NAI target of *bərə* responses can be located in the context update this way. Having shown how to represent NAI updates as proposals, in section 4, I provide a formal entry for *bərə* and give a unified account deriving the two readings.

4 Core interpretation of *bərə*

Informally, *bərə* responses express that there is a not-at-issue proposition q in the information extracted from the preceding move that is highly unexpected because speaker considers $\neg q$ to be very likely. Thus, *bərə* signals misalignment in that NAI q needs to be negotiated before the at-issue p update can be resolved. We have seen that the NAI target of *bərə* can be located in the information extracted from the preceding context update proposal. For the purposes of the analysis, I assume that *bərə* takes propositional scope. Given this, I provide a surprise-style entry (see Romero 2015) to model unexpectedness. Let a, b be agents, $(Un)exp_{b,w}$ be a selection function which pairs a proposition p with a degree of (un)expectedness d for an agent b, and $\theta_{b,c}$ be the contextually provided threshold then,

(25) $[\![\textit{bərə}\ \text{q}]\!]^{c,g}(w) = 1$ iff $\forall w' \in \bigcap Dox(b,w).\exists d[Unexp_{b,w'}(q,d) \wedge d > \theta_{b,c,high}]$
(counter-expectation)

defined only if,

a. $\exists M_{a,i}$, a move with an update proposal by a awaiting evaluation ($\mathscr{F}_c \neq \emptyset$)
b. $\exists q \in I_{a,M_i} \& q \neq Content(M_i)$ (q is not-at-issue in M_i)
c. $\forall w'' \in \bigcap Dox(b,w) : \exists d'[Exp_{b,w''}(\neg q, d') \wedge d' > \theta_{b,c,high}]$ (speaker bias)
paraphrase: In all the worlds compatible with the speaker's beliefs, q is more unexpected than a contextually provided threshold. Defined only if q is the NAI content extracted from the immediately preceding discourse move and the speaker believes that $\neg q$ is very likely.

(25-a) rightly predicts that *bərə* responses are infelicitous out of the blue. (25-b) encodes the NAI-ness constraint and with the asserted component it encodes the felicity constraint of counter-expectations. Given this formal entry for *bərə* and having achieved our goal of representing NAI updates at the proposal stage in section 3.2, I will now show how the two seemingly distinct readings of *bərə*, *incredulous rejection* and *questioning*, can be derived from the same mechanism. When our expectations are violated, often the acceptance/rejection of the unexpected information depends on the credence of the information and the communicator. The governing factor, therefore, will be interaction of the speaker's expectations with considerations on relative epistemic authority of the interlocutors.

4.1 Deriving *rejection*

The rejection reading obtains when the speaker does not consider the addressee a reliable source for the commitments expressed in their utterance. This is a scenario where the speaker of *bərə* response infers that they have more epistemic authority regarding the events under discussion. Consider our core case in (3) reproduced below as (26):

(26) CONTEXT: *A is telling B about an upcoming concert of an artist in the town. This artist only plays at big arenas. B knows that Anu doesn't like going to large concert venues. A only knows that Anu likes this type of music.*

A: Oh! I should tell Anu about this.
(Assumption: Anu will be interested in going to this concert)

B: anu bər-i itk-ya mo[ʰ]-ya concert-la dza-il
anu.F.SG BARA-F.SG DEM.much.PROX-OBL big-OBL concert-DAT go-FUT
'As if Anu will go for such a big concert'

Supposing that CG = {UPCOMING-CONCERT, ARTIST-PLAY-BIG-CONCERTS, ANU-LIKE-ARTIST, A-KNOWS-ANU-LIKES-ARTIST}, before the *bərə* response, the speaker reasons as follows:

1. **Speaker knowledge:** Apart from the CG facts, B also knows that Anu doesn't like going to big concerts at all. Her doing so would be highly unexpected.

2. **Addressee's move:** A expresses their intention to tell Anu about the concert, assuming that Anu will be interested in going to this concert.

3. **Reliability inferences:** If A knew Anu's preferences they wouldn't express such intentions at all. Which means, A's assumption is based on their ignorance of Anu's preferences. Therefore, A is determined to be an unreliable source of their implicit commitment to the assumption.

4. **Epistemic conflict:** If A's update proposal regarding their intentions is accepted, the NAI assumption also gets accepted as part of total information gain. This results in a discrepancy between the B's own beliefs and the information they must accept – It must be challenged immediately.

Following the above reasoning, speaker B utters (26)-B resulting in a rejection of A's assumption. Note that at no point has the speaker accepted the NAI assumption that *Anu will be interested in this concert*. Thus, the NAI content could not have been automatically CG.

4.2 Deriving *questioning*

The questioning reading results in the speaker appearing as accepting the prejacent along with a *questioning* effect that can be paraphrased as *'how come p is the case?'*. This reading obtains when the speaker determines that the addressee is a reliable source of their commitments i.e., the speaker does not have epistemic authority on the matters being discussed. Consider the core case in (4) reproduced below as (27):

(27) CONTEXT: *A is telling B about a recent concert of an artist in the town. This artist only plays at big arenas. I know that Anu doesn't like going to large concert venues. B asks A, "how was the concert?"*

A: It was great! Anu was singing along to every song!
(Defeasible entailment: Anu went to a big concert)

B: anu bər-i itk-ya mo$\mathfrak{t}^{h}$-ya concert-la ge-l-i
anu.F.SG BARA-F.SG DEM.much.PROX-OBL big-OBL concert-DAT GO-PERF-F.SG
'Anu went to such a big concert!? (what's up with that?)'

Suppose that CG = {RECENT-CONCERT-IN-TOWN, ARTIST-PLAY-BIG-CONCERTS, ANU-LIKE-ARTIST, A-WENT-TO-CONCERT} After A's assertion, the speaker reasons as follows:

1. **Speaker knowledge:** Apart from the CG facts, B knows that Anu doesn't like going to big concerts at all. Her doing so would be highly unexpected.

2. **Addressee's move:** In saying that Anu was singing along to every song, A also implies that Anu was at the concert. This is a first-hand report of the events at the concert.

3. **Reliability inferences:** Given that A went to the concert, they are a reliable source of information they are providing. Therefore, B must accept the proposal as in case of an open disagreement, B will lose.

4. **Epistemic conflict:** Accepting the at-issue proposal means accepting that Anu went to the concert. This would cause discrepancy between B's own beliefs and the information they must accept.

5. **Contingency of Speaker commitment:** This discrepancy must be resolved immediately. For this, A is determined as the epistemic authority in this matter, thus B is dependent on A to resolve this discrepancy.

Let's take stock. A is implicitly determined as authoritative on this matter. At the same time, B cannot easily, straightforwardly reject or accept that Anu went to the concert. B is dependent on A to resolve this conflict. This is the hallmark of *questioning* flavor in

Gunlogson (2008) cashed out in terms of the *contingent commitment criterion* wherein an utterance of a declarative is *questioning* to the extent that the speaker's commitment to it is taken to be contingent on the addressee's ratification of the same. Therefore, the *questioning* flavour here marks the dependency or contingency of the speaker's commitment on the addressee to commit to the relevant proposition. Note that at no point has the speaker already accepted the NAI update. Thus, the NAI content cannot have been CG.

Thus, the two seemingly different discourse effects can be derived from the same underlying mechanism of reasoning based on the epistemic stance of the speaker and considerations on relative epistemic authority. As argued here, we see from the reasoning that the NAI content from preceding update proposal is not directly added to the common ground. It must be part of the proposal to be targetted by *bərə* responses. Proposalhood of NAI updates was represented in B&R's *waiting room* model of discourse.

5 Future outlook

The present investigation is a step towards a general theory of understanding not-at-issue updates and strategies to challenge them. *bərə* is not the only particle that exclusively challenges not-at-issue content. Hinterwimmer & Ebert (2018), Hinterwimmer (2019) show that Bavarian particle *fei* exclusively targets not-at-issue content as well. Due to space constraints, I cannot provide a detailed description and analysis of *fei* but I briefly sketch their approach and compare it with *bərə*.

Like *bərə*, *fei* also seems to require an epistemic conflict between speaker (p) and inferred addressee beliefs ($\neg p$). Unlike *bərə*, *fei* can target conventionally triggered meanings such as presuppositions and conventional implicature (i.e., projective content). *rejecting bərə* performs incredulous rejection glossed as *as if p*. This gloss is a very close translation of the effect. Thus, it is worth exploring whether the present account extends to exclamatory *as ifs*. The main difference being that *as ifs* target both at-issue and not-at-issue content of a preceding utterance.

Hinterwimmer (2019) takes *fei* to be a marker of resistance to adding the at-issue content q to CG owing to a background misalignment on p that becomes salient due to the at-issue update. As a result, the proposed semantic account posits that *fei* is licensed when there is no recent assertion by the addressee that entails $\neg p$. Crucially, the assumption therein also equates at-issueness with proposalhood thus not reaching the conclusions argued for in this paper. Given the similarity between the two particles' function, the formalism presented in this paper provides a way towards a discourse dynamic account of *fei*.

Finally, Bledin & Srinivas (2020) provide an exclamation based account of *as ifs* relying on the EX-operator combined with a treatment of *as if* as a hypothetical comparative. This account is also purely semantic. Exclamatory *as ifs* target some content learned from the preceding move so a discourse dynamic treatment is desirable and worth exploring. The present dynamic model's conception of proposalhood offers a basis to explore the complex interaction between discourse dynamics and compositional semantics of *as if*.

6 Conclusion

The main takeaway is that the strategy of directly challenging not-at-issue updates is grammaticalized. *bərə* is one of the rare particles that exclusively targets NAI component of the preceding context update proposal. In doing so, it makes at-issue what was not-at-issue in the previous move. There are other means to turn at-issue what was not-at-issue such as bare *if*-clauses (Biezma 2025) and *que*-clauses (Biezma 2026) in Spanish which are metadiscursive claims about the shape of the context – checking that all participants are on the same page about some backgrounded content or that a proposal of a particular kind needs to be evaluated. *bərə* responses are not just checking devices. They not only specify the shape of the context i.e., what the previous move requires one to accept if the move is accepted. In addition to identifying specific not-at-issue content, they also convey the speaker's attitude towards this specific NAI meaning being treated as uncontroversial previously (that it is unexpected). This results in *rejection* or *questioning how come* interpretation. Since acceptance/rejection often involves considerations about credence of information provided. The present proposal derive these seemingly distinct effects from interaction between speaker expectations and considerations on addressee's credence.

bərə shows us that NAI updates are also proposals so they must be represented as such at the proposal stage. I have shown using the waiting room framework of Biezma & Rawlins (2017) that considering the total information extracted from a move as part of the projected update allows us to do this. Therefore, at-issueness should not be equated with proposalhood. Instead, being (not-)at-issue has to be understood in terms of whether a proposition addresses the current QUD or not. Finally, present formalism also has potential to be extended to other related and similar phenomena mentioned here.

Acknowledgements

Thanks are due to my advisors Rajesh Bhatt and María Biezma for advising on this project. I'd also like to thank the audiences at UMass Semantics workshop and the Syntax-Semantics reading group (especially, Jonathan Pesetsky, Andrea Matticchio, and Hee Joong Choi) for many fruitful discussions. Thanks to Sharvari Phadnis for corroborating judgments on the data. Finally, thanks to the three anonymous reviewers for FASAL–15 as well as participants of SNEWS @MIT and FASAL–15 for helpful comments and feedback. All errors are my innovation and mine alone.

contact: sphadnis@umass.edu

References

Anderbois, Scott, Adrian Brasoveanu & Robert Henderson. 2015. At-issue proposals and appositive impositions in discourse. *Journal of Semantics* 32(1). 93–138. doi:10.1093/jos/fft014.

Beaver, David I., Craige Roberts, Mandy Simons & Judith Tonhauser. 2017. Questions Under Discussion: Where Information Structure Meets Projective Content. *Annual Review of Linguistics* 3. 265–284. doi:10.1146/annurev-linguistics-011516-033952.

Bhatt, Rajesh & Vincent Homer. 2025. Negation of disagreement in hindi-urdu. In Frances Blanchette & Cynthia Lukyanenko (eds.), *Perspectives on negation*, 215–252. De Gruyter Mouton. doi:10.1515/9783110762068-008.

Biezma, María. 2014. The grammar of discourse: The case of then. In *Semantics and linguistic theory 24*, 373–394. doi:10.3765/salt.v24i0.2444.

Biezma, María. 2026. Non-Canonical questions and the role of context in the construction of meaning: Insubordinated that-clauses, Oxford: OUP. To appear.

Biezma, María & Kyle Rawlins. 2017. Rhetorical questions: Severing asking from questioning. In *Semantics and linguistic theory 27*, 302–322. doi:10.3765/salt.v27i0.4155.

Biezma, María. 2025. Insubordinated if-clauses as discourse subordination. *Natural Language Semantics* 33. 43–81. doi:10.1007/s11050-024-09229-0.

Bledin, Justin & Kyle Rawlins. 2016. Epistemic resistance moves. In *Semantics and linguistic theory 26*, 620–640. doi:https://doi.org/10.3765/salt.v26i0.3812.

Bledin, Justin & Sadhwi Srinivas. 2020. Exclamatory As Ifs. In *Proceedings of sinn und bedeutung*, vol. 24 1, 84–101. doi:10.18148/sub/2020.v24i1.854.

Camp, Elisabeth & John Hawthorne. 2008. Sarcastic 'Like': A Case Study in the Interface of Syntax and Semantics. *Philosophical Perspectives* 22. 1–21. https://www.jstor.org/stable/25177220.

Condoravdi, Cleo & Sven Lauer. 2012. Imperatives: Meaning and illocutionary force. *Empirical issues in syntax and semantics* 9. 37–58.

Deo, Ashwini. 2025. Take on this commitment: The particle 'bərə' in Marathi (Indo-Aryan). In *Proceedings of sinn und bedeutung*, vol. 29, 386–403. doi:10.18148/sub/2025.v29.1219.

Farkas, Donka F. & Kim B. Bruce. 2010. On Reacting to Assertions and Polar Questions. *Journal of Semantics* 27(1). 81–118. doi:10.1093/jos/ffp010.

Frana, Ilaria & Kyle Rawlins. 2019. Attitudes in discourse: Italian polar questions and the particle *mica*. *Semantics and pragmatics* 12(16). 1–55. doi:10.3765/sp.12.16.

Gunlogson, Christine. 2008. A question of commitment. *Belgian Journal of Linguistics* 22(1). 101–136. doi:10.1075/bjl.22.06gun.

Hinterwimmer, Stefan. 2019. The bavarian discourse particle *fei* as a marker of non-at-issueness. In *Secondary content*, 246–273. Leiden, Netherlands: Brill. doi:10.1163/9789004393127_011.

Hinterwimmer, Stefan & Cornelia Ebert. 2018. A comparison of 'fei' and 'aber'. In *Proceedings of sinn und bedeutung*, vol. 22 1, 469–486. doi:10.18148/sub/2018.v22.101.

Hunter, Julie & Nicholas Asher. 2016. Shapes of conversation and at-issue content. In *Semantics and linguistic theory 26*, 1022–1042. doi:10.3765/salt.v26i0.3946.

Koev, Todor. 2018. Notions of at-issueness. *Language and Linguistics Compass* 12(12). e12306. doi:10.1111/lnc3.12306.

Krifka, Manfred. 2013. Response particles as propositional anaphors. In *Semantics and linguistic theory 23*, 1–18. doi:10.3765/salt.v23i0.2676.

Murray, Sarah E. 2014. Varieties of updates. *Semantics and Pragmatics* 7(2). 1–53. doi:10.3765/sp.7.2.

Potts, Christopher. 2005. *The logic of conventional implicatures*. New York: Oxford University Press. doi:10.1093/acprof:oso/9780199273829.001.0001.

Rett, Jessica. 2021. The Semantics of Emotive Markers and Other Illocutionary Content. *Journal of Semantics* 38(2). 305–340. doi:10.1093/jos/ffab005.

Roberts, Craige. 1996. Information structure in discourse. In J.H. Yoon & A. Kathol (eds.), *OSU working papers in linguistics*, vol. 49, 91–136. Ohio State University.

Roelofsen, Floris & Donka F. Farkas. 2015. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language* 91(2). 359–414. doi:10.1353/lan.2015.0017.

Romero, Maribel. 2015. Surprise-predicates, strong exhaustivity and alternative questions. In *Semantics and linguistic theory 25*, 225–245. doi:10.3765/salt.v25i0.3081.

Simons, Mandy. 2025. Availability without common ground. *Linguistics and Philosophy* 48(1). 179–211. doi:10.1007/s10988-024-09426-4.

Snider, Todd N. 2017. *Anaphoric reference to propositions*. Ithaca, NY: Cornell University dissertation.

Stalnaker, Robert. 1978. Assertion. *Syntax and Semantics* 9. 315–332.

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy* 25(5/6). 701–721. http://www.jstor.org/stable/25001871.