Proceedings of
Formal Approaches to South Asian Linguistics 14

Held in 2024 at Stony Brook University

April 2025

Edited by Daniel Greeson, Shrayana Haldar, Anushree Mishra, & Aidan Sharma

The 14th Conference on (Formal) Approaches to South Asian Languages (FASAL-14) was held from April 4 to April 6, 2024, at Stony Brook University in Stony Brook, New York, and was sponsored by the Department of Linguistics, the Mattoo Center for India Studies, the College of Arts & Sciences, and the Anandavalli & Dr. G. Swaminathan Endowed Research Professorship.

The conference was held in a hybrid format, with virtual and in-person talks from scholars working in North America, South Asia, Europe, and Oceania. In this volume, we have collected fifteen papers from the five invited talks and twenty-five submitted presentations presented at the workshop. FASAL-14 had a special theme of locality South Asian languages, and featured plenary talks from Professors Raghavachari Amritavalli (EFLU Hyderabad), Athulya Aravind (MIT), Dustin Chacón (UGA, UCSC), Prerna Nadathur (OSU), and Paroma Sanyal (IIT Delhi). Many different facets of the study of South Asian languages were represented, with talks related to morphology (AITHA, AMRITAVALLI, DASH, SANYAL), phonology & phonetics (PARAMORE, SANYAL), syntax (AGARWAL, AKOLKAR, AMRITAVALLI, ARAVIND, BASU, BHATT-MISHRA, CHACÓN, DASH, JABBAR, KAUR-SINHA, MIHRA-SYED, OTTUR, SHAH, SHARMA, SYED-BANERJEE-BANERJEE, THALLURI, VENKATESAN) semantics & pragmatics (AKOLKAR, HALDAR, JABBAR, NADATHUR, SURESH, WANG), language contact (KULKARNI, SAYNYAL-TOM, THALLURI), and neuro- & psycholinguistics (CHACÓN, HOQUE-MCLENDON-DUNAGAN-KHOKHAR-CHACÓN, SUBRAMONY), with data from a diverse array of Indo-Aryan, Dravidian, and Austroasiatic languages. The conference concluded with a thought-provoking panel discussion of different approaches, challenges, and benefits of investigating locality and other linguistic phenomena in South Asian languages, with panelists Athulya Aravind, Dustin Chacón, Thomas Graf, Paroma Sanyal, and Sandhya Sundaresan.

We thank the faculty organizers, Sandhya Sundaresan and Thomas McFadden, student organizer Daniel Greeson, the above-mentioned University sponsors, and the many volunteers from Stony Brook University whose efforts made FASAL-14 possible. Finally, we are incredibly grateful to have learned so much from the attendees and presenters, whose top-notch work in all areas of South Asian Linguistics makes us proud to be part of this community of researchers.

– Daniel Greeson, Shrayana Haldar, Anushree Mishra, & Aidan Sharma

Contents

Phases are *Read-Only*: Evidence from Hindi-Urdu
Hashmita Agarwal **1**

Four Puzzles and Affixal N in Hindi
R. Amritavalli **21**

Malayalam long-distance anaphor *taan*: a null theory
Athulya Aravind **41**

Scrambling in Bengali: An A-/A'- Movement Distinction
Sreyoshi Basu **61**

Relative Deletion
Rajesh Bhatt & Anushree Mishra **82**

It's about time!: Relating structure, the brain, and comparative syntax
Dustin Chacón **102**

Verbalization as Re-categorization of Lexical Categories in Santali
Biswanath Dash **119**

Modal Debris: Threefold Ambiguities between Permission, Weak Necessity, & Strong Necessity in Bengali
Shrayana Haldar **135**

Borrowing and dissappearance of light verbs: Loan-verb integration in Indian languages
Aaditya Kulkarni **148**

On the interaction of aspect and ability in two Hindi/Urdu constructions
Prerna Nadathur **163**

Verb Root Allomorphy in Indo-Aryan Languages
Paroma Sanyal **184**

Malayalam-Kannada Code-mixing
Paroma Sanyal & Therese Liam Tom **204**

Familiar Definite Marking in Magahi
Aidan Sharma **222**

Processing of Relative Clauses in Malayalam
Jayakrishnan Subramony **242**

An Agree-based Analysis of Nominal Agreement: Evidence from Hindi-Urdu
Madhusmitha Venkatesan **259**

Phases are *Read-Only*: Evidence from Hindi-Urdu

HASHMITA AGARWAL, *University of California Los Angeles*

ABSTRACT

Under Chomsky (2000, 2001)'s Phase Impenetrability Condition, phases induce Transfer of their complements, rendering the complements inaccessible. As a consequence, cross-phasal dependencies are ruled out. Recent work on phases has suggested that instead of being eliminated, transferred phase complements remain in the syntax (Bošković 2003; Obata 2010; Chomsky 2012; Chomsky et al. 2019). In this paper, I expand on the idea of spelled out phase complements being visible but not completely accessible for syntactic processes. I propose a *Read-Only* view of phases, wherein phase complements are not deleted from the narrow syntactic derivation for inspection after undergoing Transfer, but the featural content of the phase complement becomes unalterable. The major consequence of this view is a nuanced conception of phase locality, such that some cross-phasal dependencies—namely those that do not require feature valuation of a transferred element—are possible. Cross-phasal dependencies that do value features of transferred elements continue to remain impossible, like in standard phase theory. I show that Hindi-Urdu φ-agreement and case assignment bear out the predictions of *Read-Only* with regard to cross-phasal dependencies. φ-agreement by a higher probe with a transferred goal, where the goal itself is not altered, is possible in Hindi-Urdu. On the other hand, accusative case assignment into a spelled out phase complement—which involves valuing the case feature of the transferred goal—is impossible. However, the same transferred DP that cannot be accusative is able to condition dative case on a DP in a higher phase. I argue that no notion of phases—other than *Read-Only*—accounts for the Hindi-Urdu pattern. The phase locality imposed by *Read-Only* offers a new way of accommodating dependencies between elements belonging to different phases in a principled way.

1 Phases and locality

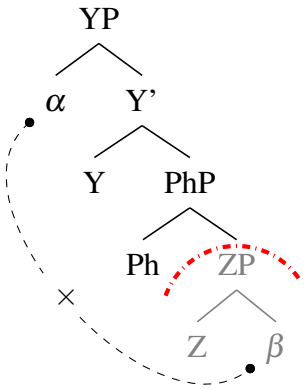
In modern syntactic theory, syntactic structure is typically constructed in chunks called phases. According to Chomsky (2000, 2001)'s Phase Impenetrability Condition (PIC) in (1), phases send their complements to the interfaces, rendering the complement inaccessible for further syntactic operations. Chomsky (2004) calls this operation of shipping off phase complements Transfer.

(1) **PHASE IMPENETRABILITY CONDITION** (Chomsky 2000:108)
In phase α with head H, the domain of H is not accessible to operations outside of α, only H and its edge are accessible to such operations.

Chomsky (2008) proposes that the cycle involved in the derivation is strong enough to prevent spelled out phase complements to even be looked into, so phase complements are entirely 'forgotten' by the narrow syntax after they are sent to the interfaces. Following Chomsky (2004, 2007, 2008), Ott (2011) interprets Transfer as an operation eliminating phase complements from the syntax—a common assumption in subsequent literature (Polinsky & Potsdam 2001; Bruening 2001; Branigan & MacKenzie 2002; Legate 2003; Grosz & Patel 2006; Epstein et al. 2009; Keine 2013, among others).

Due to phase complements being removed from the derivation upon Transfer in Chomsky's theory, dependencies between transferred elements (β) and elements active in the derivation (α) are ruled out in his PIC framework, (2):[1]

(2) *No cross-phasal dependencies*



Despite the prevalence of Chomsky (2004, 2007, 2008)'s phase complement-eliminating view of Transfer in the syntactic literature, more recent work has suggested that phase complements are *not* in fact eliminated from the syntax. Instead, transferred phase complements remain fixed in place in the narrow syntactic derivation under this alternative view (Bošković 2003, 2007; Fox & Pesetsky 2005; Obata 2010, 2017; Chomsky 2012 and Chomsky et al. 2019).

Bošković (2003, 2007) in particular argues that phases induce cyclic linearisation of their complements, thus constraining movement out of them, but linearised elements remain visible to Agree and other processes. Cross-phasal φ-agreement is then permitted (and shown to be possible in some languages) in Bošković (2003, 2007)'s view, raising questions about the sensitivity of other syntactic dependencies to phases.

In this paper, I explore the behaviour of case assignment in relation to phases. While the early-Chomskyan view of phases does not capture the possibility of cross-phasal φ-agreement, I argue that phases inducing cyclic linearisation alone à la Bošković (2003, 2007) also does not sufficiently model their behaviour, because phases also constrain case assignment. Case assignment to a transferred nominal does not affect the fixed order of linearised elements at that phase, so spelled out phase complements that remain present to

[1]The Chomsky (2001) version of the PIC also rules out dependencies with material in phase complements, the difference being that the inaccessibility of a phase complement is delayed until the next higher phase head merges.

be agreed with under Bošković (2003, 2007)'s view should also in principle be available for case assignment.

Keeping in mind case assignment alongside φ-agreement and movement, I propose an extension and refinement of Bošković (2003, 2007)'s proposal, termed *Read-Only*, (3).

(3) ***Read-Only***
Upon Transfer, phase complements undergo cyclic linearisation and feature freezing, but remain visible from outside.

The important addition of *Read-Only* is that, in addition to cyclic linearisation, phases induce feature freezing: the featural content of material inside a phase complement remains visible but cannot subsequently be altered. In relation to cross-phasal dependencies, *Read-Only* offers a middle ground between Chomsky (2000, 2001, 2004, 2007, 2008)'s conception of phases, and Bošković (2003, 2007)-style cyclic linearisation. The distinctive properties of *Read-Only* are stated in (4):

(4) a. Dependencies relating transferred item β and phase-external α that modify α are possible;

b. Dependencies relating transferred item β and phase-external α that modify β are impossible.

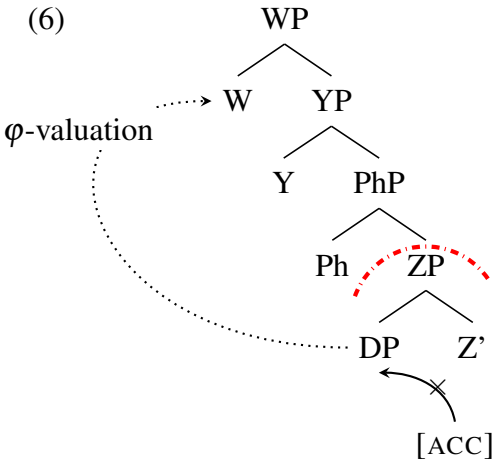
Crucial empirical evidence for *Read-Only* in (4) comes from Hindi-Urdu case and φ-agreement patterns. In particular, I show that in Hindi-Urdu:

(5) a. A φ-agreement relation valuing a phase-external probe in response to a transferred goal DP is possible—the probe gets modified.

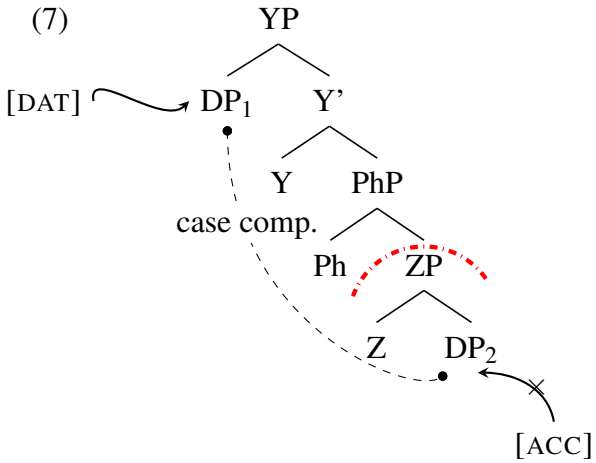
b. Accusative case assignment to a transferred DP conditioned by a phase-external element is impossible—the transferred DP gets modified.

c. Dative case assignment to a DP in a higher phase conditioned by a transferred DP is possible—the higher DP gets modified.

(6)



(7)



The following structural assumptions about Hindi-Urdu are used to implement *Read-Only*:

(8) a. $fseq = \langle C \succ T \succ Asp \succ Voice \succ v\ (\succ Appl) \succ V \rangle$

b. C, Voice, and *v* are phase heads

To study the behaviour of case assignment in relation to phases, I now turn to a theory of accusative case assignment in Hindi-Urdu, whose sensitivity to phases is contrasted with φ-agreement.

2 The accusative case–φ-agreement asymmetry in Hindi-Urdu

This section addresses the asymmetry between cross-phasal accusative case assignment and φ-agreement. I first show that the former cannot target a nominal that has undergone Transfer, and then show that the latter can. Against this background, I argue that *Read-Only*—unlike other accounts of phase locality—captures the disparity between the two processes.

2.1 Accusative case

Hindi-Urdu has differential object marking (Aissen 2003; Montaut 2018; Kalin 2018; Kalin & Weisser 2019), which I analyse as accusative case following Baker & Vinokurova (2010); Baker (2015, to appear).[2] (9) shows that some direct objects are obligatorily accusative, while (10) shows that some objects are optionally accusative, depending on certain semantic properties of the DP (Mahajan 1990; Butt 1993; Mohanan 1994; Butt & King 2004; Kachru 2006; Keine 2007; Mahajan 2017a; Kidwai 2022).

(9) *Obligatorily accusative object*
Komal=ne Tina*(=ko) dekhaa
Komal=ERG Mina*(=ACC) saw
'Komal saw Tina'

(10) *Optionally accusative object*
Komal=ne fuul(=ko) dekhaa
Komal=ERG flower(=ACC) saw
'Komal saw a/the flower'

The pattern for accusative case on a direct object in Hindi-Urdu—similar to other differential object marking (DOM) systems—is roughly summarised in (11). (Aissen 2003; Bhatt 2007; Davison 2014)

[2]By 'accusative case' I specifically mean the *-ko* marker that appears on many direct objects of transitive clauses and becomes optional under passivisation. I distinguish accusative *-ko* from abstract/null case assigned to direct objects, and from dative *-ko* which appears on indirect objects/goals and does not alternate.

(11) *Accusative case in Hindi-Urdu*

a. Specific animate DOs are obligatorily accusative

b. For inanimate DOs, accusative case correlates with specificity

c. Nonspecific inanimate DOs are never accusative

Neither unaccusative arguments nor unergative arguments can receive accusative case, (12).

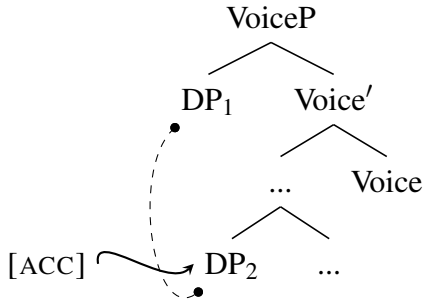
(12) *No accusative on intransitive arguments*
{laṛkaa / *laṛke=ko} giraa / khããsaa
{boy / *boy=ACC} fell / coughed
'The boy fell / coughed'

The appearance of accusative case on objects only in the presence of a higher argument suggests a dependent case analysis of the Hindi-Urdu accusative, following Baker & Vinokurova (2010); Baker (to appear). Under a head case analysis, assigning accusative case to an object only in the presence of a higher DP would be a coincidence. A dependent accusative case rule for Hindi-Urdu is stated in (13) (c.f. Baker to appear).

(13) **DEPENDENT ACCUSATIVE CASE**
If a case-unmarked DP_1 c-commands DP_2 in VoiceP, assign accusative to DP_2

The external argument introduced in Spec,VoiceP is the case competitor of the accusative object, (14).

(14) *Dependent accusative case*



The rule in (13) then makes correct predictions for intransitives like (12) and transitives with obligatorily accusative objects like (9)—In the case of intransitives, there is no case competitor so no accusative case is assigned. In transitive clauses with obligatorily accusative objects, the presence of the external argument within VoiceP triggers the accusative case rule on the direct object.

However, (13) alone does not derive the optionality of accusative case on many inanimate DPs like *fuul* in (10). I claim that the optionality of accusative case in cases like (10) is a *v*P phase effect—in combination with independently-motivated object shift, following (Baker & Vinokurova 2010; Baker to appear)'s analysis of DOM in Sakha and other languages.

Bhatt & Anagnostopoulou (1996) note that different relative orders of the direct and indirect object in ditransitives are possible in Hindi-Urdu, importantly showing accusative

direct objects to be structurally higher than case-unmarked direct objects, (15):

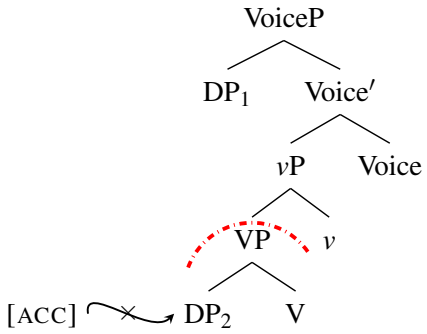
(15) *No -ko on unmoved direct objects*

a. S IO DO*(*=ko)* V
miinaa=ne ṭiinaa=ko **kitaab**(*=ko) dii
Mina=ERG Tina=DAT book(*=ACC) gave
'Mina gave Tina a/the book'

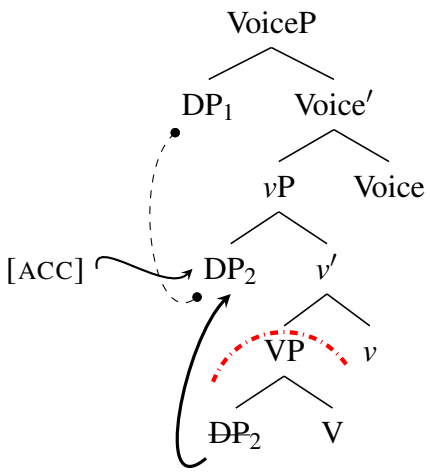
b. S **DO**=*ko* IO V
miinaa=ne **kitaab=ko**$_j$ ṭiinaa=ko ___$_j$ diyaa
Mina=ERG book=ACC$_j$ Tina=DAT ___$_j$ gave
'Mina gave Tina the book'

Since the syntactic properties of accusative direct objects in ditransitives are identical to those of accusative direct objects in monotransitives, it must be the case that accusative objects in monotransitives are also higher than case-unmarked objects.[3] Therefore, I follow Baker (to appear) in claiming that object shift feeds accusative case in Hindi-Urdu. Accusative case is never assigned to a direct object in its base position, (16). Following Bhatt & Anagnostopoulou (1996) as well as Baker & Vinokurova (2010), I assume that accusative case is fed by semantically-motivated Diesing (1992)-style movement. I further assume that Diesing (1992)-style movement lands in the specifier of *v*P, (17). Specific animate objects always undergo Diesing (1992)-style movement, and so are always accusative. Inanimate direct objects may or may not undergo movement out of VP, depending on their specificity, resulting in variable accusativity. Nonspecific and inanimate objects remain in VP, and do not receive accusative case.

(16) *No accusative case on unmoved DO*



(17) *Movement to Spec,vP feeds* ACC



[3]The optionality of accusative direct objects in both ditransitive and monotransitive clauses depends on semantic properties like animacy and specificity. Additionally, accusative objects in both types of clauses may become case-unmarked under passivisation.

DP_2 in (16) fails to receive accusative case because it undergoes Transfer as part of the vP phase complement by the time DP_1 is merged into the structure. In (17), DP_2 escapes *Transfer* by moving to Spec,*v*P, and thus receives accusative case by being in the same local domain as its case competitor DP_1.

2.2 φ-agreement

φ-agreement in Hindi-Urdu is obligatory, and targets the structurally highest DP that bears no case marker, (18). In (18a), the case-unmarked subject—the highest DP in the structure—controls φ-agreement. In (18b), the subject already has ergative case, so the case-unmarked object is the φ-agreement controller.

(18) a. *Subject agreement*
laṛkii chaand dekheg-ii / *-aa
girl.F.SG moon.M.SG see.FUT-F.SG / *-M.SG
'The girl will see the moon'

b. *Object agreement*
laṛke=ne tasviir dekh-ii / *-aa
boy.M.SG=ERG photograph.F.SG saw-F.SG / *-M.SG
'The boy saw a photograph'

Case-marked DPs in Hindi-Urdu never control agreement. In (19), the subject is ergative and object is accusative. Thus there is no case-unmarked DP to agree with, and the verb must show default (masculine singular) agreement.

(19) *Default (masculine singular) agreement*
laṛkii=ne kitaabõ=ko paṛh-aa / *-ii / *-ĩĩ
girl.F.SG=ERG book.F.PL=ACC read-M.SG / *-F.SG / *-F.PL
'The girl read the books'

Following Bobaljik (2008); Preminger (2014), I assume that the Hindi-Urdu φ-probe on T is *case-discriminating*, such that only DPs without a valued case feature can be targeted by the φ-probe.[4] The φ-agreement algorithm for Hindi-Urdu is given in (20):

(20) **φ-AGREEMENT ALGORITHM**
Agree with a subject iff it is case-unmarked;
or else agree with an object iff it is case-unmarked;
or else show default (masculine singular) agreement

φ-agreement in Hindi-Urdu does not need to be fed by moving the agreement controller, as Davison (1991); Boeckx (2004); Bhatt (2005); Bhatt & Keine (2017); Keine (2020) have argued (pace (Mahajan 1989, 2017b)).

In the idiom in (21a)—from Bhatt & Keine (2017)—X-*kii khuub marammat karna* 'give X a beating', the idiomatic object *marammat* must stay in its base position for the idiomatic reading to be preserved. Movement of the idiomatic object, as in (21b), destroys the idiomatic reading and results in a sentence that can only be interpreted literally. When the subject is ergative, the idiomatic object in its base position in (21a) must be a φ-agreement target, so movement is not required for φ-agreement in Hindi-Urdu.

[4]There is good reason to claim that the φ-probe is on T and not on a lower head like *v* or Voice. First, when both arguments in a transitive clause are unmarked, subject agreement is the only possibility, as shown in (18a). The preference of subject agreement over object agreement follows straightforwardly if the probe is higher than the subject and object, since the subject is more local to the probe. Object agreement is then correctly predicted to only be possible when the subject is invisible for φ-agreement due to being case-marked, as in (18b). Béjar & Rezac (2009) use a similar line of argumentation for Basque, where they claim that the Basque φ-probe is on *v*, given that object agreement has precedence over subject agreement in the language. Additionally, if the φ-probe were on *v*/Voice, the Agree relation involved in φ-agreement would be predicted to be upward or downward, instead of just downward, making the system more complex.

The second argument for the Hindi-Urdu φ-probe being on T comes from the interaction between ergative case assignment and φ-agreement. If φ-agreement precedes ergative case assignment, we would expect that φ-agreement-controlling subjects can receive ergative case. However, such a pattern where a subject is both ergative and controls φ-agreement is completely unattested in any variety of Hindi-Urdu (i), suggesting that the φ-probe is at least as high as the ergative-assigning head.

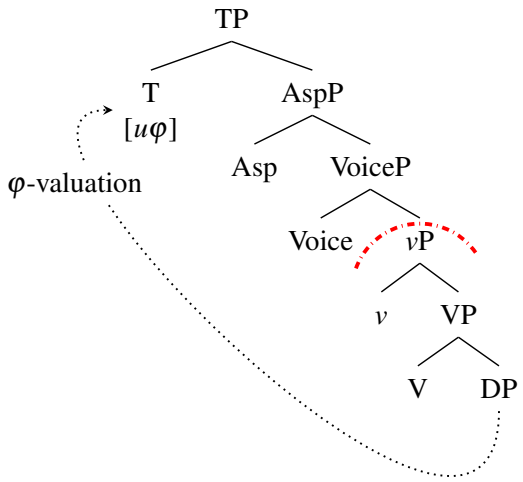
(i) *Ergative subjects never control agreement*
larkii=ne seeb khaay**-aa** / ***-ii**
girl.F.SG=ERG apple.M.SG saw-M.SG / *-F.SG
'The girl ate an apple'

(21) a. *Idiom (with obligatory object agreement)*
aamir=ne aman=kii khuub marammat ki-**i**/*ki-**yaa**
Aamir=ERG Aman=GEN many repairs.F.SG did-F.SG/*did-M.SG
'Aamir gave Aman a good beating' (lit.: 'Aamir did Aman's many repairs')

b. *No movement of idiomatic object marammat*
#[khuub marammat]$_j$ aamir=ne aman=kii ___ $_j$ kii
many repairs$_j$ Aamir=ERG Aman=GEN ___ $_j$ did
'Aamir did Aman's many repairs' (*no idiomatic reading*)

Since φ-agreement in Hindi-Urdu is obligatory and does not require the agreement controller to move, at least some (object) DPs control agreement from within a phase complement, after they have already undergone Transfer, (22).

(22)



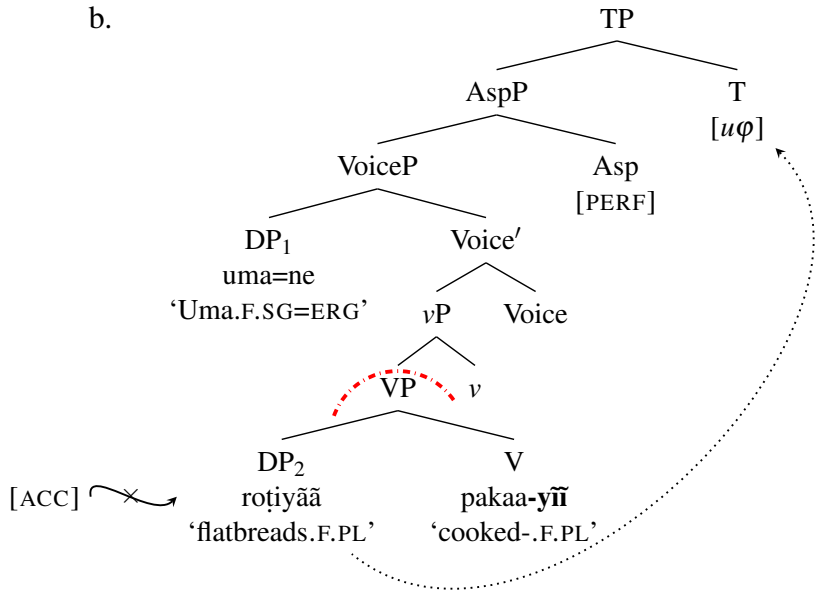
Recall from §2.1 that *v* is a phase, as shown by the inability to assign accusative case into its complement. Then, the lack of analogous *v*P phase effects for Hindi-Urdu φ-agreement is surprising.

2.3 Accusative case–φ-agreement asymmetry

Accusative case assignment is local, and cannot target an object in a transferred phase complement. φ-agreement, on the other hand, may target unshifted objects in VP after it has already undergone Transfer. In the monotransitive example (23a)—illustrated in (23b)—*roṭiyãã* 'flatbreads' cannot receive accusative case (conditioned by external argument *Uma*) in its base position but *roṭiyãã* 'flatbreads' obligatorily controls agreement from VP.

(23) a. *Case-unmarked agreeing object*
uma=ne roṭiyãã pakaa-**yĩĩ** / *-**yii** / *-**yaa**
Uma.F.SG=ERG flatbreads.F.PL cooked-F.PL / *-F.SG / *-M.SG
'Uma cooked (the) flatbreads'

b.



Read-Only is the only conception of phases that is able to account for the differences between cross-phasal accusative case assignment and φ-agreement. A PIC view of phases does not predict the opacity of phases for accusative case assignment but not φ-agreement in (23). If *v*P is a phase and phase complements are thus inaccessible for both case assignment and φ-agreement, agreement with *roṭiyãã* 'flatbreads' in (23) is unexpected. If *v*P is not a phase under the PIC conception, accusative case assignment is expected on *roṭiyãã* 'flatbreads' in its base position. There is no derivational option under a PIC conception of phases that permits an element in a phase complement to be targeted for φ-agreement but not case assignment.

Bošković (2003)'s cyclic linearisation alone also fails to make the correct predictions with regard to the accusative case–φ-agreement asymmetry in (23). Even with *v*P phasehood, *roṭiyãã* 'flatbreads' in VP is incorrectly predicted to be accessible for accusative case assignment under this view.

Due to feature freezing of transferred phase complements under *Read-Only* repeated in (24), there is a straightforward account of the accusative case assignment–φ-agreement asymmetry. Accusative case assignment to *roṭiyãã* 'flatbreads' into a VP is impossible because valuing the DP's case feature requires overwriting a transferred phase complement after the features in it are already frozen in place. φ-agreement with *roṭiyãã* 'flatbreads' in its base position in the *v*P phase complement, on the other hand, remains possible because it only involves inspecting a phase complement without tampering with its frozen features.

(24) ***Read-Only***
Upon Transfer, phase complements undergo cyclic linearisation and feature freezing, but remain visible from outside.

Now that I have contrasted the impossibility of assigning case into a transferred phase complement with the seemingly exceptional ability of φ-agreement to target a transferred DP,

I bring in dative case and its ability to be conditioned by a transferred element to provide further support for a *Read-Only* system.

3 The accusative case–dative case asymmetry

In this section, I argue for a dependent case analysis of dative case in Hindi-Urdu, and show that dative case can be conditioned by a transferred case competitor. However, the same transferred case competitor that may trigger dative case on a higher case competitor cannot itself receive accusative case, which only a *Read-Only* view of phases can account for.

3.1 Dative case

Dative case in Hindi-Urdu appears in ditransitives (25), experiencers (26), and causatives, among other constructions. In ditransitives, the indirect object invariably receives dative case, which is syncretic with accusative *-ko*. Following Larson (1988); Davison (2004); Pylkkänen (2008), I assume that indirect objects as well as experiencers are introduced in the specifier of an Appl(icative) head, which *v* takes as its complement.

(25) *Dative on Indirect Objects*
miina=ne ṭiina*(=ko) kitaab dii
Mina=ERG Tina*(=DAT) book gave
'Mina gave Tina a/the book'

(26) *Dative experiencer*
chhatr*(=ko) ḍigrii milii
student*(=DAT) degree got
'The student got the degree.'

Evidence for dative *-ko* and accusative *-ko* being different cases (pace Kalin 2014) comes from dative case being obligatory in contrast with the often optional accusative case. Dative *-ko* is obligatory on indirect objects even if they are inanimate (27a), but accusative *-ko* is optional on inanimate objects (27b). Unlike with accusative case, a DP's semantic properties have no influence on the obligatoriness of dative case.

(27) a. *Inanimate DO, optional Accusative -ko*
nisha=ne fuul(=ko) dekhaa
NishaERG flower(=ACC) saw
'Nisha saw a/the flower'

b. *Inanimate IO, obligatory Dative -ko*
nisha=ne fuul*(=ko) paanii diyaa
Nisha=ERG flower*(=DAT) water gave
'Nisha watered the flower' (Lit:'Nisha gave water to the flower')

In passives of transitives, accusative *-ko* becomes optional on a direct object (28a), even on DPs that require it in active voice. When an indirect object is passivized however, dative *-ko* remains obligatory on the indirect object, (28b).

(28) a. *Passivised DO, optional Accusative -ko*
raam(=ko) bulaayaa gayaa
Ram(=ACC) called PASS
'Ram was called'

b. *Passivised IO, obligatory Dative -ko*
raam*(=ko) kitaab dii gayii
Ram*(=ACC) book given PASS
'Ram was given a book'

It is clear from (27-28) that although dative and accusative case is Hindi-Urdu are both realised as *-ko*, they are structurally different.

The dative case data we have seen so far is compatible with both, a head case analysis of the Hindi-Urdu dative (where Appl assigns dative case to its specifier), and a dependent case analysis where dative is assigned to the higher of two DPs that *v* c-commands. Causatives are crucial in showing that only a dependent case analysis correctly accounts for the distribution of Hindi-Urdu dative case, as I have also argued in Agarwal (2024).

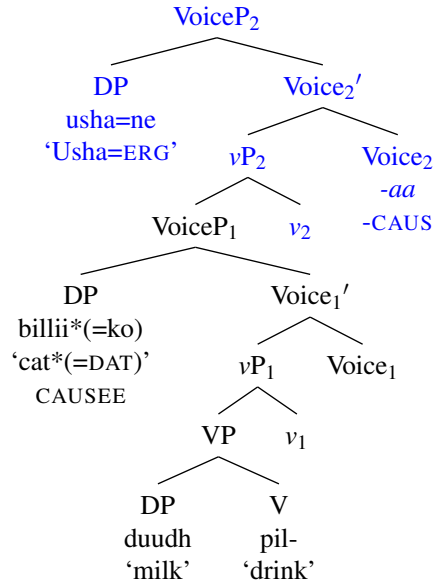
Ingestives—like in (29a)—are a class of transitive verbs in Indic that are made causative by adding a causative morpheme *-aa* to the verb stem, and introducing a causer argument, as in (29b). *Usha* is the added causer argument in (29b), while *billii* 'cat' becomes the causee. Importantly, dative case is found in causativised ingestives—the causee DP *billii* 'cat' in (29b), which is sandwiched between *Usha* and *duudh* 'milk' is marked dative.

(29) a. *Ingestive*
billii=ne duudh pii-yaa
cat=ERG milk drink-PFV
'The cat drank milk'

b. *Causativised ingestive*
usha=ne billii*(=ko) duudh pil-aa-yaa
Usha=ERG cat*(=DAT) milk drink-CAUS-PFV
'Usha made the cat drink milk'

Following Baker & Vinokurova (2010), Harley (2008), and Bhatt & Embick (2017) for Sakha, Japanese, and Hindi respectively, I assume that the structure of transitive causatives is derived from the structure of plain transitives by adding a causer and a causative morpheme, (30).

(30) *Causative structure*



Under functional head case theory, there are two contenders for dative case assigners in the causative: First, Voice_1, which introduces *billii* 'cat' in both the ingestive in (29a) and its causative in (29b). Second, the causative head Voice_2, which embeds VoiceP_1 and introduces the causer. I will now argue that neither of these heads assign dative case in the causative in Hindi-Urdu.

Voice_1 does not assign dative case to *billii* 'cat' in the simple ingestive in (29a), so it cannot assign dative case to *billii* 'cat' in the causative in (29b). As seen in (31), the simple ingestive is ungrammatical with a dative subject in place of an ergative subject.

(31) *Ingestive with dative subject*
*billii=ko duudh pii-yaa
cat=DAT milk drink-PFV
Intended: 'The cat drank milk'

Voice_2—the causative head—also does not assign dative case. Consider the unergative in (32a), and its causativised counterpart in (32b). Notably, the causee in the causative in (32b) is not dative. It is either case-unmarked or accusative, but not dative, as evidenced by the optionality of *-ko*. Recall from (28) that any instance of optional *-ko* is accusative. Passivising (32b) corroborates that *-ko* on *kutta* 'dog' is accusative, and not dative, (32c).

(32) a. *Unergative*
kutta dauṛ rahaa hai
dog run PROG AUX.PRES
'The dog is running (around)'

b. *Causativised unergative*
salma kutte(=ko) dauṛ-aa rahii hai
Salma dog(=ACC) run-CAUS PROG AUX.PRES
'Salma is making a/the dog run'

c. *Passive of causativised unergative*
kutta / kutte=ko dauṛ-aa-yaa jaa rahaa hai
dog / dog=ACC run-CAUS-PFV PASS PROG AUX
'The dog is being made to run.'

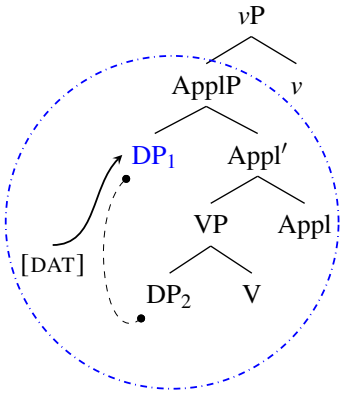
Thus, the causative head *-aa*—represented as Voice_2 in (30)—also cannot assign dative case. Then, no head is available to assign dative case to the causee in causatives of transitive ingestives like (29b), and dative case in Hindi-Urdu cannot be a functional head case.

Since dative case only ever appears in the presence of a lower DP in (25-29), its distribution is best captured with dative being a dependent case. In Agarwal (2022, 2024), I propose the rule in (33) for dependent dative case in Hindi-Urdu, following Baker & Vinokurova (2010) for Sakha.

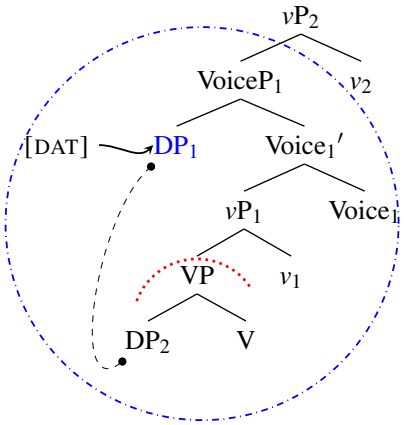
(33) **DEPENDENT DATIVE CASE RULE**
If DP_1 c-commands DP_2 in the complement of *v*P, assign dative to DP_1

Along with goals in ditransitives and experiencer arguments (34a), (33) also accounts for the assignment of dative case to causees in ingestivised causatives like *billii* 'cat' in (29b) as in (34b), due to the presence of the lower case competitor DP_1. Similarly, (33) correctly accounts for the absence of dative case on the embedded arguments in (32), since a lower case competitor is absent.

(34) a. *Dative case on goals/experiencers*



b. *Dative case in causatives*



I have argued here that the dependent dative case analysis exemplified in (34b) is the only

viable account of dative in ingestivised causatives. Importantly for our purposes, DP_2—the case competitor of the dative DP_1—is in a lower phase, but—as (29b) showed—dative case is still obligatorily assigned in the causative.

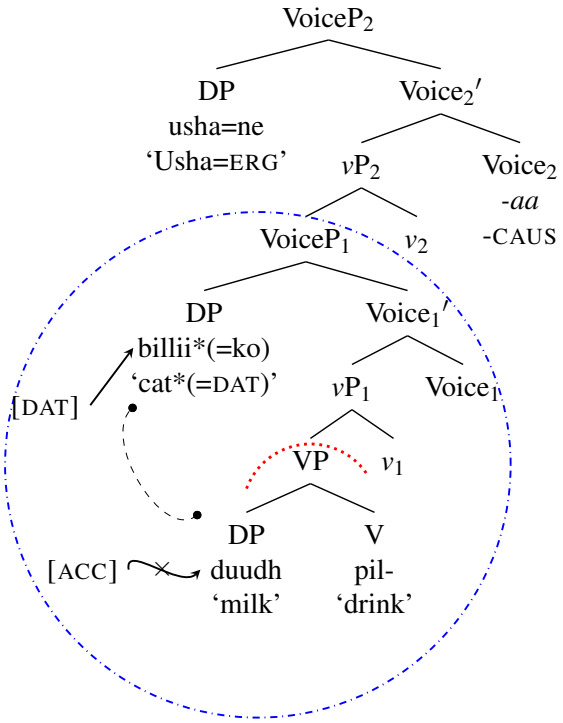
3.2 Accusative-dative asymmetry

In this subsection, I square the ability of a transferred DP in the ingestivised causative construction to condition dative case on an active DP with the inability of the same transferred DP to receive accusative case. I then show that *Read-Only* derives this asymmetry between locally-assigned accusative case and nonlocal dative case.

Reconsider the example of the ingestivised transitive causative construction in (29b), repeated here as (35) and illustrated in (36). Here, *duudh* 'milk' cannot receive accusative case due to being trapped in VP, but nevertheless conditions dative case on *billii* 'cat' when vP$_2$ merges and the structural description of the dative rule is met.

(35) *Causativised ingestive*
usha=ne billii*(=ko) duudh pil-aa-yaa
Usha=ERG cat*(=DAT) milk drink-CAUS-PFV
'Usha made the cat drink milk'

(36)



In (35), *duudh* 'milk' stays in-situ and undergoes *Transfer* as a part of VP when v_1 enters the structure. Then, *duudh* 'milk' has already been spelled out by the time the accusative case competitor *billii* 'cat' is merged in VoiceP$_1$. I noted in §2 that accusative case cannot be assigned to a transferred DP. Thus, the unmoved object *duudh* 'milk' cannot receive

accusative case via the rule in (13) even though it has a higher case competitor in VoiceP$_1$.

Turning to dative case assignment in (35), the causee *billii* 'cat' obligatorily receives dative case, which (34b) showed is triggered by the presence of a lower DP, specifically the object. In this instance, *duudh* 'milk'—being the only lower DP—must be the case competitor for dative case on *billii* 'cat'. Despite *duudh* 'milk' having undergone *Transfer* in the *v*P cycle$_1$, it still exceptionally conditions dative case on *billii* 'cat' when v_2 merges.

As was the case with the asymmetry between cross-phasal φ-agreement and accusative case assignment, the PIC and Bošković (2003) are unable to account for the pattern observed with regard to accusative and dative case competition in Hindi-Urdu, while *Read-Only* derives it. For the PIC account, if *v*P$_1$ is a phase under this view, *duudh* 'milk' in (35) is expected to be unavailable to condition dative case on *billii* 'cat', since *duudh* 'milk' should already be deleted from the syntax when the causee merges. Note that *duudh* 'milk' in (35) also controls agreement from its base position, which would be disallowed if it were completely inaccessible. If *v*P is not a phase given the PIC, *duudh* 'milk' in its base position is incorrectly predicted to receive accusative case conditioned by *billii* 'cat' in addition to conditioning dative case on *billii* 'cat' (and controlling φ-agreement).

Bošković (2003)'s cyclic linearisation also fails to accommodate the asymmetry between cross-phasal accusative case assignment and dative case competition. Whether or not *v*P phasehood is assumed under this view, the unmoved object *duudh* 'milk' is wrongly expected to be available for accusative case assignment, alongside conditioning dative case on *billii* 'cat'.

Once again, *Read-Only* is the only conception of phases that encompasses the asymmetry between accusative and dative case in Hindi-Urdu. *duudh* 'milk' in (35), which undergoes *Transfer* in the *v*P phase cycle, cannot receive accusative case under *Read-Only* due to its case feature being frozen in place. *duudh* 'milk', however, is still visible post-*Transfer* to condition dependent dative case on *billii* 'cat', since a case competition relation merely requires read-access to the features of *duudh* 'milk' in VP.

4 Discussion and conclusion

Using case and agreement data from Hindi-Urdu, I have argued for a *Read-Only* view of phases (37), under which complements of phases are still present in the narrow syntax after Transfer, but are not modifiable.

(37) ***Read-Only***
Upon Transfer, phase complements undergo cyclic linearisation and feature freezing, but remain visible from outside.

Major evidence for *Read-Only* phases came from the asymmetric visibility of transferred nominals for φ-agreement and case competition, but not accusative case assignment. Under this conception of phases, cross-phasal dependencies that modify a phase-external element are permitted, therefore correctly permitting φ-agreement or case competition with a transferred element, which only modifies the active element in the higher phase. Cross-phasal

dependencies that modify a transferred element are not allowed under *Read-Only*. Therefore, cross-phasal accusative case assignment—which requires valuing and thus modifying the features of a transferred DP—is correctly ruled out.

As shown in Table 1 below, existing notions of phase locality, like the Chomsky (2000, 2001)'s PIC or Bošković (2003)'s cyclic linearisation are unable to account for the selective sensitivity of the three dependencies to phases in Hindi-Urdu, necessitating the nuanced view of phase locality offered by *Read-Only*.

Post-spellout / View of locality	Visibility	Agreement	Case assign.	Case comp.
Chomsky (2000, 2001)'s PIC	*	*	*	*
Bošković (2003)	✓	✓	✓	✓
Read-Only	✓	✓	*	✓
Observed in Hindi-Urdu	✓	✓	*	✓

Table 1: Phase effects under different theories

The strong predictions of *Read-Only* for the (im)possibility of cross-phasal φ-agreement, case assignment, and case competition post-Transfer raises interesting questions about how this notion of phase locality regulates other cross-phasal syntactic dependencies. In Hindi-Urdu in particular, which has a rich case system, the sensitivity of accusative case assignment to phases begs the question of whether ergative case assignment also respects phases. In Agarwal (2022), I bring in light verb constructions to show that ergative case—a functional head case—cannot in fact be assigned into a transferred phase complement, in contrast to φ-agreement. The ergative case–φ-agreement asymmetry replicates the results of §2, providing further support for a *Read-Only* view of phases.

Like case assignment, movement out of a transferred phase complement also remains correctly ruled out in this system by the cyclic linearisation clause of *Read-Only*, much like it is ruled out under the PIC and Bošković (2003).

Further work in phase locality would test the predictions of *Read-Only* for φ-agreement, case assignment, and case competition in other languages with similarly attested long distance dependencies, especially in cases where existing theories of phase locality do not suffice. Exploring further syntactic dependencies, like NPI-licensing, *wh*-licensing, control etc., and their sensitivity to phases would also be revelatory for the *Read-Only* system.

References

Agarwal, Hashmita. 2022. *Phases are Read-Only*: University of California, Los Angeles MA thesis.

Agarwal, Hashmita. 2024. Dependent dative case in Hindi-Urdu. In Samopriya Basu & Aaditya Kulkarni (eds.), *Proceedings of Formal Approaches to South Asian Languages 13*, 1–16.

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 21(3). 435–483.

Baker, Mark. 2015. *Case*. Cambridge University Press.

Baker, Mark. to appear. On dependent case and the sometimes independence of ergativity and differential object marking. In Elena Anagnostopoulou, Christina Sevdali & Dionysios Mertyris (eds.), *On the place of case in grammar*, Oxford.

Baker, Mark C & Nadya Vinokurova. 2010. Two modalities of case assignment: Case in Sakha. *Natural Language & Linguistic Theory* 28(3). 593–642.

Béjar, Susana & Milan Rezac. 2009. Cyclic agree. *Linguistic Inquiry* 40(1). 35–73.

Bhatt, Rajesh. 2005. Long distance agreement in Hindi-Urdu. *Natural Language & Linguistic Theory* 23(4). 757–807.

Bhatt, Rajesh. 2007. Unaccusativity and case licensing. *Talk presented at McGill University* .

Bhatt, Rajesh & Elena Anagnostopoulou. 1996. Object shift and specificity: Evidence from ko-phrases in Hindi. *Papers from the main session of CLS* 32. 11–22.

Bhatt, Rajesh & David Embick. 2017. Causative derivations in Hindi-Urdu. *Journal of Indian Linguistics* 78(1-2). 93–151.

Bhatt, Rajesh & Stefan Keine. 2017. Long-distance agreement. *The Wiley Blackwell Companion to Syntax, Second Edition* 1–30.

Bobaljik, Jonathan David. 2008. Missing persons: A case study in morphological universals. *The Linguistic Review* (25(1-2)). 203–230.

Boeckx, Cedric. 2004. Long-distance agreement in Hindi: Some theoretical implications. *Studia linguistica* 58(1). 23–36.

Bošković, Željko. 2003. Agree, phases, and intervention effects. *Linguistic Analysis* 33(1). 54–96.

Bošković, Željko. 2007. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic inquiry* 38(4). 589–644.

Branigan, Phil & Marguerite MacKenzie. 2002. Altruism, ā-movement, and object agreement in Innu-Aimûn. *Linguistic Inquiry* 33(3). 385–407.

Bruening, Benjamin. 2001. *Syntax at the edge: Cross-clausal phenomena and the syntax of Passamaquoddy*: Massachusetts Institute of Technology dissertation.

Butt, Miriam. 1993. Object specificity and agreement in Hindi/Urdu. In *29th regional meeting of the chicago linguistic society*, 80–103. Chicago Linguistics Society.

Butt, Miriam & Tracy Holloway King. 2004. The status of case. In Veneeta Dayal & Anoop Mahajan (eds.), *Clause structure in south asian languages*, 153–198. Dordrecht: Kluwer.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. 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. 2004. Beyond Explanatory Adequacy. In Adriana Belletti (ed.), *Structures and beyond: The cartography of syntactic structures, volume 3*, Oxford University Press.

Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces+ recursion= language?*, 1–30. Berlin: de Gruyter.

Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos Otero & Maria Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2012. Foreword. In Ángel J Gallego (ed.), *Phases: Developing the framework*, 1–7. Walter De Gruyter.

Chomsky, Noam, Ángel J Gallego & Dennis Ott. 2019. Generative grammar and the faculty of language: Insights, questions, and challenges. *Catalan Journal of Linguistics* 229–261.

Davison, Alice. 1991. Feature percolation and agreement in Hindi-Urdu. *Panel on Agreement in South Asian Languages at the SALA Conference* .

Davison, Alice. 2004. Non-nominative subjects in Hindi/Urdu: VP structure and case parameters. In Peri Bhaskararao & Karumuri Venkata Subbarao (eds.), *Non-nominative subjects, Vol. 1*, 141–168. Amsterdam: John Benjamins.

Davison, Alice. 2014. Non-finite complements and modality in *de-na* 'allow' in Hindi-Urdu. *Natural Language & Linguistic Theory* 32(1). 137–164.

Diesing, Molly. 1992. *Indefinites*. MIT Press.

Epstein, SD, H Kitahara & TD Seely. 2009. The necessity, but invisibility of counter-cyclic outputs: Deducing extraction constraints and transfer-application from 3rd factor conditions on language design. *Glow Newsletter* 34–36.

Fox, Danny & David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31. 1–45.

Grosz, Patrick & Pritty Patel. 2006. Long distance agreement and restructuring predicates in Kutchi Gujarati. *Ms., University of Vienna and UC London* .

Harley, Heidi. 2008. On the causative construction. In Shigeru Miyagawa & Mamoru Saito (eds.), *The handbook of Japanese linguistics*, 20–53. Oxford University Press.

Kachru, Yamuna. 2006. *Hindi*. John Benjamins Publishing.

Kalin, Laura. 2014. *Aspect and argument licensing in Neo-Aramaic*: University of California, Los Angeles dissertation.

Kalin, Laura. 2018. Licensing and differential object marking: The view from Neo-Aramaic. *Syntax* 21(2). 112–159.

Kalin, Laura & Philipp Weisser. 2019. Asymmetric DOM in coordination: A problem for movement-based approaches. *Linguistic Inquiry* 50(3). 662–676.

Keine, Stefan. 2007. Reanalysing Hindi split ergativity as a morphological phenomenon. *Linguistische Arbeits Berichte* 85. 73–127.

Keine, Stefan. 2013. On the role of movement in Hindi/Urdu long-distance agreement. In *Proceedings of nels*, vol. 42, 273–284.

Keine, Stefan. 2020. *Probes and their horizons*. MIT Press.

Kidwai, Sana. 2022. The Urdu active impersonal. In Samir Alam, Yash Sinha & Sadhwi Srinivas (eds.), *Proceedings of formal approaches to South Asian languages*, vol. 11, .

Larson, Richard K. 1988. Scope and comparatives. *Linguistics and philosophy* 1–26.

Legate, Julie Anne. 2003. Some interface properties of the phase. *Linguistic inquiry* 506–516.

Mahajan, Anoop. 1989. Agreement and agreement phrases. *MIT working papers in linguistics* 10. 217–252.

Mahajan, Anoop. 1990. *The A/A-bar distinction and movement theory*: Massachusetts Institute of Technology dissertation.

Mahajan, Anoop. 2017a. Accusative and ergative in Hindi. In *The oxford handbook of ergativity*, 86–108. Oxford.

Mahajan, Anoop. 2017b. Locality in Hindi agreement. *Journal of Indian Linguistics* 78(1-2). 75–92.

Mohanan, Tara. 1994. *Argument structure in Hindi*. Center for the Study of Language (CSLI).

Montaut, Annie. 2018. The rise of differential object marking in Hindi and related languages. *Diachrony of differential argument marking* 281–313.

Obata, Miki. 2010. *Root, successive-cyclic and feature-splitting internal merge: Implications for feature-inheritance and transfer*: University of Michigan, Ann Arbor dissertation.

Obata, Miki. 2017. Is transfer strong enough to affect labels. In Leah Bauke & Andreas Blümel (eds.), *Labels and roots*, 117–126. De Gruyter Mouton Berlin & Boston.

Ott, Dennis. 2011. A note on free relative clauses in the theory of phases. *Linguistic Inquiry* 42(1). 183–192.

Polinsky, Maria & Eric Potsdam. 2001. Long-distance agreement and topic in Tsez. *Natural Language & Linguistic Theory* 19(3). 583–646.

Preminger, Omer. 2014. *Agreement and its failures*. MIT press.

Pylkkänen, Liina. 2008. *Introducing arguments*. MIT press.

Four Puzzles and Affixal N in Hindi

Hi Sunt Dracones 'Here be dragons' (Hunt-Lenox Globe, c.1510 CE, eastern Asian coast)

R. AMRITAVALLI, *The English & Foreign Languages University, India*

ABSTRACT

Four puzzles in Hindi morphosyntax are "default agreement," an "oblique stem," suppletion in the nominative plural, and number-case synthesis. These phenomena are not unrelated. I analyze T's φ-features as nominal category features: N (the so-called "gender φ-feature(s)"), Num, and D, or Person. The default value *-aa* of the categorial feature N ("Affixal N," "N-stem") is reflexively overt in Hindi. "Default agreement," which surfaces in the absence of Agree, is this affixal N at T. Taking Case as a categorial feature, I argue that the nominative case feature is minimally [N-on-T]. N surfacing as *-aa* on T without Agree, an *-aa*-marked subject noun must check [N-on-T] by "reverse" Agree to get nominative case; N-*aa* becomes a "nominative stem" (N-*ee* is the 'elsewhere' stem). The Hindi Number suffixes are "portmanteaux," not "syncretic." NUM is a nasal feature spelt out on a structural Case feature: on [N-on-T] (spelt out *aa*) in the nominative, and elsewhere, on [V], the accusative case-feature common to all oblique cases (spelt out *-oo*). This explains the various plural suffix-shapes, and why plural agreement on V manifests as just a [nasal] feature (as NUM has no vowel *-aa* prior to nominative case assignment, the vowel in the NOM PL suffix *-ãã* is absent in agreement). The form expected as the NOM.M.PL noun is *N-*aa-ãã*. This form suffers spell-out failure, and suppletion occurs. This analysis of Hindi Number and Case explains the near-universal silence of nominative case as a 'direct' or self-licensing case that manifests only in "φ-agreement" at T.

1 Introduction

Hindi morphosyntax presents puzzles, not unique to this language, such as "default agreement," a putative "oblique stem" (N-*ee*), the morphological synthesis of number and case, and a suppletive nominative plural (N-*ee* again), distinct from the oblique plural N-*õõ*. I analyze these four phenomena as consequences of a property of Hindi "affixal N" (to use Pesetsky's (2013) term). I argue that the "gender phi-feature(s)" of N are affixal N, i.e., N-stems or N-category feature(s). Taking Case as a syntactic category feature, the "uninterpretable phi-feature of gender" at T is a formal, N-category feature, and the "uninterpretable phi-features" of person and number at T are the categorial features D and Num. These nominal features make up the "nominative case feature" (*pace* Pesetsky 2013), but only N among these "*u*phi" or categorial features is inevitable, so nominative case is minimally [N-on-T].

In Hindi, the N-category feature has a default, reflexive, spell out *-aa*, traditionally termed the "3M.SG." gender-number feature. Gerundive nominalizations have an overt *-aa*. Here, *-aa* is not masculine or singular, but a recognizable reflex of nominalization (an overt categorial feature N with a default value), the default spell out of affixal N. The reflexive overtness of default affixal N has three consequences: "default agreement" *-aa*, a

"nominative stem" N-*aa* (with N-*ee* an "elsewhere" form), and suppletion in the nominative masculine plural. What is spelt out at T as its "gender phi-feature" is, as in the gerund, an affixal N whose default value -*aa* appears on the verb as putative "3MSG default agreement" in the absence of Agree. The N at T, being valued by default as -*aa*, fails to probe for an identical "3MSG" subject N-*aa*. Therefore, a subject N-*aa* cannot get nominative case except by "reverse Agree" with T. Thus, -*aa* must be specified with an uninterpretable [*u*N-on-T] feature (and not an [αCase] feature). This makes N-*aa* a "nominative stem" and necessitates an "elsewhere" stem N-*ee* (the putative "oblique stem").

Turning next to the Hindi plurals, these suffixes are known to vary by case (in the nominative and non-nominative paradigms), but their various forms are thought to be arbitrary. Assuming nominative and accusative case to be the categorial features N and V, Hindi plural suffixes are seen to be portmanteaux spell-outs of NUM's lexically specified nasal feature on the structural case feature-vowels -*aa* (*aa* being the default N vowel) and -*oo.* As for suppletion, the expected nominative plural suffix is -*ãã*. When -*ãã* suffixes to N-*aa*, the resulting sequence *N-*aa-ãã* (I claim) fails to be spelt out (perhaps due to feature-conflict). The elsewhere stem N-*ee* occurs as a suppletive plural, supporting a view of suppletion as occurring where the regular affix fails (Kayne 2020).

The paper is structured as follows. Sections 2 and 3 argue that the gender features of N are its categorial features, and resolve the problem of the "two morphemes" -*ee* by positing -*aa* as the dedicated nominative (masculine) stem (*pace* McFadden 2018). Sections 4-6 motivate the nominative stem in terms of "default agreement," and explicate the nature of nominative case. Sections 7-8 specify the N-features and illustrate the absence of number concord in Hindi, and 9-10 turn to Hindi Number. I motivate the various shapes of the regular plural suffix (supporting the argument with facts from agreement), and offer an account of NOM.M.PL suppletion based on Kayne (2020) and Collins and Kayne (2020). Sections 11-12 conclude with the implications of the analysis for theorizing number-case syncretism, and locating the case hierarchy vis-à-vis the nominal functional sequence.

2 Affixal N: Gender phi-features are N-category features

The noun-root in Hindi must assume one of three canonical noun shapes ending in -*aa*, -*ee* or -*ii*, in the syntax. E.g., a root *laṛk-* surfaces with -*aa*, -*ee*, or -*ii* in the nouns *laṛk-aa*, *laṛk-ii*, *laṛk-ee*, usually glossed 'boy, M.SG.,' 'girl, F.SG.,' and 'boy, OBL.M.SG/ NOM.M.PL.' Such suffixes have been termed stem-formatives (Noyer 1992:14), noun-stem allomorphs (McFadden 2018), or even "theme vowels" (in Russian). In the syntax, where they appear in agreement and concord, they may be termed gender (and number) "phi features" that are thought to be interpretable on the noun. However, outside of human/ animate count nouns, these Hindi formatives are uninterpretable, and easily seen to be purely formal categorial features (or allo-features). Consider a class of productive nominalizations that correspond to the English infinitive/gerund paradigm. The Hindi deverbal nouns *ḍar-naa* 'to fear,' *aa-naa* 'com-ing,' end in -*aa* in the nominative paradigm (1a), and in -*ee* in the oblique (1b), but the suffixes are not interpretable as masculine or singular.

(1) a. ḍar-n-aa manaa hɛ / andar aa- n- aa manaa hɛ.
fear-NMNL-*aa* is forbidden inside come- NMNL-*aa* is forbidden
'To fear is forbidden / No entry.' [lit. 'Fearing/ coming in is forbidden.']
b. ḍar - n-ee see/ par … / andar aa- n- ee see/ par …
fear- NMNL-*ee* by/ on inside come -NMNL-*ee* by/ on
'By fearing /by coming in; on fearing /on coming in, …'

In (1), the "masculine gender" of the "phi-features" *aa* and *ee* can signal only that the nouns derived belong to a productive, unmarked subclass. Productive, regular deverbal nouns never carry the "feminine" N-feature *-ii*, but lexically idiosyncratic nominalizations do: e.g., the root *cal-* 'walk' has a regular masculine nominalization *cal-naa* 'to walk, walking,' and an idiosyncratic nominal with a (covert) feminine nominal feature in *caal-Ø* 'walk (walking style), behavior, trick.' Gender thus reflects 'genre,' or noun (sub)class, not a semantically interpretable feature. Hindi has grammatical gender. Harris (1991) too argued at length that the putative gender morphemes of Spanish were mere "word markers." So also, *aa*, *ee* are not "singular" in (1). (More evidence against *aa*, *ee*, *ii* as "singular" (*pace* Castillo 2001),[1] in (3) below, is their (covert) occurrence in the non-count/ mass nouns *paani-Ø* 'water,' *hawaa-Ø* 'breeze.')

Pesetsky (2013:4-5), proposing that Case is a syntactic category feature, refers to "the suggestion by Marantz (1997) and others" that words are formed by "categorizing morphemes" on category-neutral roots, and suggests that these are "affixal realizations" of "the various *parts of speech*" (his italics). Adopting this formulation, I shall say that Hindi *-aa*, *-ee*, *-ii* are affixal N, featural realizations of the syntactic category N, and refer to them as N-stems or N-features. Affixal categories may appear (Pesetsky suggests) either by the "lexical route" (attaching to a root), or by feature-sharing in syntax. I shall show that the Hindi N-features, though mandatory, are lexically silenceable on N-roots. However, they are obligatorily overt when they appear by feature-sharing in the syntax as concord on A (cf. (3) below).[2] I take the reflexive overtness of the default N-feature in (1) to be this same syntactic overtness of affixal N (taking productive derivational morphology as syntactic, not lexical), and extend the argument to affixal N at T that appears as agreement on V.

3 The "oblique stem" and the "two morphemes" *-ee*

Given the two "masculine" affixal N *-aa*, *-ee* in (1), is *-ee* "oblique," or *-aa* "nominative"? Blake (2001:10) cites the very Hindi nouns *laṛk-aa* 'boy, M.SG.,' *laṛk-ee* 'boy, OBL.M.SG,' to illustrate that N-stems in the world's languages may occur as "nominative, alternatively direct," or "oblique" (cf. also Mohanan 1994:61, and n. 8). The lexicalization of an oblique stem does not motivate the need for it. Notice that only masculine N in Hindi have an oblique stem. Feminine N occur unchanged in nominative and oblique contexts:

[1] Reported (in Ueda 2009:109) to designate Hindi *-aa* and *-ii* as Number morphemes, in support of a claim that "singular" may have a phonologically specified shape.
[2] Italian nouns derived from present participles (*patente* 'license,' *cantante* 'singer') are unmarked for gender, but trigger gender inflection on adjectives (Giusti 2011).

(2) a. laṛk-ii aay -ii.
girl-F.SG. came-F.SG.
'The girl came.'
b. laṛk-ii -koo deekh-oo.
girl-F.SG. -ACC. see- IMP.
'See the girl.'

In the absence of a general distinction between "direct" and "oblique" stems in Hindi, it may well be that it is N-*aa* that is a "nominative stem" prohibited in oblique contexts and N-*ee* an elsewhere stem (a possibility McFadden (2018) briefly considers but rejects on putative markedness grounds for Finnish-type languages). Designating -*ee* oblique results in an idiosyncratic "homonymy" of the oblique singular and nominative plural morphemes -*ee* (under a structuralist view of the morpheme) that is both uneconomical and circular. Consider covert masculine nouns like *bandar*-Ø 'monkey,' *paani*-Ø 'water,' that trigger overt -*aa* (M.SG) and -*ee* (OBL.M.SG/ NOM.M.PL.) concord in the expected way:

(3) a. gand-*aa* bandar-Ø / gand-*aa* paani-Ø aa-yaa
dirty-*aa* monkey-Ø dirty-*aa* water-Ø come 3M.SG.
'The dirty monkey came/ Dirty water came.'
b. gand-*ee* bandar-Ø/ gand-*ee* paani-Ø -koo mat chu-oo/-naa
dirty-*ee* monkey-Ø dirty-*ee* water-Ø -ACC. don't touch-IMP.
'Don't touch the dirty monkey/ the dirty water.'
c. gand-*ee* bandar-Ø aa-yee
dirty-*ee* monkey-Ø come-3M.PL.
'The dirty monkeys came.'

Assuming that a late rule silences the N-feature on these nouns, if -*ee* is OBL.SG in (3b) but NOM.PL. in (3c), -*ee* silencing must be specified to apply to two unrelated morphemes. Worse, the across-the-board homonymy of these putative morphemes means that only by function and distribution are they distinguishable in the first place, not morphologically. Entertaining, therefore, the view that it is -*aa* that is restricted to the nominative, I now turn to the question of why Hindi might have a dedicated nominative N-stem -*aa*.

4 "Default Agreement:" A Misnomer for Agree Failure

Descriptively, "default agreement" is an -*aa* that appears on the Hindi verb in the absence of Agree, without T feature-checking an argument and assigning nominative case. The identification of legitimate "default agreement," in fact, depends on the impossibility of Agree: cf. Preminger's (2009) "characterization" of "the relation between phi-agreement and (un)grammaticality":

(4) (Preminger's (58)) "*You can fail, but you must try*": Applying ϕ-agreement to a given structure is obligatory; but if the structure happens to be such that ϕ-agreement cannot culminate successfully, this is an acceptable outcome.

"Agreement" is thus a misnomer for the *-aa* that surfaces on V in (i) ergative clauses with an overtly *k-oo*-marked object (cf. (5a), where T can agree with neither the object nor the ergative subject); and (ii) in passive constructions where the verb carries passive morphology but the thematic object retains its *-koo* marker, (5b). Observe the mismatch of the *aa*-marked verb with the arguments in these examples.

(5) a. mɛ̃-nee rooṭi -yõõ -koo khaa.Ø(y) -aa
I-ERG. roti.F -OBL.PL -ACC. eat.PERF. -3M.SG.
'I ate the rotis.'

b. bacc-õõ -koo chaḍi -see maar.Ø-aa jaa.t -aa thaa.
child-OBL.PL-ACC. stick-INSTR. hit.PERF-3M.SG. PASS.IMPF.-3M.SG. BE.PERF.3M.SG.
'Children were beaten (=used to be beaten) with sticks.'

Anand and Nevins (2006:19), pointing out in passing the problem "default agreement" poses for classical Agree theory, assume a unique head $T_{checked}$ specific to Hindi, valued for the "3M.SG." feature. But what has checked it? The supposition is that T has an uninterpretable phi-feature set that can surface only if its value is "checked" by a corresponding interpretable occurrence. I have shown that gender on Hindi N is not an interpretable but an N-feature; and that *-aa* is a spell out of affixal N that appears reflexively in productive deverbal nominalisation in (1a). I suggest that "default agreement" *-aa* is a similar reflexively spelt out N-feature on T (its "gender phi-feature").

Here I must recapitulate two points from Amritavalli (2019) about the nature of Hindi "T," and feature-checking by Hindi "T." Hindi T is a feature complex [$T_{PERSON.NUM}$ - $P_{TCPL.NUM.GEN.}$], and the "phi-features" are not checked as a bundle. A sentence 'I run' in Hindi surfaces as 'I be.PERSON.NUM. run.IMPF.PTCPL-NUM.GEN.' T has a dummy verb *be* marked for person and number, and the verb is a participle (in nominative as well as ergative clauses). T probes only for Person. The PTCPL head probes for and values the NUM feature; this value is specified also at T, under a feature-sharing version of Agree (Pesetsky and Torrego 2007). The gender or N-feature remains at PTCPL. In (5), thus, *-aa* appears on the perfect participle 'eaten' in (5a), and in (5b), on the imperfect (habitual) passive participle 'going' (= "getting") (and the perfect participles 'hit,' 'been'). (The imperfect (non-past) participial inflection is *-t-*; I take the perfect (past) participial inflection to be null.) For a participle to have an N-feature (i.e., categorially be nominal as well as verbal) is perhaps natural; participles in English, e.g., function as adjectives, long thought to featurally be [+N, +V]. The participles in (5) are unpronounceable without their N-feature inflection.

Default agreement is thus a default, unchecked [[*u*N]-on-T] feature spelt out *-aa* on the verb that raises to pick up its participial inflection. The unchecked N-feature neither simply deletes (as has sometimes been suggested), nor stays silent.

Note that "default agreement *-aa*," conventionally notated "3M.SG.," is solely an N-feature. The informal notation "3" for Person indicates a lack of a value for Person (taking 3rd person as the absence of person, Harley and Ritter 2002). As for Number, *-aa* (we said) does not necessarily lexicalise a "singular" feature (appearing as it does on non-count N and on deverbal nominalizations). "Singular morphology" is a default that includes absence of Number (Pesetsky 2013). Therefore, Person and Number are valued at T only by Agree

with an argument that specifies values for them; else, they are absent at T. "Masculine gender," the default value of N, is not, however, the absence of N, but a default feature-subclass of N. Among the "phi-features," then, the categorial N feature is always and at least, present at the T-complex; and [[*u*N]-on-T] is spelt out on the verb with a default value even when nothing values it, and no nominative case is assigned.

5 Default *-aa* versus checked *-aa*: a problem

What differentiates "default *-aa*" from a "checked *-aa*" at T? To put it differently, what prevents [[*u*N]-on-T] from always surfacing by default as *-aa*, independent of Agree?

Neither Anand and Nevins (2006) nor Preminger (2009) frontally address this question of how to rule out illegitimate "default agreement," as their primary concern is to "rule in" or allow default agreement in structures where the goal is inaccessible, cf. (4): "if … ϕ-agreement cannot culminate successfully, this is an acceptable outcome." Preminger offers an injunction, "You can fail, but you must try." What enforces the injunction to "try"? It must be the probe's need to value its unvalued features.

Consider now three scenarios where the Hindi [N on T] probes the subject for its value, assuming a classical account of nominative case as a "reflexive" checking of the subject's *u*Case feature contingent on T's checking its "*u*phi" features with the subject's "*i*phi" features. (i) The subject N-stem has the specified feature values "feminine" or "suppletive M.PL." The goal's specified value is specified at T, spelt out *-ii* or *-ee* (as "gender-(number) agreement"), and case is assigned. (As T's "phi-features" are checked individually in Hindi, each corresponding "*i*phi" nominal head presumably has its case feature checked.) (ii) The goal is *ko-* marked (as in ergatives and passives). Agree tries and fails, and [*u*N] surfaces with its default categorial value *-aa* (as per Preminger). (iii) The subject is "masculine," and its N-stem is spelt out *-aa*. For concreteness, assume a mass noun subject as in *paaniØ beht-aa hɛ* 'Water(.M) flows.M' (with *-aa* lexically silent but appearing by feature-sharing in agreement), where "gender" is the only feature of the subject that could be valued. As *-aa* is the default value of affixal N, the subject N-*aa* has no specified feature values (assuming privative features, as in Nanosyntax; in a binary feature system, it would have unmarked, i.e. unspecified, feature values). When [*u*N] on-T (which has no specified values) probes for a value for itself, it encounters in the goal the identical N with no specified feature values. Therefore, no feature-valuation could occur. Agree would thus fail, but not obviously, because unvalued N in the syntax is spelt out by default in the absence of valuation. Thus, [*u*N] on-T will nevertheless be independently spelt out *-aa*, as "default agreement." But if Agree does not happen, how is an N-*aa* in subject position licensed as nominative?

It seems that Preminger's imposition of an additional "if … then" logic on Agree in (4), which works for scenario (ii), must be strengthened for scenario (iii) by a "match" contingency on the probe: "if you find an identical value on the goal as on yourself, check yourself and assign Nominative anyway; *only if the goal is inaccessible* may you spell out your N-feature without assigning case, as default agreement." Thus, Agree must be complicated in more than one respect (with an *if-then* logic, plus a *match* condition) to distinguish "checked *-aa*" from "default *-aa*." The same problem arises with Anand and

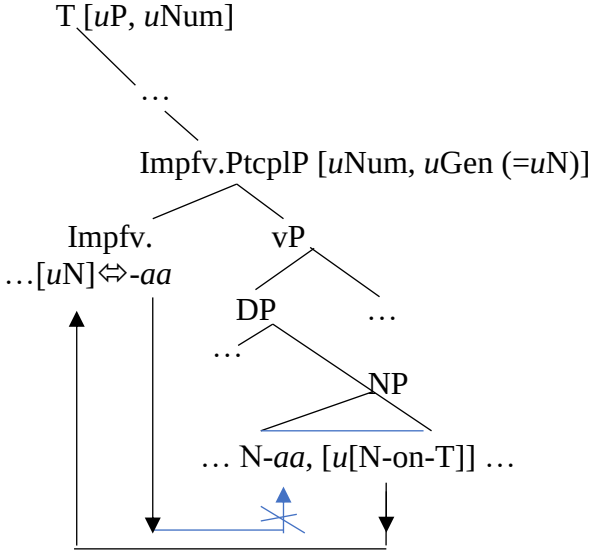
Nevins' $T_{checked}$. If $T_{checked}$ merges where phi-agreement with an N-*aa* subject is possible, Agree becomes illegitimately inoperative, but with no surface reflex of illegitimacy.

What the Hindi data tells us is that it is our premise that "default agreement" is a form of agreement that is wrong. It is irrational to modify Agree to account for a phenomenon that arises in the absence of Agree. If "default agreement" is not a value specified by feature-sharing by agreement, but (as I suggest) is an unchecked, default N-stem value reflexively spelt out in the syntax on T (as it is on gerundive nominals), the problem, rather, is how to ensure that a subject with an N-*aa* stem gets nominative case, in the face of a possible occurrence of this very N-stem value on T as "default agreement."

6 Nominative case: its assignment, and its silence

I suggest that the solution is "reverse" agree, adapting (modifying) a proposal for reciprocal checking in Pesetsky & Torrego (2001). The problem, and the proposed solution, are illustrated in (6), adapted from Amritavalli's (2019) example (12).

(6)



In (6), the IMPFV. PTCPL's [[*u*N]-on-T] probes its goal (the subject) for its N or "gender" value. The subject has an N-*aa* stem, and "gender" valuation fails. However, the subject N-stem -*aa* is endowed with a case feature. I assume the case feature here to be not a general feature [*u*/*α*Case], but a specific nominative case feature [*u* [N-on-T]]. This feature independently probes for and checks its categorial feature at T. N-*aa* licenses itself, but as a result of its [*u* [N-on-T]] case feature, N-*aa* becomes a "nominative stem."

This proposal for "reverse" agree differs from the Pesetsky-Torrego proposal in its assumptions. (i) Nominative case is not [*u*T on D] (uninterpretable Tense on D), but the subject nominal's own categorial values on T. (ii) Thus, nominative case assignment is not a reciprocal exchange of uninterpretable T and D features, and "reverse" agree is not

dependent on "phi-feature valuation by T." [3] (iii) Not all N-stems have a [*u* [N-on-T]] nominative case feature (this would restrict them all to the nominative projection.) N in general has only a *u/α*Case feature. (iii) Case is a syntactic category feature, primarily a structural-case feature: either the nominative case feature(s), or the accusative case feature, common to all oblique cases (Caha 2009).

We may now understand the "*u*T" feature (which Pesetsky and Torrego admit they find no spell out evidence for at D) to be the set of nominal categorial features {D, Num, and N} which are present at T and valued in the subject DP, but must be valued at T. The subject's presence is identified at T by its own categorial features. This explains (i) in what sense nominative is a "direct" case (no alien case-head assigns its syntactic case feature to the DP, which licenses itself on T); (ii) the apparent "silence" of Nominative case at the DP (it should morphologically be an iteration at the DP of its own categorial features, but their iterated spell-out in the same projection appears to be suppressed), and (iii) conversely, the ubiquitousness of "phi-feature agreement," i.e., valued DP-features-on-T which appear on the inflected verb (often considered an "imperfection" in language, an ill-understood phenomenon that has more recently been termed "a core case of syntactic doubling," Barbiers 2008:28).

Hindi nominal morphology is particularly transparent in that the features of the syntactic category N are expressed on the noun. In some languages, N-features are silent on the noun (as they may be in Hindi as well; perhaps because heads prefer to be silent, Kayne 2016). In languages without concord, N-features may appear only in verb "agreement." Even in agreement, the categorial N-feature is the most likely to be silent. In the Greenbergian implicational hierarchy, gender agreement is said to depend on the incidence of number agreement (Harley and Ritter 2002). i.e., among the phi-features Person, Number and N/Gender (conceptually, "deixis, countability and taxonomy" features, Harley and Ritter 2002), the "taxonomy" or N-categorial features are the least likely to appear, even though, in my analysis, only N is inevitably projected in the nominal expression and at T. The pervasive silence of the N-category feature is perhaps why "phi-agreement" has been a construct that makes no reference to categorial features.

To sum up this section, Agree between the default [N-on-T] and an N-*aa* subject can be enforced (for the purposes of nominative case assignment) by shifting the burden of checking from the N-probe at T to the goal N, by specifying a [*u*[N-on-T]] feature on -*aa*. "Default agreement" at T thus entails a dedicated "nominative stem."

7 Specifying the N-categorial features

Prior to considering the other two puzzles, I attempt to explicate the feature specifications of the Hindi N-stems. I assume that syntax puts together individual features by the operation Merge. In Nanosyntax (Starke 2009, Caha 2009), the lexicon contains sub-trees.

[3] Zeijlstra (2012: section 3.1) suggests that "(n)othing would go principally wrong if the case-checking relation was the primary one and φ-agreement secondary" in subject-verb agreement, and that upward agree "(r)evers(es) the Agree relation." Indeed, these may be just two independent mechanisms (of case-licensing and case-specification). (My proposal is compatible with upward Agree, and an assumption that a syntactic category feature on an "alien" syntactic category is in some sense "uninterpretable," or alien.)

A lexical tree matches a syntactic node if it contains the syntactic tree (the Superset principle), subject to the Elsewhere ("minimise junk") principle for competing entries, and to Cyclic override ("biggest wins"). A morpheme in Nanosyntax is a lexical sub-tree of syntactic features, paired with conceptual and phonological features (Baunaz and Lander 2018).

We noted that feminine N have only one stem N-*ii*, but masculine N have two: N-*aa*, N-*ee*. "Gender" (genre) first splits the N-category into unmarked ("masculine") and marked ("feminine") sub-classes; only the unmarked subclass undergoes further feature specification. Assuming privative features, "unmarked" is "featurally/ structurally simplest." In (7), [FEM] is specified, and [MASC] is just "affixal N."

(7) i. /-aa/ <=> N ii. /-ee/ <=> N iii. /-ii/ <=> FEM
N

What differentiates the N-stems *-aa* and *-ee*? N-*aa* is never plural. N-*ee* is "singular-only" in the oblique, and "suppletive plural-only" in the nominative projection.

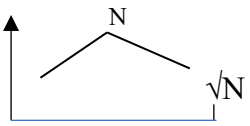
We referred to Pesetsky's (2013) characterization of "singular morphology" as "'elsewhere' number (singular or absence of NBR)." I suggest Hindi *-aa* is a default, elsewhere number form that includes both [SINGULAR] and "absence of NBR," whereas *-ee* is specified for "absence of NBR:" it is "numberless." Borer (2005) suggests that all nominal denotations are mass denotations, and a specific structure is projected that makes nouns "count." However, a putative feature [Count] could not formally sub-classify Hindi nouns; it is a contentful semantic feature that may occur in nouns in all three formal subcategories of (7). Assuming [Count] to be a conceptual feature, I require Count nouns to have a specified value "singular" or "plural" in the syntax (non-Count nouns being semantically illegible if so specified). The discussion of number that follows refers only to Count nouns.

The feature [plural] for N is introduced as the head of a functional category NumberP (Ritter 1991, 1993). For suppletive plurals, a "low" Number, in addition to a regular "high" Number, has been motivated in the literature (Collins and Kayne 2020, Pesetsky 2013). Pesetsky proposes that a "numberless," morphologically singular, Russian noun becomes plural by "immediately" merging in the syntax with a paucal numeral, which is a "free-standing instance of NBR." (This explains its triggering plural agreement on the adjective it merges with.) I require Hindi "numberless" N-*ee*, if it is a count noun, to be specified in the syntax for a number value [SINGULAR] or [PLURAL]. I, thus, first add an entry (8iii) for the feature [SINGULAR], which has no phonetic realization, and modify the lexical entries (7i) and (7ii) for *-aa* and *-ee* as (8i) and (8ii).

(8) i. /-aa/ <=> N ii. /-ee/ <=> NBRLESS iii. [SINGULAR] <=> Ø
N

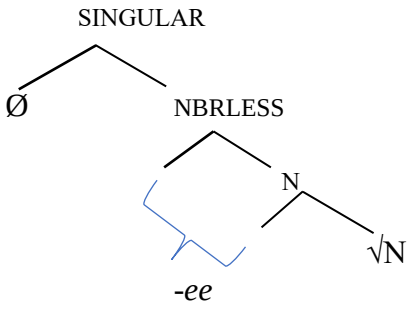
Let us say that the syntax projects the subtree (9). Both *-aa* and *-ee* qualify for insertion under N: (8i) is an exact fit, and (8ii) is a superset. (For spell out of N, √N moves out and up.)

(9)



By the Elsewhere (minimize junk) principle, (8i) *-aa* wins. This is what we want in the nominative projection (for non-plural, nominative, masculine N to be spelt out N-*aa* and not N-*ee*). In the oblique projection, however, a derivation where *-aa* merges will ultimately not converge, because of *-aa'*s [*u* [N-on-T]] feature. This allows an alternate derivation where numberless *-ee* merges, and the singular feature is separately spelt out (cf. (10)).[4]

(10)



How does *-ee* occur in the nominative projection, if it never matches the subtree (9)? Nothing prevents the syntax from projecting a structure that matches the lexical subtree (8ii). This structure is subject to a condition that a numberless count N needs to be specified for number (e.g., as in (10)). In the nominative projection, however, the independent singular feature is not available to *-ee* (as (9) privileges (8i)); but an independent (suppletive) plural feature is. I now address the question where the suppletive plural feature might merge with *-ee* by describing Number concord facts in Hindi, and return to *-ee* suppletion in section 10.2.

8 Number Concord and the Suppletive Plural in Hindi

Suppletion is often equated with idiosyncratic "lexical specification," but I adopt Collins and Kayne's (2020) suggestion that the suppletive plural is an "inner plural" PL1 that an "outer plural," the regular plural PL2, can select. Thus, PL1 is "lower than" the regular PL2. Where precisely PL1 merges seems vary in languages. In Russian, the paucal numerals trigger plural concord, so Pesetsky requires a "numberless" N to "immediately merge(s)" with a "free-standing instance" of Number (the plural projection) in the syntax.

[4] Sinha (2018:7) similarly underspecifies *-ee* for number (in a DM framework, his 13). He specifies *-aa* as [M.SG] (his (12)), but not "for direct case," because *-aa* occurs also in verb agreement, and "verbal agreement does not involve case features." To make *-aa* ineligible for the oblique projection, Sinha needs an impoverishment rule that deletes the singular feature in the oblique, "leading to the insertion of the underspecified *-e*" (his (14), [singular, oblique] → [oblique]). On my account, *-aa, -ee,* and *-ii* all have a case feature, but *-aa's* is more specific than [*u/α*case]; so that the other N-features may, but *-aa* must, check nominative as its case feature.

Hindi, however, has no number concord. Concord is with the N-stems *-aa*, *-ee*, and *-ii*. If a suppletive plural feature merges with *-ee* before an adjective does, *-ee* concord in the nominative would be (an exceptional instance of) "plural concord." Not only that: because *-ee* concord in the oblique is not plural concord, the "two *-ee*s" that I argue against would re-incarnated. I shall thus say that Hindi PL1 and PL2 both select the NP, i.e., the layer of the DP where N-stem concord takes place.

We have seen concord in the DP and the KP (adopting the term in Bittner and Hale 1996) in example (3), repeated as (11).

(11) a. gand-*aa* bandar-Ø / gand-*aa* paani-Ø aa-yaa
dirty-*aa* monkey-Ø dirty-*aa* water-Ø come 3M.SG.
'The dirty monkey came/ Dirty water came.'
b. gand-*ee* bandar-Ø/ gand-*ee* paani-Ø -koo mat chu-oo
dirty-*ee* monkey-Ø dirty-*ee* water-Ø -ACC. do not touch-IMP.
'Don't touch the dirty monkey/ the dirty water.'
c. gand-*ee* bandar-Ø aa-yee
dirty-*ee* monkey-Ø come-3M.PL.
'The dirty monkeys came.'

Adjectives (and the possessive phrase) show *-aa* concord with count (singular) and mass nouns in the DP (11a). In the oblique (KP), these same nouns trigger *-ee* concord (11b). But *-ee* concord also occurs in the DP (11c), and here it looks like "number concord," because N-*ee* occurs in the DP only as a masculine plural count noun. However, concord with a putative suppletive plural in (11c) is the sole instance of putative Number concord in Hindi. The regular Number suffixes do not trigger concord. In the DP (12), the feminine plural suffix on N-*ii* is *-ãã*, and on N-Ø, *-ẽẽ*; but the concording morpheme on the adjective is the N-stem *-ii* in (12a), and (12b) is illicit.

(12) a. us-k -ii acch -ii laṛk-iy-ãã / ããkh-Ø-ẽẽ
3P.-GEN.STEM-*ii* good-*ii* girl- F. -PL. (*ãã*) / eye -Ø-F.PL. (*ẽẽ*)
'His/ her good girls / good eyes'
b. *us-k-iy-ãã acch-iy-ãã laṛk-iy-ãã / *us-k-ẽẽ acch-ẽẽ ããkh-Ø-ẽẽ

In the KP (13), the plural suffix is *-õõ*, regardless of N's gender. The noun's modifiers do not show *õõ*-concord, but *-ee* concord or *ii*-concord, as in (13-14).

(13) a. us-k -ee acch-ee laṛk-õõ -koo/ bandar- õõ- k-oo deekh-oo.
3P.-GEN.STEM-*ee* good-*ee* boy-OBL.PL-ACC./monkey-OBL.PL-ACC. see-IMP.
'See his/ her good boys/ good monkeys.'
b. *us-k-õõ acch-õõ laṛk-õõ -koo/ bandar- õõ- k-oo deekh-oo.

(14) a. us-k -ii acch-ii laṛk-iy-õõ -koo deekh-oo.
3P.-GEN.STEM-*ii* good-*ii* girl-OBL.PL-ACC. see -IMP.
'See his/ her good girls.'

b. *us-k-õõ acch- õõ laṛk-iy-õõ -koo deekh-oo.

Thus, there is no concord with the regular plural suffix. Concord is with an N-stem *-aa*, *-ee* or *-ii*, and appears to be spelt out before N picks up its plural suffix. If *-ee* is marked a suppletive plural in the DP before concord occurs, *-ee* concord would be the sole instance of plural concord, and *-ee* concord with a suppletive plural (11c) would be different from *-ee* concord in the oblique (11b). (These data raise interesting questions that I do not enter into in this paper: is there a concord "phase" in the Hindi DP? Do Agree and Concord represent the same feature-sharing mechanism (Carstens 2000, Baker 2008, Norris 2011, Guisti 2011)?)

9 The Hindi plural morphemes: Number as a portmanteau morpheme

The Hindi plural suffixes evidently vary by case, but are thought to do so arbitrarily. Table 2 illustrates the Number vowel varying with Case; the four rows indicate noun subclass by gender, and overt/ covert spell out of the N-feature.

Table 2. Hindi singular and plural nouns, nominative and oblique

	Nominative		Oblique	
	Singular	Plural	Singular	Plural
Masculine 1	N-aa *laṛk-aa* 'boy'	N-ee *laṛk-ee*	N-ee *laṛk-ee*	N-õõ *laṛk-õõ*
Masculine 2	N-Ø *bandar* 'monkey'	N-Ø *Bandar*	N-Ø *bandar*	N-õõ *bandar-õõ*
Feminine 1	N-ii *laṛk-ii* 'girl'	N-ii-ãã *laṛk-iy-ãã*	N-ii *laṛk-ii*	N-ii-õõ *laṛk-iy-õõ*
Feminine 2	N-Ø *ããkh* 'eye'	N-ẽẽ *ããkh-ẽẽ*	N-Ø *ããkh*	N-õõ *ããkh-õõ*

The plural suffixes vary only with the two structural cases. The *-õõ* that occurs in the accusative KP in examples (13) and (14) occurs also in the genitive, ablative, ergative and vocative KPs (*bhai-(y)õõ aur behen-õõ*! 'Brothers and sisters!'). (Vocative case may be either direct or oblique in the world's languages, Hilda Koopman, p.c., 2014.) This follows from the accusative case feature being common to all oblique cases (Caha 2009). Caha does not specify the content of any of the case features, but Pesetsky (2013) does: accusative case is a V-feature. This explains an apparent coincidence of vowels in the Hindi accusative/ dative case marker *k-oo* and the oblique plural *-õõ*. The V-feature spells out the vowel *-oo* on the accusative/ dative case-stem *k-*, and in the oblique plural, the V-feature *-oo* hosts the nasal feature. Nominative case, I have argued, is [N-on-T], and masculine N is spelt out *-aa* or *-ee*. We may now notice that the overt vowels in the nominative vowels are *-aa* and *-ee*. In this analysis, case is transparently instantiated in the regular plural suffixes. The Hindi plural suffix specifies its vowel as a structural case feature nominative ("direct") or accusative ("oblique"). Only the nasal feature that rides on these case vowels

is lexically specified for Number, not the vowels themselves. The regular plurals are not merely "syncretic" but portmanteau case-and-number suffixes.

A word about genitive case is in order. In Caha's hierarchy, genitive intervenes between the accusative and the dative, but the latter two are syncretic in Hindi, entailing their adjacency.[5] Jayaseelan (2013, 2017: 518-520) points out that the Dravidian genitive is accepted as morphologically the "oblique stem," and illustrates that it occurs closer to the noun than dative/ accusative in Malayalam. He posits a hierarchy that manifests on the noun as GEN-ACC-DAT, i.e., the case hierarchy must allow for some variation.

Secondly, not all Hindi oblique cases overtly spell out the V feature on the case stem. Instrumental/ablative case *-see* and ergative *-nee* instantiate oblique as *-ee*, in perhaps a (yet-to-be-explained) spell out alternation *-ee/k-oo* seen also in the oblique pronouns, e.g., *us-ee, us-koo* (3SG.)*, mɛ̃-nee*, *mujh-koo* (1SG). Apparently, if the pronoun stem is "oblique" (incorporates a V-feature), the elsewhere N-feature *-ee* may suffix to it as the non-nominative case feature. The Hindi genitive clearly has both a V-feature and an N-feature. The V-feature manifests in the genitive pronominal stem (*us-*, not *wo*) and the occurrence of an oblique a case-stem *k-*. The N-feature manifests on the case head *k-* as agreement with the head N: *us-k-aa kutt-aa* (3OBL.-GEN.M. dog.M.) 'his/her/its dog,' *us-k-ii rooṭ-ii* (3OBL.-GEN.F. *roti*.F.) 'his/her/its *roti*.' (Possibly, an underlying V-feature is "elided," cf. Caha's (2013) discussion of "elision" of an underlying case during case-agreement.) Note also that Genitive is an oblique case in (i) Caha's case hierarchy, (ii) Dravidian: cf. Herur and Amritavalli (2022) for parallels between the Kannada genitive and the English *of*-genitive that do not hold for the English "Saxon genitive" *'s*, and Caha (2009:110) who notes that it is the English *of*-phrase that "shares the distribution of unambiguous genitives of other languages"), and (iii) in traditional analyses of Russian (mentioned by Pesetsky 2013, for whom Genitive is non-oblique, and the N-feature!).

10 The absence of the plural suffix vowels in verb agreement

In support of my claim that the plural suffix vowel is not part of the lexical spell out of Number, but a case vowel that gets specified on Number when Case is assigned, I now show that the number vowel of the nominative plural suffixes does not occur in the corresponding verbal number agreement morphemes. Subsequently, I attempt an account of the variety of nominative plural suffixes (section 10.1), and (in section 10.2) motivate suppletion in the nominative masculine as a failure of regular plural morphology in this paradigm (Kayne 2020, quoting a suggestion in Barbiers 2007).

Recall that only feminine nouns take the regular number suffix in the DP. Consider now number agreement with feminine plural nouns, comparing the singular agreeing forms of the verbs *uʈh-**ii,*** *khul-**ii,*** in (15a), with the corresponding plural forms *uʈh-**ĩĩ,*** and *khul-**ĩĩ*** in (15b). The only difference between (15a) and (15b) is the addition of nasalization in (15b) on the "feminine singular" stem vowel *-ii*. The plural suffixes in (15b) are *ãã* and *ẽẽ* (on 'girls' and 'eyes' respectively), but the vowels *aa* and *ee* do not occur in plural verb

[5] As an anonymous reviewer had kindly pointed out.

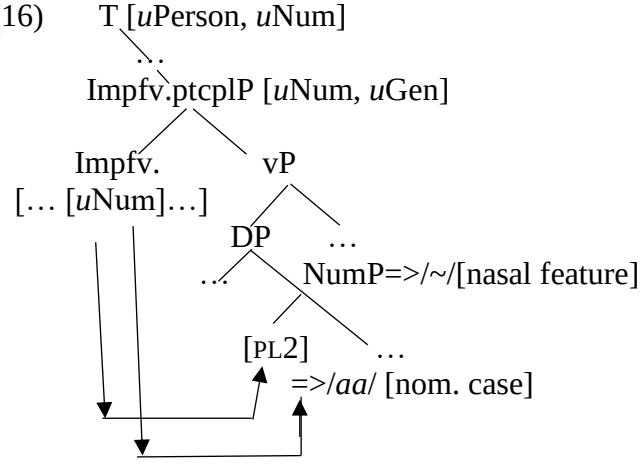
agreement; only the feature of nasality does, spelt out on the "singular" N-stem, as *ĩĩ*. Thus (15c), with surface identity of the plural suffixes and plural verb agreement, is illicit.

(15) a. us -k -ii acch-ii laṛk-**ii** uṭh **-ii** / ããkh-Ø khul- **ii.**
3P.-GEN.STEM-*ii* good-*ii* girl-F.SG. got up-F.SG/ eye.F.SG. opened-F.SG.
'His/ her good girl got up / His/ her good eye opened.'
b. us -k -ii acch -ii laṛk-**iy-ãã** uṭh **-ĩĩ / ããkh-Ø-ẽẽ** khul-**ĩĩ.**
3P.-GEN.STEM-*ii* good-*ii* girl-F.PL. got up-F.PL. eye- F.PL. opened-F.PL.
'His/ her good girls got up / His/ her good eyes opened.'
c.* us -k -ii acch -ii laṛk-**iy-ãã** uṭh- **iy-ãã /** ããkh-Ø-**ẽẽ** khul-Ø-**ẽẽ**.

Number agreement does not manifest the entire plural suffix shapes *(iy)-ãã* and (-Ø)-*ẽẽ*. It does not manifest the vowel of the plural suffix, but only a nasal feature, on the "feminine singular" verb. We noted earlier (section 8) the absence of number concord in Hindi. Plural agreement does occur in (15b), but it unpacks the plural suffix into a nasal feature and a vowel; in my analysis, it is a case vowel that is not integral to the Number morpheme. Nasality, on the other hand, is a robust feature of the Hindi regular plural (PL2), whether in the plural suffixes (oblique *-õõ*, "direct" *-ãã*, *-ẽẽ*), or in verb agreement *-ĩĩ* with the regular suffixes; whereas suppletive PL1 *-ee* has no nasal feature.

In (16), by the (classical) Agree algorithm for Hindi outlined in section 6, T's PTCPL complement successfully probes for a value for *u*Number (PL2), Agree takes place, and PL2 receives the nominative case feature [N-on-T]. N in the nominative case feature is invariantly spelt out as the default *aa*, and bears Num's lexical nasal feature.

(16)



Nominative Number is thus spelt out *ãã*, but the vowel is the outcome of nominative case assignment to the DP. It is not available when Num in the DP values Num at T; *u*Num's feature-valuation spells out only Number's lexical feature of nasality at T. In the spell out of plural verb agreement in (15), therefore, the nasal plural feature piggybacks on the N-feature "gender" vowel *-ii* that has been checked at ImpfvP.

Not all nominative plural suffixes in Table 1 are spelt out *ãã*, however. I now consider how the other spell outs of nominative number could occur.

10.1 The superficial shapes of the Nominative plurals

The nominative plurals vary according to N-features and their spell out: unlike the oblique plural, they are particular to the noun's gender and its overt/ covert expression. Nominative feminine plural suffixes have two shapes, *-ãã* or *-ẽẽ*, depending on whether the singular noun (*laṛk-ii, ããkh-Ø*) suppresses the spell out *ii* of its N-feature. In contrast, in the nominative masculine plural paradigm (17), the stem vowel is not pronounced on 'monkey' in (17b) in either the singular (N-*aa*) or the suppletive plural (N-*ee*). If Ø-stem nouns obey a simple rule "do not spell out the N-stem vowel," this would also delete PL1, the suppletive plural *-ee*, as a stem vowel, notwithstanding its function as a plural.

(17) a. laṛk-aa 'boy' ~ laṛk-ee 'boys'
b. bandar-Ø 'monkey' ~ bandar-Ø 'monkeys'

Why does the rule seem to not apply to feminine nouns like *ããkh-Ø*, to yield the (non-occurring) paradigm *ããkh-Ø, ããkh-Ø* 'eye, eyes' in (18)?

(18) *do not spell out N-stem vowel*: ããkh-~~-ii~~ =>Ø 'eye' ~ *ããkh~~-iy-ãã~~ => Ø

The plural in (18) is a regular number suffix PL2, spelt out as a nasal feature. If the deletion rule applies consistently, as expected, Num's nasal feature gets stranded, as both vowels *ii*, and *aa*, get deleted. A phonological repair rule thus seems to insert *-ee* as a dummy vowel for Num's nasal feature to manifest.

There are dialects of Hindi where the plural of *ããkh-Ø* does not undergo stem-vowel suppression, and surfaces with the expected plural suffix *-ãã*. For example, in classical Hindi poetry, the plural surfaces as *akh-iy-ãã*, suggesting that vowel length in the initial syllable is a factor in pronouncing *-ii*. In Dakkhini Hindi (on the other hand), null-marked masculine and feminine N both surface with *ãã* in the plural: *bandar(Ø)-ãã* 'monkeys,' *ããkh-(Ø)-ãã* 'eyes;' *admi-(Ø)-ãã* 'men, people,' *aurat-Ø-ãã* 'women;' *loog(Ø)-ãã* 'people,' *baat-Ø-ãã* 'words, speech,' arguing that the deletion rule is limited to stem vowels on roots in this dialect.

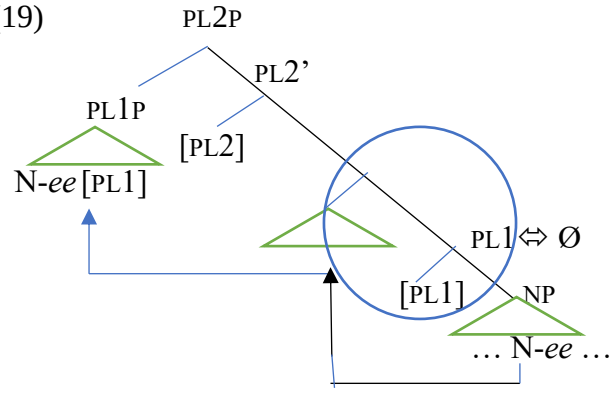
10.2. The suppletive masculine plural *-ee*

If "...suppletion applies when a regular morphological process is blocked for independent reasons..." (Kayne 2020, quoting Barbiers 2007), what is blocked in the NOM.M.PL, and why? What is blocked is **laṛk-aa-ãã* 'boys' (a default N feature on the N-stem and in the regular nominative plural suffix). (**Bandar-aa-ãã* 'monkeys' must also be blocked: if not, its derivation would parallel that of *ããkh* 'eye,' and yield **bandar-~~aa-ãã-ẽẽ~~* with the nasal feature of PL2.) Why is *-aa*$_{SG}$*-ãã*$_{NUM}$ blocked? I tentatively suggest a feature-clash when *-aa* with default number specification SG ("elsewhere number") morphologically combines with *-aa* spelt out under Num as a case vowel. Whereas *-ii*$_{SG}$ *-aa*$_{NUM}$ is fine, the vowel reiteration *-aa*$_{SG}$ *-aa*$_{NUM}$ appears to be illicit. Perhaps there is a general prohibition in languages against identical spell out of adjacent singular and plural feature values (or any

two different feature values).[6] (There is also a rule suppressing the reiterated spell out of the [N-on-T] case-feature *aa* on the N-stem, i.e., *N-*aa-aa*, *N-*ii-aa* in the singular, but it is unlikely that this is responsible for *N-*aa-ãã*.)

Given my assumptions about how N-*ee* merges in the syntax, an alternative derivation for Count noun as PL1 with this stem is available for masculine N. A Count noun is required to be marked singular or plural. N-*ee* in the nominative projection can combine only with a suppletive plural feature PL1, selected by PL2, as in (19).

(19)



In (19), N-*ee* merges with PL1 spelt out Ø, and moves up above this node for PL1 to be spelt out, as per Collins and Kayne (2020). PL1 moves up again above PL2, but PL2 is not spelt out on N-*ee* (neither its case vowel, nor its nasal feature, are spelt out: *N-*ee-ãã* (**lark-ee-ãã*), *N-*ẽẽ* (**lark- ẽẽ*). This is the difference between suppletive nominative masculine plural *-ee* and feminine plural *-ẽẽ* in *ããkh-Ø-ẽẽ*. In the latter, the number suffix is spelt out (but its vowel is deleted along with the stem vowel for this (class of) roots, and a resulting stranded nasal feature is supported by vowel insertion of *-ee*.)

Collins and Kayne (2020), motivating the PL1 - PL2 structure from Amharic (where an irregular plural suffix is spelt out inside the regular plural suffix), note that only PL1 is spelt out in English, for which they invoke a "No Crowding Condition:"

(20) (= their (27)) No Crowding Condition (relativized to formal features FF)
If X and Y both have FF, then if YP appears in the specifier of XP, X is not spelled-out.

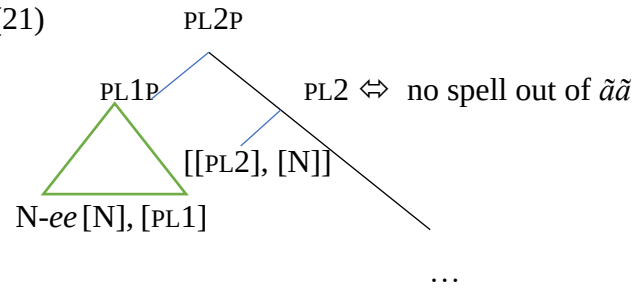
This spell out condition applies in Hindi. As shown in (21), N-*ee* is a lexical N-feature that merges with the plural feature PL1. PL2 is a lexical plural feature that receives an N-feature by nominative case assignment. Thus, PL1 and PL2 are featurally identical in their case and plural features, and PL2 stays silent.[7]

---

[6] Cf. *horses* (PL), *horse's* (GEN SG), *horses'* (**horses*-/iz/) (GEN PL).

[7] Conversely, singular *-ee* is absent in the oblique masculine plural: *lark-õõ*, **lark-ee-õõ*. Sinha (2018) explains this a general rule of *-ee* deletion preceding a back vowel (his (6)). Whether *ãã* deletion in the nominative plural, *-ee* deletion in the oblique plural, and the *-ee/ k-oo* alternation in oblique case spell-out noted in section (9), last paragraph, are three phenomena or facets of a single phenomenon is a question I leave open.

(21)



11. What Number-Case syncretism might tell us

That number and case syncretize is well-known (Blake 2001:18, Caha 2009: 73ff.). I have, for expository purposes, adopted a conservative Agree mechanism for the syncretism in the nominative, to show that the Hindi Number vowel is a case vowel. But there are larger questions at stake. Caha's (2009) syntactic account of case-syncretism (only adjacent nodes in the case hierarchy can syncretize) excludes number. He resorts to lexical specification of Finnish nominative-accusative plural *-t*, first spelling out Num, before its "phrasal lexical insertion."[8] For Hindi, this approach appears to require initial lexical specification of *ãã*, *õõ*, and *-ee* as Number, with some explanation offered for excluding the Number vowel from feminine plural verb agreement. This would miss out on a more general account of syncretism as restricted by a universal functional sequence.

Giusti (1995) suggested that case is part of the D-system in the DP. I have suggested (Amritavalli 2021) that the case hierarchy is integral with the functional heads Number and D(efiniteness) in the DP. Case and definiteness correlate in Hindi (Bhatt and Anagnostopoulou 1996), and in the Dravidian languages (which have "differential object marking"). In Bulgarian, where "full nouns bear only a nominative or accusative case suffix," "the distinction between nominative and accusative is made only with definite DPs" (Caha 2009: 34, and n. 22), reminiscent of Hindi. The Hindi absolutive object (termed nominative in Mohanan 1994) spells out Number as in the nominative, but does not tolerate pronouns, which must occur in the accusative; conversely, Hindi oblique plural *õõ* occurs only with overt ACC/DAT case *k-oo*; where *k-oo* fails to occur, so does *õõ* (Amritavalli 2019, 2021). It may be that objects in Bulgarian, can similarly appear in an absolutive case that is licensed by number and gender features only (excluding Person), whereas definite DPs require Person-licensing, as pronouns do, and appear as accusative.

12. Conclusion

I have tried to argue that the problem of the "two morphemes *-ee*" in Hindi has a bearing on four larger questions: (i) the nature of suppletion, (ii) the syncretization of Number and

[8] Baunaz and Lander assume adjacency of K and Num (in Hindi, Dem may intervene) to illustrate Latin *-ās* (ACC.FEM.PL.) target "the entire phrase [KP K [NumP Num]] … for spellout" (2018: sec. 3.1), claiming that phrasal spell makes it "possible to model portmanteau morphology as larger chunks of structure." This (and their German) example illustrates syncretism rather than a transparently portmanteau morpheme, and syncretism is again driven by lexical specification of its components.

Case in languages, (iii) the nature of Case, in particular, the structural cases nominative and oblique, and their occurrence in a hierarchy that may be integral with the functional hierarchy in the DP, and (iv) the nature of the categorial feature N, and its manifestations in the syntax. Suppletion is shown to be syntactic, not merely lexical, and to possibly occur late enough in the derivation to avoid an exceptional instantiation of suppletive plural concord in Hindi. Taking T's "phi-features" to be nominal head features may allow for a new understanding of nominative case as "direct" and not "oblique," and of feature-redundancy in concord and agreement (including "default agreement") as constituency indications, instead of an imperfection in language, i.e., the "doubling" of nominal features in syntactic feature-sharing provides clues to syntactic structuring and constituency (a point made by Norris 2011). Its overtness contrasts with the silence of a doubled set of nominative case features on DP heads such as N and Dem (but not Num) in my analysis of Hindi (perhaps by a condition similar to No Crowding, that silences PL2 when PL1 occurs). The factors that govern these choices of silence and spell out of redundant syntactic features, or of heads, which are perhaps central to the acquisition of spoken languages, remain to be addressed.

Acknowledgements

The ideas in this paper have had varied and chequered lives. Thanks to: Audiences at Glow-in-Asia (Taiwan, 2014), IIT Delhi (2013), Nanzan University (2010), and EFL-U (2010): particularly Hilda Koopman, Ayesha Kidwai, Pritha Chandra, Paroma Sanyal; also Probal Dasgupta, Giuliana Giusti, Rajesh Bhatt; K.P Mohanan, Tara Mohanan, Rama Kant Agnihotri; numerous anonymous reviewers, including those whose puzzlement and displeasure (at my "pretentiousness"), helped sharpen bits of the argument (such as it is); my students, Vineet Chaturvedi and Saurabh Kumar Mishra, for re-kindling my interest in Hindi; and as always, Rahul Balusu and K. A. Jayaseelan for their support, as they struggled with my proposals and their Hindi.

References

Amritavalli, R. 2019. Ergative in Hindi/Urdu: Reconciling the perfective oblique and the heavy imperfective analyses. *Journal of South Asian Linguistics,* 9, 1-23.

Amritavalli, R. 2021. Number and Case in Hindi. *Nanzan Linguistics* 17.

Amritavalli, R., and Sindhu Herur. 2022. Gradability and Comparison in Kannada. In R. Amritavalli and Bhuvana Narasimhan, eds., *The Oxford Handbook of Dravidian Languages* (online edn., Oxford Academic, 19 Dec. 2022). https://doi.org/10.1093/oxfordhb/9780197610411.013.9, accessed 24 Aug. 2023.

Anand, P. and A. Nevins 2006. The locus of ergative case assignment: evidence from scope. In A. Johns, D. Massam & J. Ndayiragije, eds., *Ergativity: Emerging Issues*, Springer: Dordrecht, 3-24.

Baker, Mark C. 2008. *The syntax of agreement and concord*. Vol. 115. Cambridge University Press.

Barbiers, S. 2007. Indefinite Numerals One and Many and the Cause of Ordinal Suppletion. *Lingua* 117, 859-880.
Barbiers, S. 2008. Microvariation in syntactic doubling: an introduction. In S. Barbiers, O. Koeneman, M. Lekakou, & M. van der Ham, eds., *Microvariation in syntactic doubling,*1- 35. *Syntax and Semantics* vol. 36, Bingley, U.K.: Emerald.
Baunaz, Lena, and Eric Lander, 2018. Nanosyntax: the basics. In Baunaz, L., De Clercq, K., Haegeman, L., and Lander, E., eds., *Exploring Nanosyntax*. New York: Oxford University Press, 3-56. DOI: https://doi.org/10.1093/oso/9780190876746.003.0001
Bhatt, R., and E. Anagnostopoulou. 1996. Object shift and specificity: Evidence from *ko*-phrases in Hindi. *Papers from the main session of CLS, 32,* 11-22.
Bittner, M. & K. Hale. 1996. The structural determination of case and agreement. *Linguistic Inquiry* 27:1 1-68.
Blake, B. 2001. *Case*. Cambridge University Press: Cambridge.
Borer, Hagit. 2005. *Structuring sense*: Volume 1: In name only. Oxford University Press, 2005.
Caha, Pavel. 2009. *The nanosyntax of Case*. PhD. dissertation, *CASTL*, Tromso.
Caha, Pavel. 2013. Explaining the Structure of Case Paradigms by the Mechanisms of Nanosyntax: The Classical Armenian Nominal Declension. *Natural Language and Linguistic Theory* 31: 1015- 1066.
Castillo, J. 2001. Thematic relations between nouns. Doctoral thesis, U. of Maryland.
Carstens, V. 2000. Concord in minimalist theory. *Linguistic Inquiry* 31: 319–355.
Collins, C., and R. S. Kayne. 2020. Towards a Theory of Morphology as Syntax. Ms., NYU.
Giusti, Giuliana. 1995. A unified representation of Case and article: Evidence from Germanic. In H. Haider, Susan Olsen and Sten Vikner, eds., *Studies in comparative Germanic syntax*, 77-93. Dordrecht: Springer Netherlands.
Giusti, Giuliana. 2011. On Concord and Projection. *Bucharest Working Papers in Linguistics* 13:103-124.
Harley, Heidi, and Elizabeth Ritter. Person and number in pronouns: A feature-geometric analysis. *Language* (2002): 482-526.
Harris, J. (1991) The exponence of gender in Spanish. *Linguistic Inquiry* 22: 27-62.
Kayne, R. S. 2016. The Silence of Heads. *Studies in Chinese Linguistics* 37: 1-37. DOI: 10.1515/scl-2016-0001
Kayne, R. S. 2020. On a More Demanding Approach to Suppletion. Ms., NYU.
Marantz, Alec .1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. University of Pennsylvania working papers in linguistics 4.2 (1997): 14.
McFadden, T. 2018. *ABA in stem-allomorphy and the emptiness of the nominative. *Glossa* 3(1): 8. 1–36, DOI: https://doi.org/10.5334/gjgl.373
Mohanan, T. 1994. *Argument Structure in Hindi*. Stanford, CA: CSLI Publications.
Norris, M. 2011. Towards an analysis of concord (in Icelandic). In Jaehoon Choi et al., eds., *Proceedings of the 29th West Coast Conference on Formal Linguistics*, eds. 205-213. Somerville, MA: Cascadilla Proceedings Project.

Noyer, R. 1992. Features, Positions and Affixes in Autonomous Morphological Structure. Doctoral Dissertation, MIT.
Pesetsky, D. 2013. *Russian case morphology and the syntactic categories*. Cambridge, MA: MIT Press.
Pesetsky, D. & E. Torrego. 2001. T-to-C movement: causes and consequences. In *Ken Hale: a Life in Language*, ed. M. Kenstowicz, 355-426. Cambridge, MA: MIT.
Pesetsky, D. and E. Torrego. 2007. The syntax of valuation and the interpretability of features. In S. Karimi, V. Samiian & W. Wilkins, eds., *Phrasal and clausal architecture: Syntactic derivation and interpretation*, pp. 262–294. Amsterdam: John Benjamins.
Preminger, O. 2009. Failure to Agree is Not a Failure: ϕ-Agreement with Post-Verbal Subjects in Hebrew. Ms., MIT / Harvard.
Ritter, E. 1991. Two functional categories in noun phrases: evidence from Modern Hebrew. *Syntax and Semantics* 25, 37-62. San Diego: Academic Press.
Ritter, E. 1993. Where's gender? *Linguistic Inquiry* 24: 795–803.
Sinha, Yash. 2018. Hindi nominal suffixes are bimorphemic: A Distributed Morphology Analysis. *Proceedings of the Linguistic Society of America* 3. 20: 1-10. DOI: https://doi.org/10.3765/plsa.v3i1.4301
Starke, Michal. 2009. Nanosyntax: A short primer to a new approach to language. *Nordlyd* 36 (1): pp. 1-6. [online] Available at: http://septentrio.uit.no/index.php/nordlyd/index.…
Ueda, Yasuki. 2009. Number in Japanese and Chinese. *Nanzan Linguistics* 5: 105-130.
Zeijlstra, H. 2012. There is only one way to agree. *The Linguistic Review 29*(3): 491-539. doi: https://doi.org/10.1515/tlr-2012-0017

Malayalam long-distance anaphor *taan*: a null theory[1]

ATHULYA ARAVIND, *Massachusetts Institute of Technology*

ABSTRACT

The Dravidian languages, Malayalam included, have a third-person pro-form *taan*, with a surprising locality profile. Like pronouns, *taan* cannot be bound locally; like reflexives, it seems to require a sentence-internal antecedent. Nearly three decades years ago, Jayaseelan (1997) argued that Malayalam *taan* is a Condition B-obeying pronoun, but this analysis has since fallen out of favor. A prominent alternative instead treats *taan* as a reflexive, bound by a silent pronoun instantiating a syntactically-represented perspectival center (Jayaseelan 1998 for Malayalam; Sundaresan 2012, Sundaresan 2018 for Tamil). In this paper, I will reexamine the syntactic and interpretive profile of *taan* and argue that the evidence favors the '*taan*-as-pronoun' analysis. Minor amendments to *taan*'s semantics — specifically, encoding its perspective-sensitivity as a presupposition — capture much of its distribution.

1 Introduction

1.1 Two approaches to *taan*

The 'long distance anaphor' *taan* in Malayalam[2] and related Dravidian languages has been well-studied (Mohanan 1982; Amritavalli 1984; Lidz 1995; Jayaseelan 1997, 1998; Sundaresan 2012, 2018; Jayaseelan 2017; Aravind to appear. It is a third-person, +human anaphor, which typically requires an utterance-internal antecedent that can be arbitrarily far, but cannot be local. Thus, (1-a), where the only utterance-internal antecedent is a clause-mate of *taan*, is ill-formed. In (1-b) and (1-c), *taan* can be co-construed with a non-local subject. When multiple long-distance antecedents are available, a sentence with *taan* is ambiguous (1-c).[3]

[1]Unless otherwise noted, all Malayalam data reflect my own judgment, confirmed with two other native speakers.

[2]Malayalam is spoken primarily in Kerala, a province of India that stretches along its southwest coast. It has SOV word order (with relative word-order freedom), a nominative-accusative case system, and is head-final (has postpositions, final complementizers). Exceptional for a Dravidian language, it lacks agreement.

[3]The language also has a set of polymorphemic reflexives that are licensed only in the presence of a clause-mate co-construed antecedent, as in (i). These are ordinary Condition-A obeying reflexives, and I will put them aside in this paper.

(i) a. *Raman*$_i$ ***tann-e-tanne***$_i$ *sneehikk-unnu*
Raman self-ACC-self love-IMPF
'Raman loves self.'

b. *Raman*$_i$ ***avan-e-tanne***$_i$ *sneehikk-unnu*
Raman 3MSg-ACC-self love-IMPF

(1) a. *$Raman_i$ ***tann-e***$_i$ *sneehikk-unnu*
Raman ANAPH-ACC love-IMPF
'Raman loves self.'

b. *$Raman_i$ vicaarichu [peNkuTTikaL$_j$* ***tann-e***$_{i/*j}$ *sneehikk-unnu ennu]*
Raman thought [girls ANAPH-ACC love-IMPF COMP]
✓'Raman thought the girls love him.'
✗'Raman thought the girls love themselves.'

c. *$Raman_i$ vicaarichu [Amma$_j$ [peNkuTTikaL* ***tann-e***$_{i/j}$ *sneehikk-unnu*
Raman thought [Mom [girls ANAPH-ACC love-IMPF
ennu] paranju ennu]
COMP] said COMP]
✓'Raman thought Mom said the girls love him.'
✓'Raman thought Mom said the girls love her.'

The problem posed by *taan* is the same as that posed by long-distance anaphora generally: it does not fit neatly into classical binding theory (Chomsky 1981 *et seq.*). It seemingly has a freer distribution than Condition A obeying reflexives, which require a co-indexed local antecedent. It also has a stricter distribution than Condition B obeying pronouns, which can pick up antecedents from the discourse.

Prior attempts to reconcile *taan*'s distribution with binding theory have fallen into two camps. One approach, pursued by Jayaseelan (1997), argues that *taan* is a species of Condition B-obeying pronoun.[4] Crucially, it is ϕ-featurally deficient, which forces the presence of an utterance-internal antecedent for interpretation. He observed that modulo discourse anaphora, *taan* is in free variation with third-person personal pronouns in many contexts. Both *taan* and ordinary pronouns can participate in cross-clausal anaphora, as in (1-b) and (1-c). Both *taan* and personal pronouns can appear inside a DP (2) or a PP (3) and be felicitously anteceded by a phrase-external DP.

(2) *Pronoun*

a. $\boxed{Raman_i}$ *[$_{DP}$* ***avan-te***$_i$ *kutti-e] snehikkunnu*
Raman 3Msg-GEN child-ACC loves
'Raman loves his child.'

b. $\boxed{Raman_i}$ *[$_{PP}$* ***avan-te***$_i$ *munn-il] oru aana-ye kaNDu*
Raman 3Msg-GEN front-LOC one elephant-ACC saw
'Raman saw an elephant in front of him.' [Jayaseelan 1999, ex.71]

(3) *Taan*

a. $\boxed{Raman_i}$ *[$_{DP}$* ***tan-te***$_i$ *kutti-e] snehikkunnu*
Raman ANAPH-GEN child-ACC loves
'Raman loves his own child.'

'Raman loves himself.'

[4]Jayaseelan was building here on prior observations in Mohanan (1982) and Amritavalli (1984)

b. $\boxed{Raman_i}$ $[_{PP}$ ***tan-te***$_i$ *munn-il]* *oru aana-ye* *kaNDu*
Raman ANAPH-GEN front-LOC one elephant-ACC saw
'Raman saw an elephant in front of himself.' [Jayaseelan 1999, ex.24]

Both, furthermore, can be anteceded by a non c-commanding nominal external to its clause.

(4) a. $\boxed{Raman\text{-}te_i}$ *aagragraham* ***[avan***$_i$ *manthri aakaN-am* *ennu]* *aaNu*
Raman-GEN wish [3MSg minister become-MOD COMP] COP
'Raman's wish is that he become minister.'
b. $\boxed{Raman\text{-}te_i}$ *aagragraham* ***[taan***$_i$ *manthri aakaN-am* *ennu]* *aaNu*
Raman-GEN wish [ANAPH minister become-MOD COMP] COP
'Raman's wish is that he become minister.' [Jayaseelan 1999, ex.41]

But Jayaseelan himself had a change of heart soon, thereafter. In Jayaseelan (1998), he noted a set of interpretive restrictions on *taan* that suggested that the anaphor is *perspective-sensitive*.[5] Its antecedent has to be identified with the 'point-of-view' holder or 'perspectival center' of the sentence. To account for this perspective-sensitivity, he proposed that *taan* is bound by a silent perspectival element at the clause-edge. On this analysis, *taan* is a Condition A-obeying reflexive, with a local binder that happens to be silent.

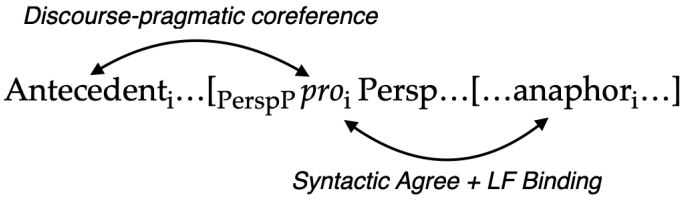
In the three decades since, the earlier, *taan*-as-pronoun analysis has fallen out of favor. In contrast, there has been much interest in the perspective-bound reflexive analysis. The general idea — that apparent long-distance anaphora with *taan* in fact involves local binding by silent material — has since been developed and extended to account for long-distance anaphora in many languages, including other Dravidian languages (see in particular, Sundaresan 2012, 2018; Nishigauchi 2014; Charnavel 2019). Here I will sketch in more detail Sundaresan's (2018) proposal, as it pertains to the anaphor *taan* in the closely related Dravidian language Tamil.[6]

1.2 Sundaresan's (2018) account of perspectival anaphora

Sundaresan (2018) defines perspectival anaphora as cases where an anaphor "is properly contained within a predication which is evaluated relative to the perspective, mental or spatial, of some sentient individual... and the antecedent of the anaphor must denote this individual" (p.6). Long-distance anaphora, she notes, seems to be regulated both by structural constraints (anti-locality) and pragmatic constraints (perspective-sensitivity). To capture both effects, she develops a "two-stage" model, schematized in (5).

[5]Jayaseelan's characterization of perspective is similar to the notion of "empathy" discussed by Kuno and Kaburaki (1977), Kuno (1987) and Oshima (2006) in relation to the Japanese long-distance anaphor *zibun*.

[6]It's worth noting, though, that everything I mention about Sundaresan's proposal should equally apply to other similar proposals on the market.



(5) (Sundaresan 2018, p.2)

Core to Sundaresan's proposal is the idea that perspectival anaphors are syntactically bound by a silent pro-form, the perspectival *pro*. Perspectival *pro* is introduced in the specifier of a Perspective Phrase (PerspP). The head Persp^0 assigns the pronoun a θ-role and semantically relates the individual denoted by the pronoun to an event argument as the perspective-holder of that event.

Perspectival *pro* itself receives its reference from the context and co-refers with the most salient antecedent. The tendency for long-distance anaphors to have utterance-internal antecedents is simply a consequence of the greater salience of these mentioned antecedents compared to those that are not mentioned. In this system of multiple dependencies, only one element — perspectival *pro* — is formally the perspectival holder, but the nature of binding and co-reference is such that its antecedent (the overt "antecedent" of *taan*) and the element that it antecedes (*taan*) will also track the perspective-holder.

Whenever a perspectival reflexive is licensed, that means that there is a sufficiently local PerspP projection and a perspectival *pro* that can bind the reflexive. Sundaresan ties perspective to phasehood: CPs, *v*Ps, DPs and PPs can all host their own PerspPs. This is consistent with data we already saw: *taan* is licensed not just in clauses, but also inside DPs and PPs.

The primary evidence for Sundaresan's proposal comes from Tamil. The Tamil long-distance anaphor *taan* has a highly similar distribution as its Malayalam counterpart. A critical difference between the two languages is that Tamil has subject-verb agreement, and Tamil *taan* can appear in configurations involving so-called "monstrous" agreement. *Taan* can optionally control first-person agreement morphology on the verb. Thus in (6), the embedded verb displays first-person agreement morphology, though neither *taan* nor the overt antecedent *Sai* is first-person.

(6) *Sai$_i$ [taan$_i$ djej-pp-**een**-nnu] so-nn-aan*
Sai ANAPH win-FUT-1SG-COMP say-PST-3SG
'Sai said that he would win.' (Sundaresan 2018, ex.57)

Sundaresan (2018) argues that the unexpected agreement pattern arises due to the interaction of the agreeing T, the anaphor *taan* in subject position, and the silent perspectival *pro*. Building on earlier work (e.g. Jayaseelan 1997), *taan* is taken to lack ϕ-features altogether. This means that when the subject is *taan*, T will fail to find ϕ-features on the subject DP, leading to an expansion of its search domain. T will then probe upwards and agree with the next closest DP — the perspectival *pro*. Monstrous agreement takes place in attitude and

speech contexts, contexts that are known to trigger shifted readings of indexicals in certain languages. Sundaresan argues that in monstrous agreement scenarios, the perspectival *pro* in the embedded CP is a shifted first person indexical. Thus, monstrous agreement involves genuine agreement with a silent first person element, whose reference, given indexical shift, tracks *not* the speaker coordinate of the utterance context, but the author of the attitudinal context.

1.3 Present work

The perspective-bound reflexive approach is appealing on both theoretical and empirical fronts. In unifying long-distance and local anaphora, it simplifies the typology of anaphoric dependencies. Empirically, the approach accounts for the interpretive quirks of long-distance anaphors and provides an explanation for otherwise recalcitrant data, such as monstrous agreement in Tamil. Despite these virtues, I will argue in this paper that this approach is insufficient for Malayalam *taan*. Specifically, it fails to fully capture its distribution. The *taan*-as-pronoun approach fares better in this regard. I will therefore attempt to rescue this approach by making certain modifications to deal with *taan*'s perspective-sensitivity. The crucial move will be to shift away from a "syntactification" of perspective, and instead treat perspective as a contextual parameter — a strategy pursued by various semantic treatments of perspectival phenomena (Lasersohn 2005; Stephenson 2007; Percus 2011; Sudo 2015). This revised *taan*-as-pronoun approach effectively captures many, though not all, of *taan*'s properties.

I begin in the next section by presenting data that showcase *taan*'s perspective-sensitivity and outlining how these data might be explained on a perspective-bound reflexive approach. In §3, I'll discuss patterns that are more problematic for the perspective-bound reflexive approach. The main problem is that the postulated perspectival domains are sometimes too small and other times too large. I turn to my own proposal in §4, which incorporates insights of both camps of prior analyses. In §5, I discuss some shortcomings of this proposal.

2 Evidence for perspective-sensitivity

Despite otherwise having a similar distribution as pronouns, *taan* shows certain restrictions on its antecedents that are not shared by ordinary third-person pronouns. *Taan*, unlike regular pronouns, tends to be subject-oriented. This contrast is demonstrated in (7): a non-subject DP in the matrix clause can antecede a regular pronoun (7-b), but not *taan* (7-a).

(7) a. *Raman* [*Sita-yode*$_j$] ***tan-te***$_{*j}$ *bhaavi-e patti paranju*
Raman Sita-soc ANAPH-GEN future-ACC about told
✗'Raman told Sita about her own future.'

b. *Raman* [*Sita-yode*$_j$] ***avaL-uDe***$_j$ *bhaavi-e patti paranju*
Raman Sita-soc 3FSg-GEN future-ACC about told
✓'Raman told Sita about her future.'

Sundaresan (2012) argues this is not a syntactic restriction but a conceptual one. The real requirement is that the antecedent of *taan* is sentient, as only sentient beings can be perspective-holders. Once the sentience requirement is met, subject-orientation can be explained on the perspective-bound reflexive approach if the more salient antecedent for the perspectival *pro* is the matrix subject. Indeed, subjects have been argued to be more prominent than non-subjects for pronoun resolution (Crawley et al. 1990; Grosz et al. 1995).

But as previously noted by Jayaseelan (1998), subject-orientation is only a descriptive tendency for *taan*. When the perspectival center is clearly something else, e.g. the "protagonist" of the narration in Free Indirect Discourse (FID) contexts, *taan* can appear *sans* an overt antecedent and get co-identified with that protagonist.

(8) $\boxed{John_i}$ *manassilaakki, ee bandham awasaaniccu ennu.* ***Taan***$_i$ *ini*
John understood, this relationship finished COMP. ANAPH now-on
Mary-e orikkalum kaaN-illa.
Mary-ACC ever see-NEG
'John understood that this relationship is finished. He [John] will never see Mary again.' (Jayaseelan 1998, ex. 14)

This, too, is unproblematic of the perspective-bound reflexive approach. Recall that the relation between *taan* and its surface antecedent is claimed to be *mediated*. In actuality, the apparently unbound *taan* in (8) is bound by the silent perspectival *pro* at the clause-edge. FID involves special contexts where a narrator uses a sentence (what appears to be a root clause) to communicate the thoughts and perceptions of a character who inhabits the world of the (potentially fictional) narrative. In such circumstances, the FID-protagonist is presumably highly salient, and a suitable antecedent for the perspectival *pro*.

Yet another piece of evidence pointing to *taan*'s perspective-sensitivity is the way it interacts with other perspectival elements. Certain verbs of transfer in Malayalam encode whether the goal *should* or *should not* be identified with the perspective holder.[7] When *taan* co-occurs with these verbs, perspectival consistency is enforced (original observations due to Jayaseelan 1998). As an illustration, consider the contrast in felicity between the (a) and (b) sentences in (9), which involve the perspectival verbs *tar-* and *koDukk-*, both of which roughly translate to 'give'.

(9) a. $\boxed{Raman_i}$ *[Sita sammaanam* ***tan-ikku***$_i$ *<u>tar-um</u> ennu] vicaariccu*
Raman [Sita prize ANAPH-DAT give-FUT COMP thought
'Raman thought that Sita would give the prize to him.'

b. # $\boxed{Raman_i}$ *[Sita sammaanam* ***tan-ikku***$_i$ *<u>koDukk-um</u> ennu] vicaariccu*
Raman [Sita prize ANAPH-DAT give-FUT COMP thought
'Raman thought that Sita would give the prize to him.'

The verb *tar-* requires that the goal of transfer is also the perspective-holder. The verb *koDukk-* requires that the goal is *not* the perspective-holder. On the perspective-bound

[7] These verbs are similar to the better-studied Japanese empathy verbs (Kuno 1987).

reflexive account, the oddness of (10-b) is explained if the perspectival requirements of the binder of *taan* conflict with those of *koDukk*: *taan* is bound by an element that represents the perspectival center, but the verb mandates against such an element being the goal.

A final interpretive restriction worth noting is that *taan* cannot co-occur with co-referential personal pronouns in the same clause (10-a). It can, however, co-occur with other co-construed instances of *taan* (10-b).

(10) a. * $\boxed{\textit{Raman}_i}$ ***[taan***$_i$ ***avan-te***$_i$ *viiTT-ileekku pook-um ennu] paranju*
Raman ANAPH 3Msg-GEN house-LOC go-FUT COMP said
'Raman said that he will go to his house.'

b. $\boxed{\textit{Raman}_i}$ ***[taan***$_i$ ***tan-te***$_i$ *viiTT-ileekku pook-um ennu] paranju*
Raman ANAPH ANAPH-GEN house-LOC go-FUT COMP said
'Raman said that he will go to his own house.'

On the perspective-bound reflexive account, this restriction can be explained by Condition B. The sentences in (10) both contain an occurrence of *taan*, co-construed with the matrix subject. To capture this co-construal, the account posits a structure like (11) for these sentences.

(11) [Raman$_i$ .... [*pro*$_i$ Persp^0 [... anaphor$_i$/pronoun$_i$]]]

Because the perspectival *pro* is co-indexed with Raman in this structure, only a reflexive — i.e. *taan* — can occur in its scope. A pronoun would be too local to a co-indexed antecedent, violating Condition B. See related discussion in Sundaresan (2018), though for different types of data.

3 Locality troubles

We have seen that many characteristics of Malayalam *taan* can be straightforwardly accounted for on the perspective-bound reflexive account. In this section, I turn to properties whose explanation is less clear, which have to do with *taan*'s locality profile. I will focus in particular on two. The first involves situations where the PerspPs we need to posit to license *taan* fail to correspond to the domains where perspectival constraints are enforced. The second has to do with the behavior of *taan* in infinitival complements. Here, *taan*'s distribution fully parallels that of ordinary pronouns, and does not follow from a treatment of the expression as a reflexive bound by silent material at the edge of the infinitival clause.

3.1 Conflicting perspectival domains

We saw in the previous section that when there are multiple perspective-sensitive elements, their perspectives need to resolve to the same center. The perspective-bound reflexive approach can make sense of this, so long as the elements are within a single PerspP. In addition to clauses, DPs and PPs are argued to constitute independent perspectival domains (Sundaresan 2012, 2018). Postulating PerspPs inside DPs and PPs is necessary to account

for the fact that *taan* is licensed inside them with phrase-external, but still clause-internal, antecedent. However, perspectival consistency requirements do not seem to correspond to every domain hypothesized to host PerspPs. Rather, they seem to be operative uniformly over *an entire clause.*

Consider the simplex sentence in (12). The sentence contains an occurrence of *taan* inside the possessive DP, understood to co-refer to the sentential subject *Sita*. The fact that *Sita* can antecede *taan* despite being part of the same clause is taken to show that the possessive DP hosts its own PerspP. This sentence also involves a perspective-sensitive verb of motion, *var-*, which requires that the perspective-holder be at the goal-location. Here, both perspectives — the referent of *taan* and the individual at the goal-location — naturally resolve to the subject. Note that this is so despite there being two distinct PerspPs.

(12) *Sita$_i$ [[**tan-te**$_i$ kuTTi-uDe] veeTT-ileekku] vannu*
Sita ANAPH-GEN child-GEN house-LOC walked
'Sita came to her child's house'

This alone is uninformative. But consider (13-a), where we have embedded the above sentence under the attitude verb *vicaarikk-* 'think', thereby introducing another potential antecedent for *taan*. Not all predicted readings for this sentence are available. If Raman is at the goal-location, then *taan* must also resolve to Raman. Sita is no longer an option, despite being a possible antecedent for *taan* in the unembedded variant.

(13) *Raman [Sita [[**tan-te** kuTTi-uDe] veeTT-ileekku] <u>vannu</u> ennu] vicaariccu*
Raman Sita ANAPH-GEN child-GEN house-LOC came COMP thought
✓*taan* = Raman; Raman in child's house
✗*taan* = Sita; Raman in child's house

If we replace the perspective-sensitive *var-* with a non-perspectival verb like *naDakk-* 'walk', as in (14), the "mixed" reading becomes available. Here, both Raman and Sita are possible antecedents for *taan*.

(14) *Raman [Sita [[**tan-te** kuTTi-uDe] veeTT-ileekku] <u>naDannu</u> ennu]*
Raman Sita ANAPH-GEN child-GEN house-LOC walked COMP
vicaariccu
thought
✓*taan* = Raman
✓*taan* = Sita

The contrast between (13) and (14) suggests that the loss of a reading is due to a requirement for perspectival consistency across *taan* and *var-*. But if a DP can in principle host its own PerspP, it is not clear why there needs to be consistency between a DP-internal *taan* and the perspectival-verb, which is outside the perspectival domain of the anaphor.[8]

[8]Sundaresan (2018) reports that mixed readings are in fact available in similar structures (involving PPs) with Tamil *taan*. I'm not sure what to make of this variation.

One possible response is to blame extra-grammatical factors for the lack of mixed readings. Perhaps salience considerations could result in the perspectival pronouns in both domains being resolved to the same referent. It is difficult to fully control for salience factors, but here is an attempt using conjoined subjects. (15) serves as the baseline. It shows that the individual conjuncts, *Raman* and *Ravi*, are both salient enough to serve as antecedents for ordinary pronouns within the sentence. Importantly, they can each antecede a different pronoun, yielding the mixed reading indicated by the indices in (15).[9]

(15) $\boxed{\textit{Raman}_i\textit{-um}}$ $\boxed{\textit{Ravi}_j\textit{-um}}$ ***avan-te***$_i$ *Amma-ye-um* ***avan-te***$_j$
Raman-CONJ Ravi-GEN-CONJ 3Msg-GEN mother-ACC-CONJ 3Msg-GEN
aniyatti-ye-um *New York-il* *koNDuvannu*
sister-ACC-CONJ New York-LOC brought
'Raman and Ravi brought Raman's mother and Ravi's sister to NY.'

In principle, these DPs should also be salient enough to serve as antecedents for silent pronominals, such as the perspectival *pro* that binds *taan*. However, when we replace the overt pronouns with *taan*, as in (16), the mixed reading disappears. Notice, though, that each DP remains sufficiently salient to serve as the antecedent for *taan*, so long as it antecedes *both* occurrences of the anaphor, not just one.

(16) # $\boxed{\textit{Raman}_i\textit{-um}}$ $\boxed{\textit{Ravi}_j\textit{-um}}$ ***tan-te***$_i$ *Amma-ye-um* ***tan-te***$_j$
Raman-CONJ Ravi-CONJ ANAPH-GEN mother-ACC-CONJ ANAPH-GEN
aniyatti-ye-um *New York-il* *koNDuvannu*
sister-ACC-CONJ New York-LOC brought
✗'Raman and Ravi brought Raman's mother and Ravi's sister to NY.'
(✓Raman's mother and Raman's sister)
(✓Ravi's mother and Ravi's sister)

The loss of the mixed reading here is difficult to explain on the perspective-bound reflexive account. Both possessive DPs should be able to contain their own PerspPs, from whose specifier a silent pronoun can locally bind each instance of *taan*. This pronoun, furthermore, should be able to pick up reference from the context, as any ordinary pronoun. The prediction then is that (16) should have all the same readings as (15), but this prediction is not borne out. Rather, the generalization appears to be that the relevant domain for perspective is the whole clause, and not the proposed PerspPs.

3.2 Infinitives

Inside infinitival complements, *taan* has the same distribution as ordinary pronouns. This was already observed by Jayaseelan (1997) and used as evidence for his *taan*-as-pronoun account. In this subsection, I will argue that the distributional restrictions on *taan* inside

[9]The sentence is most naturally read with contrastive stress on both genitive pronouns. Crucially, stress does not save the *taan* variant.

infinitives cannot be explained if it is in fact a reflexive bound by a perspectival *pro* at the edge of the infinitival clause.

Let us first consider ECM-infinitives, arguably the least problematic for the perspective-bound reflexive approach. A matrix clause subject cannot antecede *taan* or pronouns when the latter are ECM subjects; compare (17-a) with the finite complement in (17-b).

(17) a. *$Raman_i$ [**tan-ne$_i$/avan-e$_i$** *oru miDukkan aayi] karuthi*
Raman$_i$ ANAPH-ACC/3Msg-ACC a clever.person COP considered
✗'Raman considered self/him a smart person.'

b. *$Raman_i$ [**taan$_i$/avan$_i$** oru miDukkan aaNu ennu] karuthi*
Raman ANAPH/3Msg a clever.person COP COMP considered
✓'Raman thought that he is a smart person.' [Jayaseelan 1999]

The inability of pronouns to occur in this environment is standardly explained by Condition B. The pronoun is too close to a co-indexed, c-commanding DP. But what about *taan*, if it is indeed a reflexive? On the perspective-bound reflexive account, the inability of *taan* to occur as ECM-subjects can be explained if the embedded subject moves past the embedded PerspP to a position in the matrix, as schematized in (18).

(18) [pro_i Persp^0 [... [Raman$_i$ .... [$_{VP}$ taan$_i$ [... [$_{TP}$ ... pro_i Persp^0 [... <~~taan$_i$~~>]]]]]]]

In such a scenario, *taan* would be too high to be bound by the embedded perspectival *pro*. The higher perspectival *pro*, while in principle a suitable binder, cannot co-refer with Raman, as that would cause a Condition C violation. Thus, the account explains *taan*'s inability to serve as an ECM subject co-indexed with the matrix subject, although the reasoning differs from why pronouns cannot be ECM subjects under the same circumstances.

Once we move beyond ECM, however, further problems emerge. In non-ECM contexts, a matrix subject can antecede an embedded non-subject *taan* or pronoun when a non-coreferential embedded subject intervenes (19-a).[10] However, the absence of this intervening subject blocks co-reference between *taan* and the matrix subject; see (19-b).

(19) a. *$Raman_i$ [Sita **tan-ne$_i$/avan-e$_i$** pukazht-aan] aagrahiccu*
Raman Sita ANAPH-ACC/3Msg-ACC praise-INF wanted
✓'Raman wanted Sita to praise him'

b. *$Raman_i$ [PRO_i **tan-ne$_i$/avan-e$_i$** pukazht-aan] aagrahiccu*
Raman ANAPH-ACC/3Msg-ACC praise-INF wanted
✗'Raman wanted to praise himself'

This contrast is unexpected. Why should a domain that could in principle host a PerspP and license *taan* fail to do so in the absence of an overt subject?

A possible response might be that the two infinitives in (19) are not of the same size. Perhaps infinitives that do not license an overt subject are structurally smaller than those

[10]Note that unlike ECM subjects, which receive exceptional ACC case, an embedded subject of *want*-predicates receives NOM case.

that don't, and they in turn do not project their own PerspPs. This is conceivable, but the same argument can be made with obligatory control environments involving only PRO-subjects. A matrix subject cannot antecede an embedded non-subject *taan*, or for that matter a third-person pronoun, in *subject-control* environments. This is illustrated in (20-a) with the subject-control predicate *try*. However, *taan can* be anteceded by the matrix subject in an *object-control* environment, (20-b).

(20) a. *$Raman_i$ [PRO_i **tan-ne$_i$/avan-e$_i$** *nannaakkuv-aan] shramiccu*
Raman ANAPH-ACC/3Msg-ACC improve-INF tried
✗'Raman tried to improve himself.'

b. $Raman_i$ *Ravi-ode*$_j$ [PRO_j **tan-ne**$_{i/*j}$**/avan-e**$_{i/*j}$ *nannaakkuv-aan]*
Raman Ravi-SOC ANAPH-ACC/3Msg-ACC improve-INF
paranju
said
✓'Raman said to Ravi to improve him (=Raman).'

Postulating clause-size differences for the two sentences in (20) is more methodologically suspect. The fact that *taan* is licensed in (20-b) suggests that there is a PerspP at the edge of the infinitival clause. Why it becomes unavailable in (20-a) is a mystery on the perspective-bound reflexive account.[11]

All of these distributional facts are straightforwardly accounted for on a pronoun account. *Taan* and ordinary pronouns like *avan* are both subject to Condition B — hence their parallel distribution inside infinitives. The contrast between overt and covert subjects in (19) tells us that the binding domain in Malayalam is the smallest clause containing the pronoun and a c-commanding subject. In the absence of that subject, neither *taan* nor *avan* is free in their binding domain. Both *taan* and pronouns are blocked in (20-a) for the same reason: because PRO, co-indexed with Raman, causes a Condition B problem for both types of expressions. In contrast, with object control predicates ((20-b)), *taan*/*avan* can be co-construed with the subject, with which PRO is contra-indexed. In sum, *taan* appears subject to identical locality constraints as personal pronouns — Condition B. It should therefore be treated as a pronoun.

4 Proposal: *taan* as a perspectival pronoun

A perspective-bound reflexive account of *taan* captures the perspective-sensitivity of the anaphor and related interpretive properties, but makes the wrong distributional predictions. A simplistic pronoun account, while accounting for the locality profile of *taan*, does not say anything about its interpretive restrictions. My diagnosis is that the issue lies not in treating *taan* as a perspectival element, but with the "syntactification" of perspective. In what follows, I will try to amend the pronoun approach to capture *taan*'s perspective-sensitivity.

[11] Questions also arise regarding the interaction of PRO — another arguably perspective-sensitive element — and PerspP; I am going to ignore these here.

4.1 A semantics/pragmatics for perspective

Perspective-sensitive phenomena are abundant in natural language. Other examples beyond anaphors include relative locative terms (e.g. *to the left, come, go,* the Malayalam *tar-/koDukk-* verbs we saw earlier), predicates of personal taste (e.g. *tasty, fun*), and relative socio-cultural terms (e.g. *foreigner*). Some of these have been given a "relativist-semantic" treatment in prior work (see e.g., Lasersohn 2005; Stephenson 2007; Percus 2011; Bylinina 2014; Sudo 2015 a.o.), which I will be building on here.

It is standard to assume that the interpretation function is relativized to three parameters: a variable assignment g, an index of evaluation i, and a context c. Indices and contexts are tuples, <x,y,t,w>, where x and y are individuals, t is a time and w is a world. For the context parameter, the values of these elements are set to the utterance coordinates: <speaker/author, addressee, utterance time, world of utterance>.[12] Indexicals such as *I*, *you* and *now* make reference to the c-parameter.

The crucial innovation necessary for capturing perspectival phenomena is the addition of a second context parameter, the "d" parameter, relevant for perspective terms (Percus 2011). This can be thought of as an enriched variant of the judge parameter in Lasersohn (2005).[13] We can think of the d-parameter as also being formally identical to c, though only the author-coordinate of d — auth(d) — will be relevant for us here. The fuller representation of the interpretation function is as in (21).

(21) $[\![.]\!]^{c,d,g,i}$

Perspective-sensitive expressions make reference to the d-parameter. For example, relative location verbs like *come* can be thought of as having a perspectival presupposition that the perspective-holder is at the goal-location (Oshima 2006; Sudo 2015):

(22) $[\![$Sita is coming to Kochi$]\!]^{c,d,g,i}$
is defined iff auth(d) is in Kochi
when defined, $[\![(19)]\!]^{c,d,g,i}$ = 1 iff Sita is traveling to Kochi in w_i.

Generally speaking, a speaker s uttering a sentence S in w while adopting the perspective of an individual y is saying that S evaluated at their own context c and at y's context d (the perspectival context) holds in w. In cases where the speaker is not obviously adopting anyone else's perspective, we can assume that they are taking their own perspective and identify auth(d) with auth(c). This feels correct for Malayalam, as it is for a language like English. Both the Malayalam sentence in (23) and the English one in (22) seem to convey that the speaker is currently in Kochi.

[12] The formally identical treatment of indices and contexts is argued for in, e.g., Anand 2003, von Stechow and Zimmermann 2005 and Deal 2020.

[13] The two context approach is also commonly adopted in discussions of FID, e.g. Doron 1991; Schlenker 2004; Sharvit 2008; Eckardt 2014; Abrusan 2020.

(23) *Sita Kochi-yileekku var-unnu*
Sita Kochi-LOC come-IMPF
'Sita is coming to Kochi.'

The motivation for postulating a separate perspectival parameter in addition to the usual context parameter is that perspectival and indexical terms can diverge in behavior. In languages like English and Malayalam, indexicals do not shift under attitudes. This means that in (24), the first person indexicals get their reference from the utterance context c and resolve to the speaker, even though the indexical is embedded under an attitude verb. Perspectival terms, on the other hand, do shift. The perspectival location relevant for the embedded perspective-sensitive verb *come* is most naturally that of the attitude holder Raman. Perspective shifting under attitudes will be discussed in more detail later in the paper.

(24) a. Raman thinks that I am coming to Kochi.
I = speaker of utterance; Raman at goal-location
b. *Raman [njaan Kochi-yileekku var-unnu ennu] vicaarikk-unnu*
Raman I Kochi-LOC come-IMPF COMP think-IMPF
'Raman thinks that I am coming to Kochi'
njaan 'I' = speaker of utterance; Raman at goal-location

4.2 A pronoun analysis for *taan*

I assume that all pronouns, including first and second pronouns, are uniformly interpreted as variables, following Heim & Kratzer (1998); Sauerland (2003); Heim (2008), among others. Pronouns come furnished with indices in the syntax. Indices are variables and mapped to semantic values by the contextually given assignment function g. This approach contrasts with treatments of first and second person pronouns are constants with an indexical semantics. Pronominal ϕ-features, including person features, constrain the range of possible referents by triggering presuppositions. For instance, first person pronouns presuppose that their referent is the speaker of the current context, (25).

(25) $[\![\mathrm{I}_7]\!]^{c,g,i}$ is defined iff g(7) is auth(c).
when defined, g(7) (Heim 2008)

We now have all the ingredients to propose a lexical entry for Malayalam *taan*. The proposal is simple: *taan* is a pronoun with a perspectival presupposition; see (26).[14]

(26) $[\![\mathrm{taan}_7]\!]^{c,d,g,i}$ is defined iff g(7) is auth(d).
when defined, g(7)

[14] A similar analysis has been proposed for the silent experiencer argument of taste-predicates by Stephenson (2007). Anand (2003) also treats *taan* as having a contextual presupposition, but one tied to the utterance context c:
(i) $[\![\mathrm{taan}_7]\!]^{c,d,g,i}$ is defined iff g(7) is auth(c); when defined, g(7).

Like ordinary third-person pronouns, *taan*'s interpretation is assignment-dependent. Its referent is whatever the contextually-given assignment function assigns to its index. Unlike ordinary third-person pronouns, *taan* carries a perspectival presupposition that its referent is identified with the author of d.

4.3 Welcome consequences

Various properties of *taan* fall out straightforwardly from this analysis. To start, its locality profile comes for free: *taan* is a pronoun and we expect its distribution to be governed by Condition B, modulo perspectival requirements.

One such requirement noted earlier is perspectival consistency: if multiple perspectival elements are part of the same sentence, their associated perspective sites typically have to be the same. We saw in §3, that this is a requirement enforced at the clause-level, independently of the locality conditions governing *taan*. On the present view, this is because perspective is a contextual parameter, and contextual parameters become relevant when evaluating the truth of a sentence.

This makes perspective-dependency, like other types of context-dependency, a phenomenon at the semantics-pragmatics interface. The semantics tells you where in the computation of the expressed proposition there are gaps for the pragmatics to fill in (e.g. in the form of presuppositions on pronouns). The pragmatics establishes how the contexts involved in sentence use (context of utterance, perspectival context) are identified and how such contexts determine a domain of discourse.

This approach also allows for a rethinking of the *quasi*-subject-orientation of *taan* as a kind of perspective "shift" under attitude verbs. The literature on indexical shift has taught us that contexts can systematically shift under attitude or speech situations. Attitude and speech predicate as quantifiers over indices: the lexical entry for a verb like *think*, for instance, says that every index that constitutes the attitude holder's doxastic alternatives are those where ϕ is true (27).

(27) $[\![\text{think } \phi]\!]^{c,d,g,i} = \lambda \text{x}.\ \forall \text{i'} \in \textbf{DOX}(\text{x})(\text{w}_i).\ [\![\phi]\!]^{c,d,g,i'} = 1$
where i' $\in$ **DOX**(x)(w_i) iff i' is compatible what x believes in w_i

Indexical shift happens when the c-parameter of the attitude complement is overwritten with the (attitude-bound) index parameter. In the framework for indexical shift originally proposed by Anand and Nevins (2004) and subsequently developed by many others (e.g. Anand & Nevins 2006; Sudo 2012; Deal 2020, a.o.), this over-writing is done by an intermediate element, a context shifter operator.

Recall that in under attitude verbs, *taan* is most naturally understood as co-referring with the attitude-holder. We see this in (28) (repeated from (1b)).

(28) $\boxed{\textit{Raman}_i}$ *vicaarichu [peNkuTTikaL* ***tann-e***$_i$ *sneehikk-unnu ennu]*
Raman thought [girls ANAPH-ACC love-IMPF] COMP
'Raman thought the girls love him.'

We can explain this tendency if under attitude contexts, the *d*-parameter gets overwritten in a manner similar to indexical shift. Attitude predicates may (optionally) combine with a perspective shifter as in (29-a), which overwrites the coordinates of the *d*-parameter with those of the index (see e.g. Percus 2011, Sudo 2015). In the context of this perspective shifter, auth(*d*) of an attitudinal complement will be identified with the attitude holder, the author of the attitudinal context (29-b).

(29) a. $[\![\text{OP}_d\ \phi]\!]^{c,d,g,i} = [\![\phi]\!]^{c,\underline{\mathbf{i}},g,i}$

b. $[\![\text{Raman thinks [Op}_d \text{ the girls like } \textit{taan}]]\!]^{c,d,g,i}$
$\forall \text{i'} \in \textbf{DOX}(\text{Raman})(\text{w}_i)$. $[\![\text{Op}_d \text{ the girls like } \textit{taan}]]\!]^{c,\underline{\mathbf{i}'},g,i'}$
all contexts that constitute Raman's doxastic alternatives are ones in which the girls like Raman's counterpart in those contexts[15]

5 Problems and (partial) solutions

The perspectival pronoun analysis predicts the distribution of *taan* to be relatively free as long as its perspectival presupposition is satisfied. In this section, I will highlight two puzzles for this prediction in Malayalam, both having to do with when *taan* can be co-construed with a speech act participant. The first problem is the *absence* of speaker-oriented readings of *taan*. I will propose a solution that will also rule out addressee-oriented readings. This makes way for a different problem. Addressee-oriented readings of *taan* are in fact available (in informal registers).

5.1 Absence of speaker-oriented readings

One of the pragmatic assumptions in the previous section was that at the matrix level, the *d*-parameter is identified with the *c*-parameter. This captures the fact that we naturally take the individual whose perspective the speaker adopts to be the speaker themself, if there are no clear indications otherwise. For *taan*, then, we predict that (30)— with an unembedded occurrence of *taan* — should have a reading where *taan* is co-construed with the speaker. But the sentence does not have this reading.

(30) ***Taan*** *oru linguist aanu*
ANAPH a linguist COP
✗'I am a linguist'

The absence of the intended reading in (30) may be related to another characteristic of *taan*: even in anaphoric contexts, *taan* cannot be anteceded by a participant. Co-construal of an embedded clause *taan* with a matrix subject that is first person (31-a) or second person (31-b) results in ill-formedness.

[15] I am collapsing the assertion/presupposition distinction here for convenience.

(31) a. *$njaan_i$ *[Sita* ***tan-ne***$_i$ *pukazhthi ennu] paranju*
I [Sita ANAPH-ACC praised COMP said
'I said that Sita praised me.'
b. *Nii_i *[Sita* ***tan-ne***$_i$ *pukazhthi ennu] paranju*
You [Sita ANAPH-ACC praised COMP said
'you said that Sita praised you.'

It is desirable to have a unified solution for both (31) and (30). Mine is the same as the one proposed by Sauerland (2003) and Heim (2008) for why (32) cannot be used to say that the speaker is a linguist.

(32) She is a linguist.

These authors propose that pronouns compete at the level of presupposition, and all else equal, the presuppositionally stronger one is forced. This is due to a principle "Maximize Presupposition" (Heim 1991), formalized in (33), which mandates that speakers should opt for forms with the strongest satisfied presupposition.

(33) **Maximize Presupposition (MP):** An utterance of a sentence S is infelicitous in a context c iff there is an alternative S' to S such that:
a. S and S' are contextually equivalent
b. The presuppositions of S and S' are both satisfied in c
c. The presuppositions of S' is stronger than the presuppositions of S

The proposed semantics for third and first person pronouns, given in (34), are such that the first person pronoun's presuppositions asymmetrically entail (trivially) that of the third person pronoun. The choice of the latter over the first gives rise to an "anti-presupposition" that g(7) does not include auth(c).

(34) a. $[\![\text{pro-3rd}_7]\!]^{c,d,g,i}$ = g(7)
b. $[\![\text{pro-1st}_7]\!]^{c,d,g,i}$ is defined iff g(7) = auth(c). when defined, g(7)

Could we simply extend this analysis to *taan*, which is also a third person pronoun? A hiccup in doing so is that *taan* also has a contextual presupposition. This means that sentences with *taan* and one containing a first person pronoun like (34-b) would fail to meet clause (c) of (33). The first person pronoun would not have a stronger presupposition than *taan*, just a different one. To solve this, I suggest that we modify the semantics of the Malayalam first person pronoun *njaan* to also encode a perspectival presupposition, as in (35). This would make it presuppositionally stronger and a suitable competitor for *taan*.

(35) $[\![\text{njaan}_7]\!]^{c,d,g,i}$ is defined iff g(7) = auth(c) = auth(d); when defined, g(7).

The move does not seem obviously wrong for Malayalam, as suggested by the oddness of the sentence in (36). In (36), we have a perspective-sensitive verb-of-transfer *koDukk-* 'give', which requires the goal not to be perspective-holder. The combination of this verb

and a first person goal is ill-formed, a pattern similar to what we saw in §2 with *taan* (compare (9b)). We can explain the ill-formedness if *njaan* and *koDukk-* have conflicting perspectival requirements.

(36) #*Sita sammaanam* ***eni-ikku*** *koDukk-um*
Sita prize 1Sg-DAT give-FUT
'Sita will give the prize to me.'

This proposal also provides a new explanation for another puzzling feature of *taan*, namely that it cannot co-occur in the same clause as participant pronouns irrespective of *taan*'s referent. See (37) and (38). This effect is often called the *participant blocking effect*, which *taan* shares with many long-distance anaphors crosslinguistically (see e.g. Tang 1989; Pan 1998; Jayaseelan 1998).

(37) Raman cannot antecede *taan* if *njaan* intervenes:

a. *$\boxed{\textit{Raman}_i}$ *[njaan* ***tan-ne***$_i$ *pukazhthi ennu] paranju*
Raman [I ANAPH-ACC praised COMP said
✗'Raman said that I praised Raman.'

b. $\boxed{\textit{Raman}_i}$ *[Sita* ***tan-ne***$_i$ *pukazhthi ennu] paranju*
Raman [Sita ANAPH-ACC praised COMP said
✓'Raman said that Sita praised Raman.'

(38) Raman cannot antecede *taan* if (non-c-commanding) *njaan* is in the same clause

a. *$\boxed{\textit{Raman}_i}$ *[****taan***$_i$ *enn-e orikkalum kaND-iTT-illa ennu] paranju*
Raman [ANAPH 1Sg-ACC ever saw-PERF-NEG COMP] said
✗'Raman said that he hasn't ever seen me.'

b. $\boxed{\textit{Raman}_i}$ *[****taan***$_i$ *Mohan-e orikkalum kaND-iTT-illa ennu] paranju*
Raman [ANAPH Mohan-ACC ever saw-PERF-NEG COMP] said
✓'Raman said that he hasn't ever seen Mohan.'

On the perspectival account of first person pronouns, we can explain these facts as the result of contradictory perspectival presuppositions. *Taan* presupposes that its referent is auth(d) and anti-presupposes that it is not the speaker; *njaan* presupposes that its referent is both the speaker and auth(*d*). These are requirements that can never be simultaneously met.

5.2 Presence of addressee oriented readings

Like the first person pronoun *njaan*, the second person pronoun *nii* also cannot antecede *taan*. Furthermore, it too, creates participant blocking effects. This, at first blush, suggests that we should extend our perspectival treatment of *njaan* to *nii*. On the other hand, treating both first and second person indexicals as encoding perspectival presuppositions would mean that one would never be able to say sentences like (39), though the sentence is acceptable.

(39) *njaan ninn-e kaNDu*
1sg 2sg-ACC saw
'I saw you.'

More puzzlingly, *taan* can in certain circumstances be understood as picking out the addressee. Typically, this happens when the address is informal (to a close friend, for instance). In the examples below, there is no second person antecedent. In fact, we already saw in (31b) that second person antecedents are banned. The addressee oriented *taan* is the highest argument in (40-b).

(40) a. *Raman$_i$ [Sita* ***tan-ne***$_i$ *pukazhthi ennu] paranju*
Raman [Sita ANAPH-ACC praised COMP] said
✓'Raman said that Sita praised you.'
b. ***Taan*** *miDukki aaNu*
ANAPH clever.person.F COP
'You are a clever (female) person.'

These addressee uses have been previously noted, but considered a case of accidental homophony in Jayaseelan (1999) and Asher & Kumari (1997) (though cf. Swenson & Marty (2014) who argue against this view). However, perspectival consistency is enforced with these uses, which points towards a unified analysis that makes reference to perspective. We see in (41) that when there are multiple occurrences of *taan* in a clause, the type of use (second-person, third-person) has to be consistent.

(41) *Sita [****taan*** *[****tan-te*** *kuTTi-ye] pukazhti ennu] paranju*
Sita ANAPH ANAPH-GEN child-ACC praised COMP said
✓you praised your child
✓Sita praised her child
✗you praised Sita's child

Perhaps these addressee-oriented uses are exactly what you would predict if auth(*d*) in these contexts happens to be the addressee and *taan* is free to pick it out (i.e. reference to addressee is not blocked via competition). On the other hand, the ban on second person antecedents and blocking by second person remain open problems, to which I have no concrete solutions to offer.

6 Concluding remarks

In this paper, I considered two analytic options for the Malayalam long-distance anaphor *taan*. The first involved taking the surface distribution of *taan* seriously, and treating it as a Condition B obeying pronoun. The setback of this approach was that it had little to say about *taan*'s interpretive restrictions. On the second approach, the surface distribution of *taan* was viewed as somewhat misleading. The expression's interpretive properties were taken to signal the presence of a silent local binder, making it a Condition A obeying reflex-

ive. I argued in favor of the pronoun approach. To capture *taan*'s interpretive differences from ordinary pronouns, I suggested that part of its meaning is a perspectival presupposition. This means — I believe correctly — that the distribution of *taan* is sensitive to two distinct domains: (*i*) its binding domain, which is same as that of pronouns, and (*ii*) the perspectival domain, which is uniformly the clause. Problems remain, and there are many I have not even touched on. But the hope is that pinning down the right character of the anaphor — *as a pronoun* — still helps to push the needle forward (or set it back to where Jayaseelan (1997) left it).

Acknowledgements

I am grateful to the audience members at FASAL 14, especially Sandhya Sundaresan, for feedback and discussion.

References

Amritavalli, Raghavachari. 1984. Anaphorization in dravidian. In *Central institute of english and foreign languages working papers in linguistics*, vol. 1, .

Anand, Pranav & Andrew Nevins. 2006. The locus of Ergative case assignment: Evidence from scope. In Alana Johns, Diane Massam & Juvénal Ndayiragije (eds.), *Ergativity: Emerging issues*, 3–25. Springer.

Aravind, Athulya. to appear. Locality in Malayalam anaphor binding. *Proceedings of WCCFL 39* 1–8.

Asher, Ronald & TC Kumari. 1997. *Malayalam*. Psychology Press.

Bylinina, Lisa. 2014. *The grammar of standards: judge-dependence, purpose-relativity and comparison classes in degree constructions*. Utrecht: UiL OTS dissertation.

Charnavel, Isabelle. 2019. *Locality and logophoricity: a theory of exempt anaphora* Oxford Studies in Comparative Syntax. Oxford University Press.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht, The Netherlands: Foris Publications.

Crawley, Rosalind, Rosemary Stevenson & David Kleinman. 1990. The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research* 19. 245–264.

Deal, Amy Rose. 2020. *A theory of indexical shift: Meaning, grammar, and crosslinguistic variation*, vol. 82. MIT Press.

Grosz, Barbara, Aravind Joshi & Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225.

Heim, Irene. 1991. Artikel und definitheit. In *Semantik: Ein internationales Handbuch der zeitgenossischen Forschung*, 487–535. Berlin: Mouton de Gruyter.

Heim, Irene. 2008. Features on bound pronouns. In Daniel Harbour, David Adger & Susana Béjar (eds.), *Phi theory*, 35–56. Oxford: Oxford University Press.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar*. Malden, MA: Blackwell.

Jayaseelan, K. A. 1997. Anaphors as pronouns. *Studia Linguistica* 51(2). 186–234.

Jayaseelan, K. A. 1998. Blocking effects and the syntax of Malayalam taan. In *The yearbook of south asian languages and linguistics*, 11–27. De Gruyter.

Jayaseelan, K. A. 2017. Anaphora in dravidian. In *Oxford research encyclopedia of linguistics*, .

Jayaseelan, K.A. 1999. *Parametric studies in malayalam syntax*. New Delhi, IN: Allied Publishers.

Lasersohn, Peter. 2005. Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy* 28(6). 643–686.

Lidz, Jeffrey. 1995. Morphological reflexive marking: Evidence from Kannada. *Linguistic Inquiry* 26(4). 705–710.

Mohanan, K.P. 1982. Infinitival subjects, government, and the abstract case. *Linguistic Inquiry* 13(2). 323–327.

Nishigauchi, Taisuke. 2014. Reflexive binding: awareness and empathy from a syntactic point of view. *Journal of East Asian Linguistics* 23. 157–206.

Oshima, David Y. 2006. *Perspectives in reported discourse*: Stanford University dissertation.

Pan, Haihua. 1998. Closeness, prominence, and binding theory. *Natural Language and Linguistic Theory* 16(4). 771–815.

Percus, Orin. 2011. Indexicality and compositional semantics. Lecture notes, ESSLLI 2011.

Sauerland, Uli. 2003. A new semantics for number. In *Semantics and linguistic theory (SALT) 13*, 258–275. Cornell University, Ithaca, NY: CLC Publications.

Stephenson, Tamina. 2007. *Towards a theory of subjective meaning*. Cambridge, Massachusetts: MIT dissertation.

Sudo, Yasutada. 2012. *On the semantics of phi features on pronouns*. Cambridge, Massachusetts: MIT dissertation.

Sudo, Yasutada. 2015. Come and go. Talk at the Perspectives in Context Workshop, Georg-August Universitat Gottingen.

Sundaresan, Sandhya. 2012. *Context and (co)reference in the syntax and its interfaces*: University of Tromsø and University of Stuttgart, Tromsø dissertation.

Sundaresan, Sandhya. 2018. Perspective is syntactic: evidence from anaphora. *Glossa: A journal of general linguistics* 128(1). 1–40.

Tang, C.-C. J. 1989. Chinese reflexives. *Natural Language and Linguistic Theory* 7. 93–121.

Scrambling in Bengali: An A-/A'-Movement Distinction

SREYOSHI BASU, *University of Rochester*

ABSTRACT

Bengali is an SOV language (Bhatt & Dayal 2007), known for its flexible word-order. Elements in a phrase can be moved to other positions, both within and across clausal boundaries, in a process called scrambling (David 2015). This study aims to provide a comprehensive description of scrambling in Bengali and argues that scrambling manifests in two types of movement in this language: A- and A'-. It further argues that the type of scrambling involved (A- vs. A'-) is predictable from the syntactic environment based on the following generalization: A'-movement is possible only when a Spec,CP position is available as a landing site. Given this, scrambling in Bengali supports the position-based approach to the A-/A'- distinction, recently argued for in Keine (2018). Building on previous literature on scrambling in other SOV languages, such as Hindi (Keine 2018; Dayal 1994; Mahajan 1990, 1994) and Japanese (Sato & Goto 2014; Saito 1985, 1992), this paper investigates scrambling in four syntactic environments, each with a different scrambling profile: 1) vP-internal movement; 2) clause-internal movement; 3) cross-non-finite clause movement; and 4) cross-finite clause movement. Two well-established tests are used to discern A-movement from A'-movement: i) A-movement can obviate weak crossover effects and lead to reciprocal binding; ii) A'-movement can reconstruct for Condition A. It is demonstrated that vP-internal scrambling is unambiguously A-movement, while clause-internal scrambling may be both A- and A'-movement. Additionally, cross-clausal movement out of non-finite clauses can be both A- and A'-movement, but cross-clausal movement out of finite-clauses is unambiguously A'-movement.

1 Introduction

1.1 Linguistic Description

Bengali (endonym: Bangla; ISO: *ben*) is the national language of Bangladesh and the official language of the Indian states of West Bengal and Tripura (David 2015; Lewis 2009). It belongs to the Indo-Aryan sub-group of the Indo-European language family (David 2015). Spoken Bangla exhibits considerable dialect variation, with two of the most widely documented varieties being Kolkata Colloquial Bengali (KCB) and Dhaka Colloquial Bengali (DCB), which represent the standardized dialects of Kolkata and Dhaka, respectively. (David 2015). This project focuses on an analysis of KCB. [1]

[1] All data that is not cited is provided by the author, a native speaker of Bengali.

Bangla is an SOV language with post-positions and a head-final clause structure (Thompson 2020; Bhatt & Dayal 2007). The basic word order of a declarative sentence follows the pattern: subject, indirect object, direct object, and verb (S IO DO V), as shown in (1). Auxiliaries and modals typically follow the main verb (David 2015).

(1) Apu Keya-ke ek-ta chobi dekha-lo - [S IO DO V]
Apu.NOM Keya-ACC one-CLF picture show-PST
'Apu showed Keya a picture.'

1.2 Scrambling in Bangla

Bangla has a fairly flexible word order, allowing elements of a phrase to be moved to other positions in a process known as scrambling. In free-word-order languages, scrambling can be defined as the process that allows for the derivation of non-canonical word-orders via movement of constituents from their base-generated positions to other syntactic positions (Cho 1994; Saito 1985). Scrambling operations in Bangla are generally optional, and the version of the sentence without movement, that is, the basic word order, is always available (David 2015; Keine 2018). However, despite syntactic optionality, such movement of constituents often "alter[s] the information structure in some salient way" (David 2015). For instance, scrambling is often used to achieve variable emphasis and 'contrastive focus interpretations' (Thompson 2004; Syed 2017). Focus tends to fall on the word occupying the first position in the clause, while the second position serves to emphasize the meaning of the first word. A transitive sentence like *I have read the story* can be scrambled in six different ways, as shown in (2).

(2) a. ami golpo-ta pod-e-chi - [SOV]
1SG.NOM story-CLF read-PRF-PRS
'I have read the story.'

b. ami pod-e-chi golpo-ta - [SVO]
1SG.NOM read-PRF-PRS story-CLF
'I have *read* the story.'

c. golpo-ta ami pod-e-chi - [OSV]
story-CLF 1SG.NOM read-PRF-PRS
'The story, *I* have read.'

d. golpo-ta pod-e-chi ami - [OVS]
story-CLF read-PRF-PRS 1SG.NOM
'The story, I have *read*.'

e. pod-e-chi ami golpo-ta - [VSO]
read-PRF-PRS 1SG.NOM story-CLF
'I have *read it*, the story.'

f. pod-e-chi golpo-ta ami - [VOS]
read-PRF-PRS story-CLF 1SG.NOM
'I have read it, *the story*.'

Additionally, scrambling in Bangla allows constituents to undergo both leftward and rightward-movement. The subject or object may be moved to clause-initial or clause-final positions to highlight different "discourse relevant information," such as distinguishing between new or old information or emphasizing background versus foreground information (David 2015). Clause-initial (3-a) or clause-final (3-b) positions are generally indicative of emphasis (Thompson 2004), as demonstrated in the examples below:

(3) a. **gari-ta** ami **t**$_1$ chali-e-chi gotokal
car-CLF 1SG.NOM **t**$_1$ drive-PRF-PRS yesterday
'The car I drove yesterday.'

b. am-ar **t**$_1$ ach-e **ek-ti** **darun dharona**
1SG-GEN **t**$_1$ be-PRS one-CLF.DIM great idea
'I have a great idea.' (from David 2015:248)

Existing studies on Bangla syntax have explored topics such as headedness and clause structure. For instance, according to Simpson and Bhattacharya (2003), Bangla has an underlying SVO structure. They argue that wh-questions and surface-SOV structures are derived through overt movement as opposed to an underlying SOV structure that combines wh-in-situ constructions and covert movement. Bhatt and Dayal (2007) argue against this claim, drawing upon rightward remnant movement to make their argument. Islam (2016) also offers a critical evaluation of the aforementioned claim, highlighting the need for covert movement and arguing that the analysis for Bangla remains wh-in situ. Descriptions of Bangla's free word order can be found in the literature (David 2015; Bhatt & Dayal 2007; Thompson 2004); however, the type of movement (A- or A'-) involved in different scrambling environments, both within and across clausal boundaries, the syntactic positions targeted by these movements, and the reasons for differing properties across various scrambling environments have yet to be adequately described for Bangla.

Therefore, this study aims to provide a comprehensive description of scrambling in Bengali by examining the type of movement and the syntactic positions targeted by that movement. To that end, this research builds on existing literature on scrambling in other SOV languages, such as Hindi (Keine 2018; Mahajan 1990, 1994; Dayal 1994) and Japanese (Saito 1992, 1985; Sato & Goto 2014). The analysis focuses on scrambling in four different syntactic environments: 1) vP-internal movement, 2) clause-internal movement, 3) cross-non-finite clause movement, and finally, 4) cross-finite clause movement.

Movement in Bangla manifests as either A- or A'- movement. A-movement can feed binding relations, while A'-movement cannot. Therefore, in Section 2, two well-established tests that discern A-movement from A'-movement are used to identify the types of movement involved in each scrambling environment:

i) Only A-movement can obviate weak crossover effects and lead to reciprocal binding, and ii) Only A'-movement can reconstruct for Condition A of binding.[2]

This study demonstrates that vP-internal scrambling is unambiguously A-movement, while clausal-internal movement can be both A- or A'-movement. Furthermore, cross-clausal scrambling out of non-finite clauses can exhibit both A- and A'-properties, while cross-clausal scrambling out of finite clauses is A'-movement. Additionally, in Section 3, it is argued that the distribution of movement types in different syntactic environments aligns with the position-based theory of the A-/A'-distinction that was established in Keine (2018). Specifically, it is argued that the type of movement, A- vs. A'-, is predictable from the scrambling environment and that A'-movement is only available in scrambling environments that can provide an available Spec,CP position as a landing site for such movement. Finally, potential instances of hyperraising out of finite clauses resulting from variations in grammaticality judgments are identified, and scope for further research is provided in Section 4.

1.3 A- and A'-Movement in Bangla

The movements involved in Bangla scrambling can be of two types: A- or A'-. The type of movement involved in scrambling can be identified using the following properties:

1. Only A-movement is known to obviate weak-crossover effects and lead to binding of reciprocal pronouns

2. Only A'-movement can reconstruct for Condition A of binding

An illustration of weak crossover obviation and reciprocal binding in Bangla is provided in (4) and (5), respectively:

(4) ***Weak crossover obviation***

a. o-r_1 ma **prot-ek-meye-ke$_2$** pochhondo kar-e
3SG-GEN mother.NOM **every-girl-ACC** like do-PRS
'Her mother likes every girl.' (bound reading impossible)

b. **prot-ek-meye-ke$_1$** o-r_1 ma **t_1** pochhondo kar-e
every-girl-ACC 3SG-GEN mother.NOM **t_1** like do-PRS
'For every girl x, x's mother likes x.'

In (4-a), the pronoun *or* 'his/her' cannot be co-indexed with *protek meye* 'every girl,' making a bound reading impossible. A-movement of the object, *protek meye* 'every girl' over

[2] A binding relation between two elements, A and B, is established when A c-commands B *and* both A and B are co-indexed in their binding domain. The following conditions govern the distribution of anaphors, pronouns, and R-expressions in their binding domains (from Carnie 2021):
Condition A: An anaphor must be bound in its binding domain.
Condition B: A pronoun must be free in its binding domain.
Condition C: An R-expression must be free.

the subject, *or ma* 'her mother', enables co-indexing and thereby binding of the subject-internal pronoun. This allows for a bound reading of the sort 'every girl is liked by her (own) mother' in (4-b).

(5) ***Reciprocal binding***

a. *ak-e-opor-er ma **Anup-aur-Pratap-ke** daak-lo
each other-GEN mother.NOM **Anup-and-Pratap-ACC** call-PST
'*Each other's mother, Anup and Pratap called.'

b. **Anup-aur-Pratap-ke** [ake-opor-er ma **t_1**] daak-lo
Anup-and-Pratap-ACC each.other-GEN mother.NOM **t_1** call-PST
'Anup and Pratap, each other's mother called, t_1.'

(5-a) is ungrammatical because the reciprocal pronoun (anaphor) *ake opor er* 'each other's' is unbound in its binding domain, leading to a violation of Condition A. A-movement of 'Anup and Pratap' in (5-b) provides a c-commanding antecedent to the reciprocal pronoun and enables binding.

Wh-movement is an instance of A'-movement, involving the movement of a question-word from a theta-position into a non-argument position for interpretation (Dayal 1994). That A'-movement cannot obviate weak crossover nor lead to reciprocal binding is demonstrated in (6) and (7), respectively.

(6) ***Weak crossover obviation***

a. ***o-r_1*** ma kon-meye-ke_1 bok-lo?
3SG-GEN mother.NOM which-girl-ACC scold-PST
'*Which $girl_1$ did her_1 mother scold?' (bound reading impossible)

b. *kon-meye-ke_1 **o-r_1** ma **t_1** bok-lo?
which-girl-ACC 3SG-GEN mother.NOM **t_1** scold-PST
'Which $girl_1$ did her_1 mother scold?'

A'-movement does not enable bound reading of the subject-internal pronoun.

(7) ***Reciprocal binding***

a. *ake-opor-er_1 ma-ra **kon du-to baccha-ke_1** bok-lo?
each other-GEN mother-PL.NOM **which two-CLF children-ACC** scold-PST
'*Which two $children_1$ did each other's mothers scold?'

b. ***kon du-to baccha-ke_1*** ake-opor-er_1 ma-ra **t_1** bok-lo?
which two-CLF children-ACC each other's mother-PL.NOM **t_1** scold-PST
'Which two children did each other's mother's t_1 scold?'

A'-movement of *kon duto baccha* 'which two children' over the reciprocal DP *ake-oper-er ma-ra* 'each other's mothers' does not provide an antecedent for binding.

However, A'-movement is known to be able to reconstruct. Reconstruction refers to the process where a movement operation is effectively reversed, restoring the structure to its pre-movement configuration for interpretation. This allows the binding principles to be applied as though the movement had never occurred (Barss 2001). In the example of

reconstruction provided in (8), the grammaticality of (8-b), despite an apparent violation of Condition A, demonstrates proper anaphor binding in its pre-movement structure in (8-a).

(8) a. Apu$_1$ **o-r$_1$** **kon** **chhobi** dekh-lo?
Apu.NOM **3SG-GEN** **which picture** see-PST
'Which picture of Apu$_i$ did he$_i$ see?'
b. **o-r$_1$** **kon** **chhobi** Apu$_1$ **t$_1$** dekh-lo?
3SG-GEN **which picture** Apu.NOM **t$_1$** see-PST
'Which picture of Apu$_i$ did he$_i$ see?'

2 Types of Scrambling

There are four distinct sub-classes of leftward scrambling. These are: 1) vP-internal movement; 2) clause-internal movement; 3) long-distance cross-clausal movement out of non-finite clauses; and 4) long-distance cross-clausal movement out of finite clauses.

2.1 vP-Internal Scrambling

vP-internal scrambling refers to the "permutation of the IODO order" inside the vP's domain (Sato & Goto 2014), as shown below:

(9) a. Apu [$_{vP}$ Keya-ke ***boi-ta*** di-lo] - [S IO DO V]
Apu.NOM Keya-DAT **book-CLF** give-PST
'Apu gave Keya the book.'
b. Apu [$_{vP}$ **boi-ta** Keya-ke **t$_1$** di-lo] - [S DO IO V]
Apu.NOM **book-CLF** Keya-DAT **t$_1$** give-PST
'Apu gave Keya the book.'

vP-internal scrambling in Bangla exhibits A-properties. This is illustrated using weak crossover obviation in (10). In (10-a), the pronoun *or boi* 'his book', is bound by Apu, indicating that the book belongs to Apu. Movement in (10-b) allows *protek meye* 'every girl' to bind the pronoun *or boi* 'their book,' providing a bound reading of the sort 'Apu gave every girl her book.'

(10) ***Weak crossover obviation***
a. Apu$_1$ [$_{vP}$ o-r$_{1/*2}$ boi-ta **prot-ek$_2$** **meye-ke** di-lo]
Apu.NOM 3SG-GEN book-CLF **every** **girl-ACC** give-PST
'Apu gave every girl his book.' (bound reading impossible)
b. Apu$_1$ [$_{vP}$ **prot-ek$_2$** **meye-ke** o-r$_{1/2}$ boi-ta **t$_1$** di-lo]
Apu.NOM **every** **girl-ACC** 3SG-GEN book-CLF **t$_1$** give-PST
'Apu gave every girl x, x's book.'

Converging evidence of A-movement can be found in reciprocal binding. It is shown in (11) that vP-internal scrambling provides a c-commanding antecedent to the unbound reciprocal pronoun.

(11) ***Reciprocal binding***

a. *Joy [$_{vP}$ ake-opor-er ma-baba-r-shathe **Rani-ar-Abhi-ke**
Joy.NOM each-other-GEN parent-PL-GEN-with **Rani-and-Abhi-ACC**
alap-kora-lo]
introduce-PST
'Joy introduced Rani and Abhi to each other's parents.'

b. Joy [$_{vP}$ **Rani-ar-Abhi-ke** ake-opor-er ma-baba-r-shathe **t**$_1$
Joy.NOM **Rani-and-Abhi-ACC** each-other-GEN parent-PL-GEN-with **t$_1$**
alap-kora-lo]
introduce-PST
'Joy introduced Rani and Abhi to each other's parents.'

(11-a) reflects the basic ditransitive word-order, where the reciprocal pronoun remains unbound, resulting in an ungrammatical construction because of a Condition A violation. On the other hand, in the derived structure (11-b), where the DO *Rani-ar-Abhi-ke* 'Rani and Abhi' undergoes vP-internal scrambling over the reciprocal pronoun *ake-opor-er* 'each other's', an antecendent is established for reciprocal binding. vP-internal scrambling can thus be A-movement in Bangla.

Sato & Goto (2014) similarly demonstrate that vP-internal scrambling in Japanese exhibits A-properties. Furthermore, they show that vP-internal scrambling in Japanese is unambiguously A-movement and cannot be A'-movement. An equivalent construction in Bangla demonstrates that this is also true in Bangla, as shown in (12).

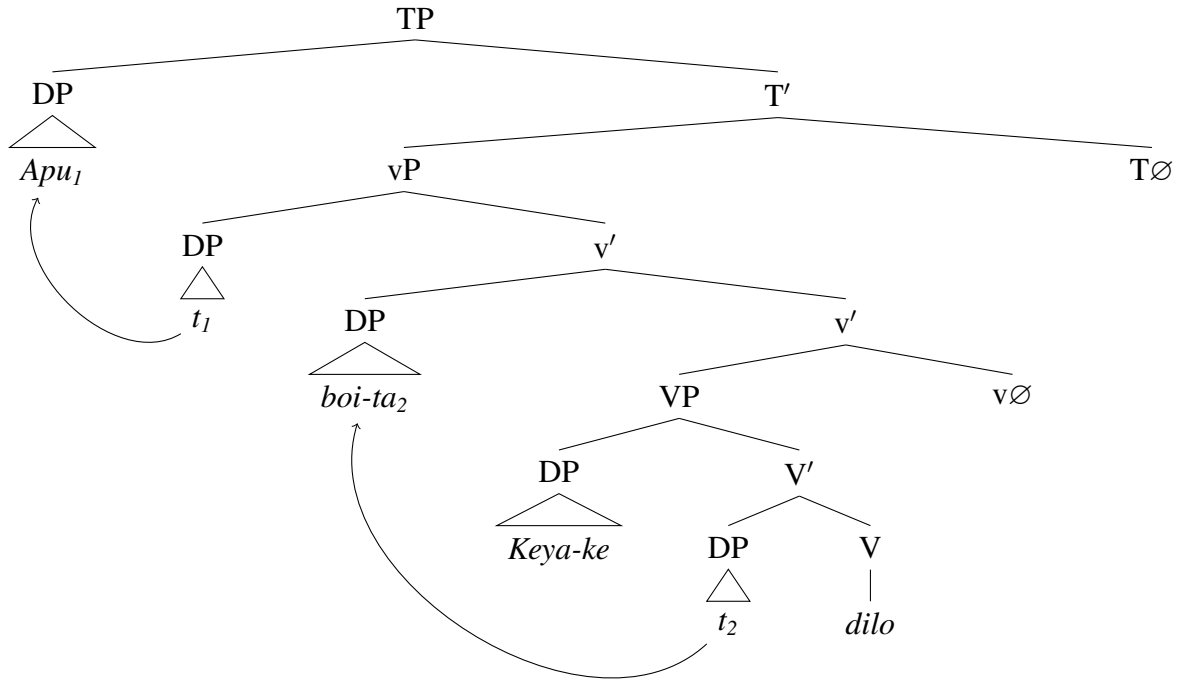
(12) a. Joy [$_{vP}$ Rani-ar-Abhi-ke **ake-opor-er-shathe** alap-kora-lo]
Joy.NOM Rani-and-Abhi-ACC **each-other-GEN-with** introduce-PST
'Joy introduced Rani and Abhi to each other.'

b. *Joy [$_{vP}$ **ake-opor-er-shathe** Rani-ar-Abhi-ke **t$_1$** alap-kora-lo]
Joy.NOM **each-other-GEN-with** Rani-and-Abhi-ACC **t$_1$** introduce-PST
'Joy introduced Rani and Abhi to each other.'

In this case, a grammatical reconstructed reading is unavailable. (12-a) provides the basic ditransitive word-order. The reciprocal pronoun, *ake-oper-er-shathe* 'with each other', is bound, making the sentence grammatical. However, movement of the reciprocal pronoun over *Rani and Abhi* in (12-b) is unacceptable. That is, such movement causes the reciprocal pronoun to A-bind the R-expression from the moved position, violating both Condition A (the reciprocal pronoun needs to be bound) and Condition C (the R-expression cannot be bound). This ungrammaticality is accurately predicted by A-movement, resulting in the exclusion of (12-b). However, if vP-internal scrambling were A'-movement, contrary to evidence in (12), the R-expression would be A-free, and Condition C violation would be evaded due to reconstruction. (12-b) shows that reconstruction by A'-movement is not available for vP-internal scrambling.

Therefore, this proves that vP-internal scrambling in Bangla is unambiguously A-movement. (13) provides the derivation of vP-internal A-movement in (9).

(13)



Apu boita Keya ke dilo 'Apu gave Keya the book.'

It is proposed that vP-internal scrambling targets an inner specifier of v, tucking in below the subject. This is necessary since the subject is seen as a more local goal by T_0 when its EPP probes.

2.2 Clause-Internal Scrambling

Clause-internal scrambling is the movement of an element across a subject to a sentence-initial position within the same clause (Sato & Goto 2014) as shown below:

(14) a. Apu **boi-ta** kin-lo
Apu.NOM **book-CLF** buy-PST
'Apu bought the book.'

b. **boi-ta** Apu **t_1** kin-lo
book-CLF Apu.NOM **t_1** buy-PST
'The book, Apu bought t_1.'

Clause-internal scrambling in Bangla exhibits both A- and A'-properties. Evidence of its A-properties comes from weak cross-over obviation, as shown in (15).

(15) ***Weak crossover obviation***

a. o-r_1 ma **prot-ek$_{*1/2}$** **baccha-ke** dekh-lo
3SG-GEN mother.NOM **every** **child-ACC** see-PST
'His/her mother saw every child.' (bound reading impossible)

b. **prot-ek$_1$** **baccha-ke** o-r_1 ma **t_1** dekh-lo
every **child-ACC** 3SG-GEN mother.NOM **t_1** see-PST
'For every child x, x's mother saw x.'

Movement of the object *protek baccha ke* 'every child' over the subject *or ma* 'his/her mother' provides a bound reading of the subject-internal pronoun. Furthermore, reciprocal binding, as in (16), also provides supporting evidence of A-movement in clause-internal scrambling environments; movement provides antecedent for reciprocal binding.

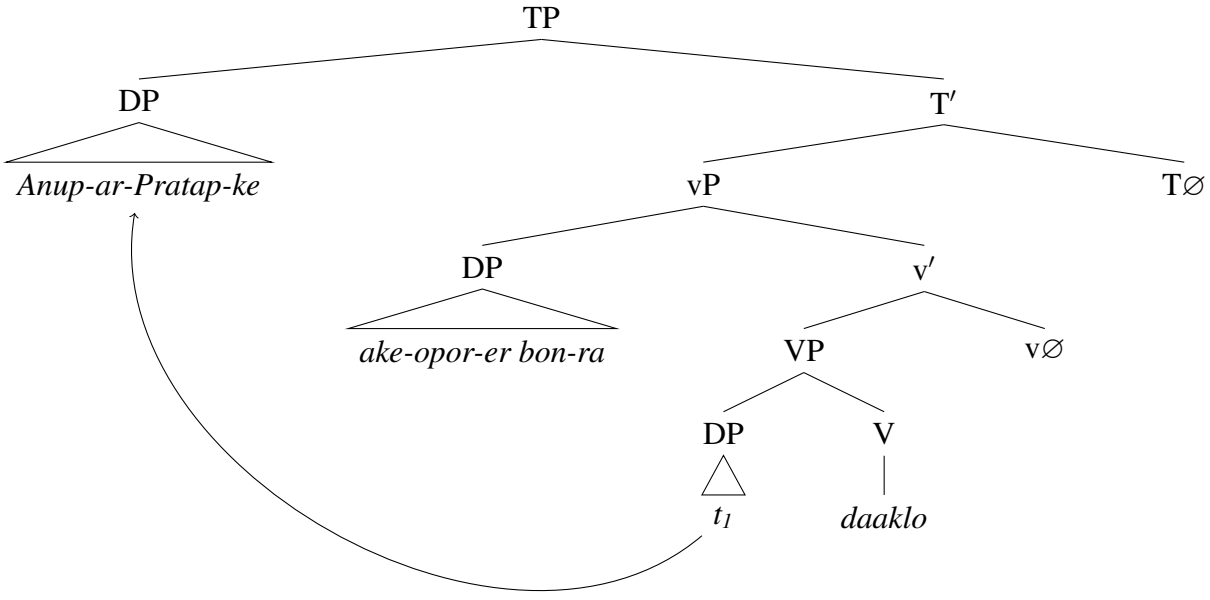
(16) ***Reciprocal binding***

a. **ake-oper-er$_1$ bon-ra* ***Anup-ar-Pratap-ke$_1$*** *daak-lo*
Each other's sister-PL **Anup and Pratap-ACC** call-PST
'*Each other's sisters called Anup and Pratap.'

b. ***Anup-ar-Pratap-ke$_1$*** *[ake-oper-er$_1$ bon-ra]* ***t_1*** *daak-lo*
Anup and Pratap-ACC Each other's sister- PL **t_1** call-PST
'Anup and Pratap, each other's sisters called t_1.'

A derivation of A-movement in clause-internal scrambling in (16) is given in (17).

(17)



Anup-ar-Pratap-ke ake-opor-er bon-ra daaklo 'Anup and Pratap, each other's sisters called t_1.'

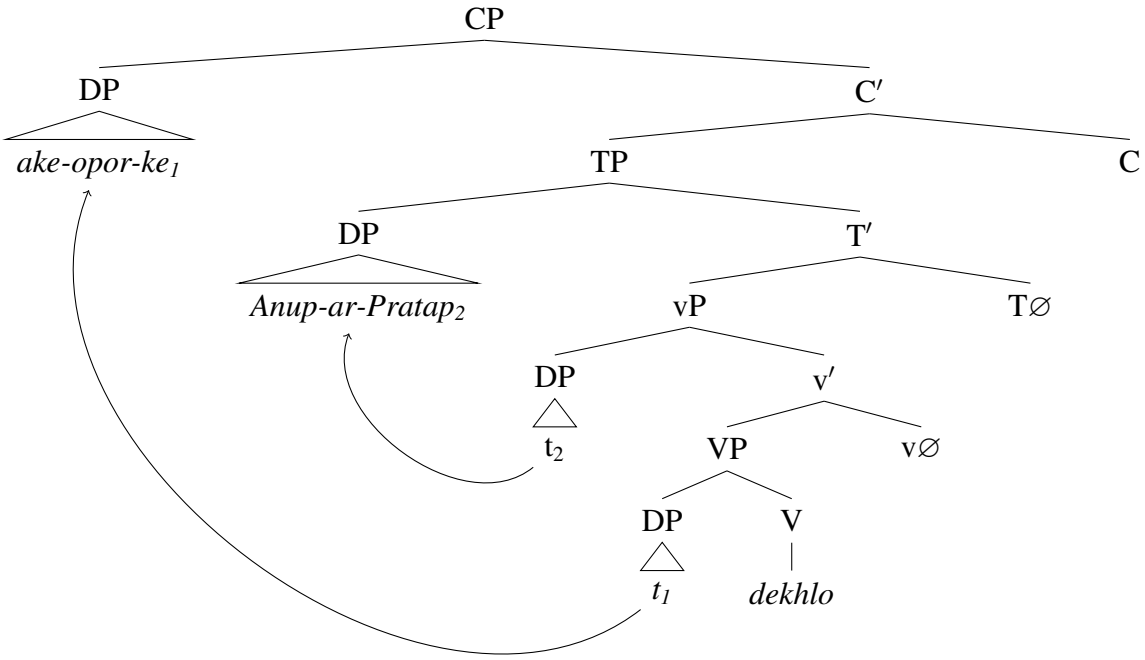
Hindi (Keine 2018) and Japanese (Sato & Goto 2014) also behave similarly in displaying A-movement in clause-internal scrambling. Furthermore, Hindi and Japanese, in their ability to reconstruct, also exhibit A'-properties in clause-internal scrambling (Keine 2018; Sato & Goto 2014). Equivalent phrases in Bangla reveal that clause-internal scrambling also exhibits A'-properties in Bangla, as demonstrated by reconstruction in (18).

(18) a. Anup-ar-Pratap **ake-opor-ke** dekh-lo
Anup and Pratap.NOM each-other-ACC see-PST
'Anup and Pratap saw each other.'

b. **ake-opor-ke** [Anup-ar-Pratap **t_1**] dekh-lo
Each-other-ACC Anup and Pratap.NOM **t_1** see-PST
'Each other, Anup and Pratap saw t_1.'

(18-a) shows the basic grammatical word order that follows both Conditions A and C in that the reciprocal pronoun is bound and, the R-expression is free. The grammaticality of (18-b) is evidence of reconstruction because the scrambled reciprocal pronoun does not induce violation of Condition C. The R-expression *Anup and Pratap* remains A-free, thereby avoiding violation of Condition C. Therefore, clause-internal scrambling can also be A'-movement.
The derivation of A'-movement in (18-b) is illustrated in (19).

(19)



ake-opor-ke Anup-ar-Pratap marlo 'Each other, Anup and Pratap saw t_1

2.3 Cross-Clausal Scrambling

Cross-clausal scrambling is the movement of an element to a sentence-initial position across a clause boundary (Sato & Goto 2014). Cross-clausal movement can occur out of both non-finite clauses (20) and finite clauses (21) (Keine 2018).

(20) ***Cross-clausal movement out of non-finite clauses***

a. Apu **Keya-ke** dekh-te chai-lo
Apu.NOM **Keya-ACC** see-INF want-PST
'Apu wanted to see Keya.'

b. **Keya-ke** Apu [$_{TP}$ **t$_1$** dekh-te] chai-lo
Keya-ACC Apu.NOM **t$_1$** see-INF want-PST
'Keya, Apu wanted to see t_1.'

(21) ***Cross-clausal movement out of finite clauses***

a. Apu bhab-lo [$_{CP}$ je Keya **shobai-ke** dekh-e-che]
Apu.NOM think-PST that Keya **everyone-ACC** see-PRF-PRS
'Apu thought that Keya had seen everyone.'

b. **shobai-ke** Apu bhab-lo [$_{CP}$ je Keya **t$_1$** dekh-e-che]
everyone-ACC Apu.NOM think-PST that Keya **t$_1$** see-PRF-PRS
'Everyone, Apu thought that Keya had seen t_1.'

The two scrambling environments vary in the types of movement they allow out of them. While movement out of non-finite clauses resembles clause-internal scrambling, allowing both A- and A'-movement, movement out of finite clauses seems to be restricted to A'-movement.

2.3.1 Cross-clausal scrambling out of non-finite clauses

As stated above, cross-clausal scrambling out of non-finite clauses exhibits both A- and A'-properties. Evidence of A-movement can be found in weak crossover obviation (22) and binding of reciprocal pronoun (23).

(22) ***Weak crossover obviation***

a. [o-r$_{1/*2}$ ma] [$_{TP}$ **prot-ek$_2$** **baccha-ke** dekh-te] chai-lo
3SG-GEN mother.NOM every child-ACC see-INF want-PST
'His/her mother wanted to see every child.' (bound reading impossible)

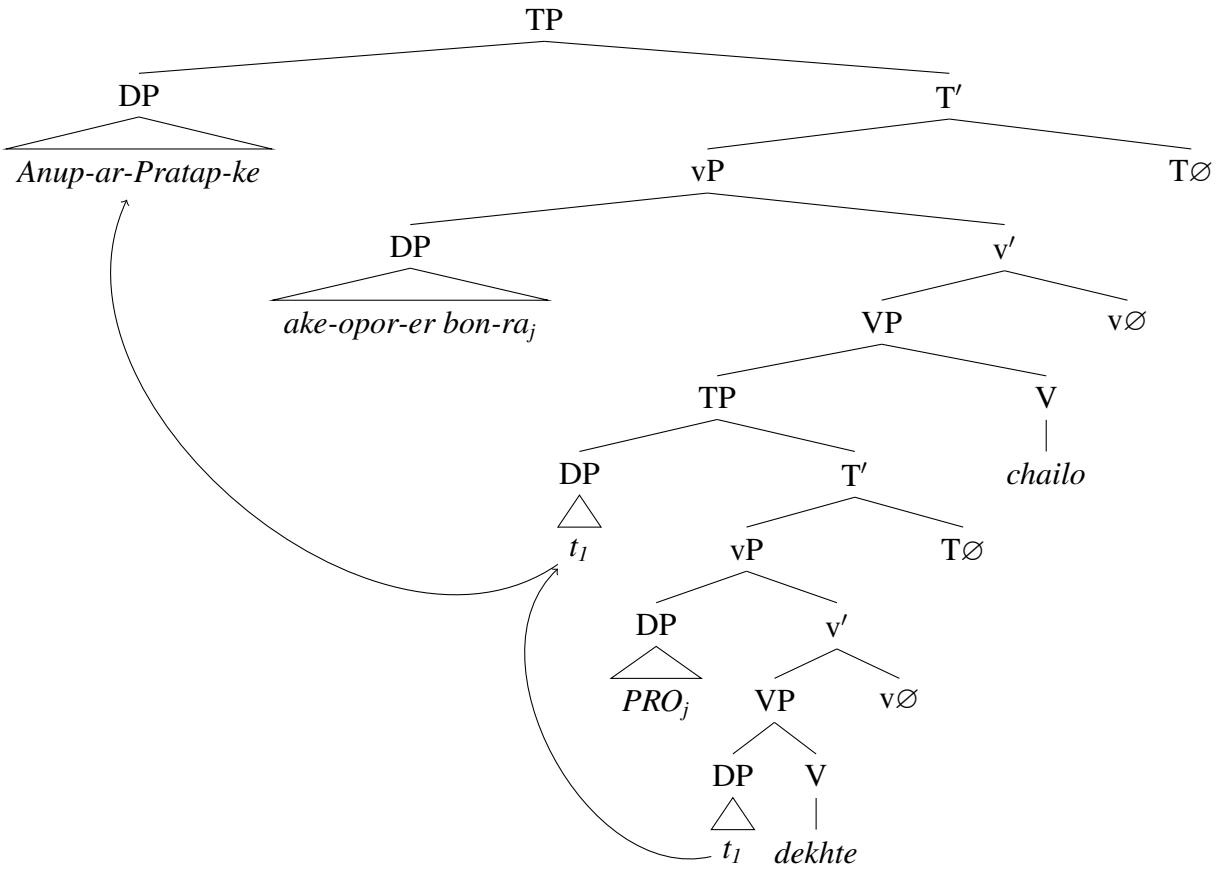
b. **prot-ek$_1$** **baccha-ke** **[o-r$_1$** ma] [$_{TP}$ **t$_1$** dekh-te] chai-lo
every child-ACC 3SG-GEN mother.NOM **t$_1$** see-INF want-PST
'For every child x, x's mother wanted to see x.'

(23) ***Reciprocal binding***

a. [*ake-oper-er$_1$ bon-ra] [$_{TP}$ **Anup-ar-Pratap-ke$_1$** dekh-te] chai-lo
Each other's sister-PL **Anup-and-Pratap-ACC** see-INF want-PST
'*Each other's sisters wanted to see Anup and Pratap.'

b. **Anup-ar-Pratap-ke$_1$** [ake-oper-er$_1$ bon-ra] [$_{TP}$ **t$_1$** dekh-te] chai-lo
Anup-and-Pratap-ACC Each other's sister-PL **t$_1$** see-INF want-PST
'Anup and Pratap, each other's sisters wanted to see t$_1$.'

The derivation of reciprocal binding as in (23) is given in (24).



(24)

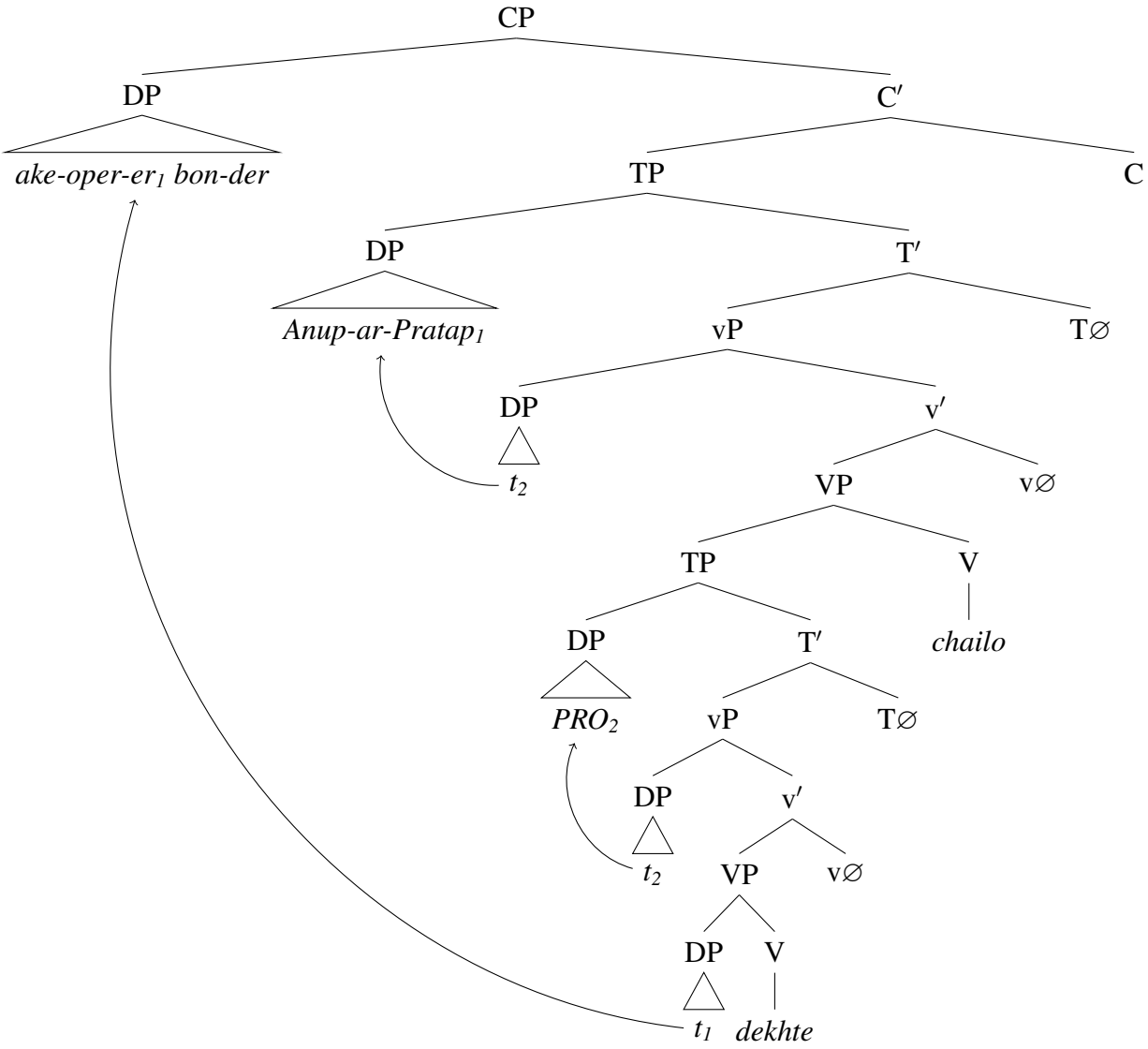
Anup-ar-Pratap ke ake-opor-er bon-ra dekhte chailo
'Anup and Pratap, each other's sisters wanted to see t$_1$.

Movement out of non-finite clauses can also be A'-movement, as shown in (25), and derived in (26).

(25) a. **Anup-ar-Pratap$_1$** [$_{TP}$ **ake-oper-er$_1$** bon-der dekh-te] chai-lo
Anup and Pratap.NOM each other's sister-PL see-INF want-PST
'Anup and Pratap wanted to see each other's sisters.'
(*Reciprocal pronoun is bound by Anup and Pratap.*)

b. **[ake-oper-er$_1$ bon-der] Anup-ar-Pratap$_1$** [$_{TP}$ t$_1$ dekh-te] chai-lo
each other's sister-PL Anup and Pratap-ACC t$_1$ see-INF] want-PST
'Each other's sisters, Anup and Pratap wanted to see.'

(26)



Ake-opor-er bon-der Anup-ar-Pratap dekhte chailo
'Each other's sisters, Anup and Pratap wanted to see.'

(25-a) presents the basic word-order, which follows both Conditions A and C of binding. (25-b) shows a grammatical sentence with scrambled word order that violates both binding conditions; the R-expression is bound, and the reciprocal pronoun is not. The grammaticality of (25-b) is evidence of reconstruction, and thereby of A'-movement.

2.3.2 Cross-clausal scrambling out of finite clauses

In Bangla, cross-clausal scrambling out of finite clauses does not display A-properties. While movement out of a finite sentence is possible, it does not lead to binding of the subject-internal pronoun *or ma* 'his/her mother' by the object *prot-ek baccha ke* 'every child', as shown in (27).

(27) ***Weak crossover obviation***

a. [o-$r_{1/*2}$ ma] bhab-lo [$_{CP}$ je Anup **prot-ek$_2$ baccha-ke**
3SG-GEN mother.NOM think-PST that Anup.NOM every child-ACC
dekh-e-che]
see-PRF-PRS
'His/her mother thought that Anup had seen every child.'

b. **prot-ek$_2$ baccha-ke** [o-$r_{1/*2}$ ma] bhab-lo [$_{CP}$ je Anup **t_1**
every child-ACC 3SG-GEN mother.NOM think-PST that Anup.NOM **t_1**
dekh-e-che]
see-PRF-PRS
'His/her mother thought that Anup had seen every child.'

A bound reading is not obtained despite movement. Since this movement does not obviate weak crossover, it is thereby classified as an A'-movement. Reciprocal binding also provides supporting evidence. In (28), movement of *Anup-ar-Pratap* 'Anup and Pratap-ACC' over the reciprocal pronoun *ake opor er* 'each other's' does not lead to reciprocal binding. Hence, scrambling out of finite clauses is unambiguously A'-movement.

(28) ***Reciprocal binding***

a. *ake-oper-er_1 bon-ra bhab-lo [$_{CP}$ je Keya
each other's sister-PL think-PST that Keya.NOM
Anup-ar-Pratap-ke dekh-e-che]
Anup-and-Pratap-ACC see-PRF-PRS
'*Each other's sisters thought Keya had seen Anup and Pratap.'

b. ***Anup-ar-Pratap-ke$_1$** ake-oper-er_1 bon-ra bhab-lo [$_{CP}$ je Keya
Anup-and-Pratap-ACC each other's sister-PL think-PST that Keya.NOM
t_1 dekh-e-che]
t_1 see-PRF-PRS
'Anup and Pratap, each other's sisters thought that Keya had seen t_1.'

In sum, Bangla exhibits the following properties in different scrambling environments:

(29) vP-internal scrambling is unambiguously A-movement.
Clause-internal scrambling can be A- or A'-movement.
Cross-clausal movement out of non-finite clauses can be A- or A'-movement.
Cross-clausal movement out of finite clauses in unambiguously A'-movement.

The varying properties of movement in the different scrambling environments can be ex-

plained based on the structure of clauses and the positions targeted by A- and A'-movement.

3 A Position-Based Account of Bangla Scrambling

The positional properties of A- and A'-movement in Bangla mirror the properties of movement in Hindi, as shown in Keine (2018). Equivalent constructions in Bangla are used to determine the structure of clauses and the positions involved in A- and A'-movement.

3.1 The Structure of Embedded Clauses

Keine (2018) has demonstrated that in Hindi, finite clauses are CPs, whereas non-finite clauses, which lack a CP layer, are TPs. This difference in structure is determined based on two observations: Firstly, Hindi finite embedded clauses may contain the complementizer *ki*, but non-finite clauses may not. Secondly, interrogative scope is associated with finite clauses and not non-finite clauses, which means that non-finite clauses lack an embedded-question reading. The standard assumption that interrogative scope is associated with C explains why it is absent in non-finite clauses, which lack a CP layer. Furthermore, complementizers are also known to sit in C, and the lack of a CP layer explains why they are absent in non-finite clauses. Therefore, non-finite clauses are structurally smaller than finite clauses (Keine 2018) and are classified as TPs.

Similarly, Bangla finite clauses also may contain the complementizer *je* (30), but non-finite clauses may not (31).

(30) Apu bhab-lo [$_{CP}$ **je** Keya shobai-ke dekh-e-che]
Apu.NOM think-PST that Keya-ACC everyone see-PRF-PRS
'Apu thought that Keya had seen everyone.'

(31) Apu [$_{TP}$ ***je** Keya-ke dekh-te] chai-lo
Apu.NOM *that Keya-ACC see-INF want-PST
'Apu wants to see Keya.'[3]

Again, in Bangla, only finite clauses provide an interrogative scope position, but non-finite clauses do not. The wh-element *ki* 'what' takes wh-scope within the embedded finite sentence, like in Hindi (Keine 2018); a matrix-question interpretation is impossible because finite-clauses are islands for wh-scope. In non-finite clauses, however, an embedded-question interpretation is impossible, and the wh-element in (33) takes mandatory matrix scope.

(32) tumi jaano [$_{CP}$ **je** o ki kor-e-che]
you know that 3SG.NOM what do-PRF-PRS
'You know what he did.'

[3] This sentence might have a relative clause reading, as in "Apu, who wanted to see Keya"; or something like "Oh, but Apu wanted to see Keya!".

(33) tumi [$_{TP}$ **ki** kor-te] jaano?
you what do-INF know
'What do you know to do?'

The evidence therefore leads to the same conclusion for Bangla (33).

(34) a. Finite clauses in Bangla are CPs.
b. Non-finite clauses in Bangla lack a CP layer; they are TPs.

3.2 Positions Targeted by A- and A'-Movement

Once again, evidence from Hindi (Keine 2018) demonstrates that A-movement lands in Spec,TP (and TP-internal positions), whereas A'-movement lands in Spec,CP. Similar evidence confirms that this distinction also applies to Bangla.

3.2.1 A-movement lands in Spec,TP (and TP-internal positions)

Keine (2018) presents novel evidence that Spec,TP, and TP-internal positions serve as landing sites of the A-movement in Hindi. To illustrate the same in Bangla, an embedded non-finite clause is extraposed to the right to demarcate the right edge in (35). This extraposition ensures that movement remains contained within the non-finite clause rather than resulting in extraction out of it.

(35) Keya cheye chilo [$_{TP}$ prot-ek meye-ke$_1$ [o-r$_1$ biye-r shomoy **t$_1$**
Keya.NOM want AUX every girl-ACC 3SG-GEN wedding-GEN time **t$_1$**
dekh-te]
see-INF
'Keya wanted to see every girl x during x's wedding.'

The embedded DO *protek meye* 'every girl' moves over the adjunct *or biyer shomoy* 'during her wedding' and can bind the internal pronoun *or* 'her' from its landing site. This is clear evidence of A-movement.

Since extraposition prevents movement outside the non-finite clause, the landing site of *protek meye* 'every girl' must be within the non-finite clause. Consequently, (35) demonstrates that A-movement can target a position internal to a non-finite clause. Furthermore, based on evidence that non-finite clauses are TPs that lack a CP layer, A-movement in Bangla must also land in Spec,TP and TP-internal positions.

3.2.2 A'-movement lands in Spec, CP

In contrast to A-movement, A'-movement targets TP-external positions in Hindi (Keine 2018). The same can be demonstrated for Bangla as well. (36) consists of sentences in a double embedding structure where a finite clause is embedded within a non-finite clause, which in turn is embedded within a finite matrix clause.

(36) ***A'-movement cannot land inside a non-finite clause***

a. [$_{CP}$ ami chai [$_{TP}$ bol-te [$_{CP}$ je ami **boi-ta** pod-e niy-e-chi]
1SG.NOM want say-INF that 1SG book-CLF read take-PRF-PRS
'I want to say that I have read the book.'

b. [$_{CP}$ *ami chai [$_{TP}$ **boi-ta** bol-te [$_{CP}$ je ami **t$_1$** pod-e niy-e-chi]
1SG.NOM want **book-CLF** say-INF that 1SG **t$_1$** read take-PRF-PRS
'*I want to **the book** say that I have read t$_1$.'

c. [$_{CP}$ **boi-ta** ami chai [$_{TP}$ bol-te [$_{CP}$ je ami **t$_1$** pod-e niy-e-chi]
book-CLF 1SG.NOM want say-INF that 1SG **t$_1$** read take-PRF-PRS
'**The book** I want to say that I have read t$_1$.'

Both (36-b) and (36-c) depict movement out of finite clauses, and hence, must be A'-movement (given that finite clauses allow only A'-movement out of them, as demonstrated in section 2.3.3) Converging with evidence in Hindi (Keine 2018), the ungrammaticality of (36-b) demonstrates that A'-movement in Bangla cannot land inside a non-finite clause. On the other hand, (36-c) shows that A'-movement can land in finite clauses.

Therefore, the ungrammaticality of (36-b) must stem from the difference in the structural properties of finite and non-finite clauses. Non-finite clauses, which obligatorily lack a CP layer, simply lack the "functional structure" needed for A'-movement landing site. In contrast, finite clauses, with their CP layer, offer this landing site for A'-movement. This, therefore, must indicate that A'-movement targets TP-external, Spec,CP positions.

In sum, A- and A'-movement target the following positions in Bangla:

(37) a. A-movement lands in Spec,TP (or TP-internal) positions
b. A'-movement lands in Spec,CP.

4 Discussion

The conclusions in (37) predict the different properties of A- and A'-movement in the different scrambling environments. Reiterating the key observations presented in Section 2: vP-internal scrambling is unambiguously A-movement, whereas clause-internal movement may be both A- and A'-movement. Further, cross-clausal movement out of non-finite clauses again exhibits properties of both A- and A'-movement, but cross-clausal movement out of finite clauses can only be A'-movement.

vP-internal scrambling can only be A-movement because the VP-internal structure does not have the functional structure necessary for providing a landing site for A'-movement. Clause-internal scrambling, on the other hand, can be both A- and A'-movement because the structure of the clause provides landing sites for both kinds of movement. Specifically, A-movement can target Spec,TP, enabling it to establish binding relations, while A'-movement can occupy a higher Spec,CP position, facilitating reconstruction in the clause.

Furthermore, in cross-clausal environments, movement out of non-finite embedded clauses exhibits properties of both A- and A'- movement. This also follows from the

fact that the structure of the non-finite clause can provide landing sites for both types of movement. A-movement out of the embedded non-finite clauses can land in the Spec,TP position of the higher clause. Again, non-finite clauses are transparent to A'-movement because movement out of a non-finite clause can land in the Spec,CP position of the higher clause, hence leading to reconstruction.

Movement out of a finite (i.e. CP) clause is A'-movement; it can only target an A'-position. That is, movement out of an embedded finite clause must obligatorily proceed through Spec,CP of the embedded clause and therefore can only land in the Spec,CP position of the higher matrix clause but not a lower TP-internal position. This is described as a *Ban on Improper Movement*.

(38) *Ban on Improper Movement*
Movement out of Spec,CP must land in Spec,CP. Movement from Spec,CP to a TP-internal position is ruled out. (from Keine 2018:22)

Converging with the evidence in Hindi (Keine 2018), finite clauses in Bangla allow A'-movement out of them because such movement lands in Spec,CP of the higher clause. The lack of a CP layer in embedded non-finite clauses allows A-movement out of them.

The ban on A-movement out of finite clauses can also be explained in terms of phase-boundaries. A'-positions (Spec,CP) are generally known to be phase-edge positions, while A-positions (Spec,TP and TP-internal) are phase-internal positions. A-movement does not cross phase boundaries, and therefore, "movement may not proceed from a phase edge to a phase-internal position" (Keine 2018).

In conclusion, this study distinguishes the different types of movement involved in Bangla scrambling, and provides an account of the properties exhibited by A- and A'-movement in four scrambling environments using a position-based account.

Bangla-scrambling has also been known to exhibit right-ward movement (David 2015; Bhatt & Dayal 2007). This can be seen in the following example (39):

(39) a. **t_1** Joy-ke boi-ta di-lo **Rani**
t_1 Joy-ACC book-CLF give-PST Rani.NOM
'To Joy gave book, Rani.'
b. am-ar **t_1** ach-e **ek-ti** **darun dharona**
1SG-GEN t_1 be-PRS one-CLF.DIM great idea
'I have a great idea.'

The properties of right-ward scrambling in Bangla form the next crucial step in this research. Additionally, Bangla scrambling is also widely noted in wh-constructions. wh-elements can remain in-situ (40-a), undergo intermediate movement (40-b) (40-c), and right-ward movement (40-d), as shown below in (40).

(40) a. **ke** dilo Rani-ke boi-ta? - Who gave Rani the book?
b. Rani-ke **ke** dilo boi-ta? - Rani who gave the book?
c. boi-ta **ke** dilo Rani-ke? - The book gave Rani who?
d. dilo Rani-ke boi-ta **ke**? - Gave Rani the book who?

A comprehensive account of A'-movement in question-constructions warrants further examination.

Furthermore, certain speakers of Bangla agree to a bound reading in constructions involving movement out of finite clauses (27) as shown below:

(41) ***Weak crossover obviation***

a. [o-$r_{1/*2}$ ma] bhab-lo [$_{CP}$ je Anup **prot-ek$_2$ baccha-ke**
3SG-GEN mother.NOM think-PST that Anup.NOM every child-ACC
dekh-e-che]
see-PRF-PRS
'His/her mother thought that Anup had seen every child.'

b. **prot-ek$_2$ baccha-ke** [o-r_2 ma] bhab-lo [$_{CP}$ je Anup **t$_1$**
every child-ACC 3PL-GEN mother.NOM think-PST that Anup.NOM **t$_1$**
dekh-e-che]
see-PRF-PRS
'Every child x's mother thought that Anup had seen x.'

Evidence in (41-b) demonstrates that movement out of finite clauses in Bangla can feed binding, providing support for A-movement. This contrasts with the ungrammaticality observed in (27), where such movement is disallowed. This indicates that Bangla can allow hyperraising out of finite clauses, contrasting with the evidence in Hindi (Keine 2018). Notably, this suggests that Bangla permits hyperraising out of finite clauses, differing from Hindi, as reported by Keine (2018). Interestingly, this variation appears to be influenced by speakers' exposure to Hindi. Speakers of Bangla from Northern Indian states, where Hindi has a greater influence, tend to disallow such constructions, while those from West Bengal accept bound readings. This phenomenon offers an intriguing avenue for exploring how Bangla's clause structure may diverge from Hindi, despite the two languages often being grouped together. The findings raise compelling questions about syntactic locality and the CP phase hypothesis, with potential implications for understanding cross-linguistic variation in clause structure. The underlying causes of this variation and its broader implications for Bangla's syntax merit further investigation.

References

Barss, Andrew. 2001. Syntactic reconstruction effects. *The handbook of contemporary syntactic theory* 670–696.

Bhatt, Rajesh & Veneeta Dayal. 2007. Rightward scrambling as rightward remnant movement. *Linguistic Inquiry* 38(2). 287–301.

Bhattacharya, Tanmoy. 2002. Peripheral and clause-internal complementizers in Bangla: A case for remnant movement. In *Proceedings of the western conference on linguistics 2000*, 100–112. Department of Linguistics, California State University, Fresno Fresno.

Carnie, Andrew. 2021. *Syntax: A generative introduction*. John Wiley & Sons.

Cho, Jai-Hyoung. 1994. On scrambling: reconstruction, crossover, and anaphor binding. *Theoretical issues in Korean linguistics,* 255–274.

Dash, Niladri Sekhar. 2015. *A descriptive study of bengali words*. Cambridge University Press.

David, Anne Boyle. 2015. *Descriptive grammar of Bangla*, vol. 2. Walter de Gruyter GmbH & Co KG.

Dayal, Veneeta. 1994. Binding facts in Hindi and the scrambling phenomenon. *Theoretical perspectives on word order in South Asian languages* 237–262.

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

Guha, Anamita. 2013. An optimality-theoretic approach to Bangla information structure. In *Paper proceedings of international conference on language, literature & linguistics 2013 (online)*, 179.

Islam, J. 2016. The necessity of covert wh-movement in sov languages: Re-thinking the overt movement proposal for Bangla.

Keine, Stefan. 2018. Case vs. positions in the locality of a-movement. *Glossa: a journal of general linguistics* 3(1).

Lewis, M. Paul (ed.). 2009. *Ethnologue: Languages of the world*. Dallas, TX, USA: SIL International sixteenth edn.

Mahajan, Anoop. 1994. Toward a unified theory of scrambling. *Studies on scrambling: Movement and non-movement approaches to free word-order phenomena* 301–330.

Mahajan, Anoop Kumar. 1990. *The A/A-bar distinction and movement theory*: Massachusetts Institute of Technology dissertation.

Ross, John Robert. 1967. Constraints on variables in syntax. .

Saito, Mamoru. 1985. *Some asymmetries in Japanese and their theoretical implications*: NA Cambridge dissertation.

Saito, Mamoru. 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1. 69–118.

Sato, Yosuke & Nobu Goto. 2014. Scrambling. In *The routledge handbook of syntax*, 264–282. Routledge.

Simpson, Andrew & Tanmoy Bhattacharya. 2003. Obligatory overt wh-movement in a wh-in-situ language. *Linguistic Inquiry* 34(1). 127–142.

Simpson, Andrew & Saurov Syed. 2014. Finiteness, negation and the directionality of headedness in Bangla. *Natural Language & Linguistic Theory* 32. 231–261.

Syed, Saurov. 2017. *The structure of noun phrases in Bengali: What it tells us about phases and the universality of DPs*: University of Southern California dissertation.

Thompson, Hanne-Ruth. 2004. *Toward a definitve grammar of Bengali: A practical study and critique of research on selected grammatical structures*. University of London, School of Oriental and African Studies (United Kingdom).

Thompson, Hanne-Ruth. 2020. *Bengali: A comprehensive grammar*. Routledge.

Relative Deletion

RAJESH BHATT, *University of Massachusetts, Amherst*
ANUSHREE MISHRA, *EFLU, Hyderabad*

ABSTRACT

Hindi-Urdu exhibits a lesser-known form of ellipsis known as Relative Deletion (RD) (Mishra 2024; van Craenenbroeck & Lipták 2006), where verbal and phrasal material in relative clauses is elided, leaving only the relative phrase and one or more remnants. This study presents an in-depth analysis of RD, examining its behavior across various syntactic structures, including equatives and temporal/locative relative clauses. We examine the influence of case-marking on, and location of, the relative pronoun on the well-formedness of RD. The study compares RD with sluicing and gapping, highlighting their locality profiles and constraints. Notably, RD requires the antecedent to originate within the clause to which the relative clause is attached, a feature that parallels restrictions found in English gapping (Johnson 2009) and not sluicing (Ross 1969). In addition, we explore apparent instances of non-local RD, where deletion seems to cross clause boundaries, posing a syntactic puzzle that raises further questions about the mechanisms of ellipsis in Hindi-Urdu.

1 Introduction

This paper explores the phenomenon of Relative Deletion (RD) in Hindi-Urdu (HU), focusing specifically on comparatives and equatives in the language, alongside other insights from relatives of degree, time, and location. Relative Deletion refers to the ellipsis of all verbal material in relative clauses, leaving behind only the relative phrase and at least one phrasal remnant. This phenomenon has been discussed in the context of Focus-based Sluicing, a framework recently extended by Mishra (2024), following van Craenenbroeck & Lipták's (2006) analysis of Relative Deletion in Hungarian (2). (1)) below contains an instance of Relative Deletion in Hindi-Urdu in the form of a relative clause that emphasizes the retention of key elements while the rest of the verbal structure is elided.

(1) mɛ̃=ne Seema=ko vo=hi: film dɪkʰa-yi tʰi: [jo=(ki) Rita=ko
I=ERG Seema=DAT that=ONLY film show-PFV.F be.PST.F REL=that Rita=DAT
(*tʰiː)]
be.PST.F
'I showed Seema the movie which I showed to Rita.

As can be seen from both the syntactic structure and the translation, the construction is marked by emphatic focus, signaled either prosodically or morphologically through the use of *hi:* 'FOC-only that'. The ellipsis that occurs here reflects the deletion of the verbal material after the focused element moves to a higher position in the left periphery, as proposed

by Mishra (2024). Relative Deletion in Hungarian follows a similar pattern, as shown in Van Craenenbroeck and Lipták (2006):

(2) Péternek AZT A FOTÓT mutattam meg amit ANNÁNAK
Péter-DAT that-ACC the photo-ACC showed PV REL-what-ACC Anna-DAT
'The photo I showed to Péter was the one that I showed to Anna.'

(van Craenenbroeck & Lipták, 2006, 3)

In both languages, Relative deletion operates as a type of Focus-Based TP/IP Deletion, where the focused element, along with the relative complementizer is moved to a higher position, with the rest of the relative clause being elided. The remnant, often a (case-marked) NP, survives deletion, providing the sentence's necessary interpretive content.

This paper seeks to extend the existing analysis of Relative deletion by applying it to other relativization structures in Hindi-Urdu, specifically focusing on constructions involving degrees, such as comparatives and equatives, as well as temporal and locative relatives. Additionally, we aim to contrast Relative deletion with other ellipsis phenomena, such as sluicing and gapping, to figure out its place in cross-linguistic syntax.

2 RD in other relativization structures

In Hindi-Urdu, relativization structures are not limited to relativization of individuals. Relativization can involve time, space and degrees yielding *when/until* clauses, *where* clauses and equative/comparative clauses respectively and we find that relative deletion is possible with non-individual relativization structures.

2.1 Relativization Structures in Hindi-Urdu

Let's start with the observation from Srivastav (1991) that finite relativization structures in Hindi-Urdu come in three distinct structures/orders.

(3) a. Correlative: left adjoined to a clausal projection
[jo laṛki: khaṛi: hɛ] [vo laṛki: lambi: hɛ]
REL girl.F tall.F BE.PRS.3SG DEM girl.F tall.F BE.PRS.3SG
'The girl who is standing is tall.' (Which girl is standing, she is tall.)

b. Embedded: adjoined to a nominal projection
vo laṛki: [jo (*laṛki:) khaṛi: hɛ] lambi: hɛ
DEM girl.F REL girl.F tall.F BE.PRS.3SG tall.F BE.PRS.3SG
'The girl who is standing is tall.'

c. Extraposed: right adjoined to a clausal projection
vo laṛki: lambi: hɛ [jo (*laṛki:) khaṛi: hɛ]
DEM girl.F BE.PRS.3SG tall.F BE.PRS.3SG REL girl.F tall.F
'The girl is tall who is standing.'

Srivastav (1991) shows that left adjoined relatives (3a) constitute a distinct relativization strategy, correlativization, where the relative clause picks out a maximal entity that the main clause is predicated of. She also shows that embedded relatives (3b) can be seen as restrictive relativization as familiar from English and that the right adjoined cases (3c) pattern with embedded relatives. Note that the relative clause in (3b, c) does not allow for internal heads while the relative clause in the correlative structure in (3a) does. The correlative structure and the right adjoined structure can be constructed naturally with equatives.[1]

(4) a. correlative:

[jitni: (kita:bẽ) Mina k^hari:de-gi:] [utni: (kita:bẽ)
HOW.MANY.F books.F Mina.F buy-FUT.3FSG THAT.MANY.F books.F
Ram k^hari:de-ga:]
Ram.M buy-FUT.3MSG

'Ram will buy as many books as Mina.' (literally: [how many books Mina will buy], Ram will buy that many books.)

b. right adjoined:

[Ram utni: (kita:bẽ) k^hari:de-ga:] [jitni: (???kita:bẽ)
Ram THAT.MANY.F books.F buy-FUT.3MSG HOW.MANY.F books.F
Mina k^hari:de-gi:]
Mina.F buy-FUT.3FSG

'Ram will buy as many book as Mina.'

As with (3), we see that the relative clause in the correlative in (4a) allows for an internal head while the right adjoined structure, which we take to be derived from an externally headed relativization structure, does not (4b).

2.2 The distribution of Relative Deletion

As noted earlier, relative deletion is attested with instances of non-individual relativization and we will turn to these cases in this section. The following generalization emerges: relative deletion is possible inside extraposed relative clauses but not inside correlative clauses. A correlative clause precedes the main clause while an extraposed clause follows the main clause. There are also structural differences between correlative clauses and extraposed clauses.

[1]The embedded structure feels degraded for reasons we do not understand.

i. ???[Ram utni: kita:bẽ [jitni: Mina k^hari:de-gi:] k^hari:de-ga:]
Ram.M THAT.MANY.F books.F HOW.MANY.F Mina.F buy-FUT.3FSG buy-FUT.3MSG
intended: 'Ram will buy as many books as Mina.'

In addition, our focus in this paper is on ellipsis and ellipsis is blocked in embedded structures because of irresolveable antecedent containment. Therefore we will not consider embedded structures further.

2.2.1 Relative Deletion in Equatives

In (4), we have seen that equatives can be realized as a correlative and also as a right adjoined relative. We know from §1 that relative deletion structures are most natural when the modified XP is focus marked with the particle *-hi:*. This particle brings in a range of exclusive meanings and so we will gloss it as 'only' (see Bajaj (2016)) but the one that seems most prominent in the relative deletion context corresponds to 'the same', 'the one', expressing identity between the element in the main clause and its counterpart in the relative clause. Relative deletion is possible with equatives but only when the HOW.MANY clause follows the main clause i.e. in the right adjoined structure but not in the correlative structure.

(5) a. right adjoined: relative deletion is possible

Ram utni:=hi: kita:bẽ kʰariːde-gaː [jitni: Mina
Ram THAT.MANY.F-ONLY books.F buy-FUT.3MSG HOW.MANY.F Mina.F
~~kʰariːde-giː~~]
buy-FUT.3FSG

'Ram will buy as many bookz as Mina.'

b. correlative: relative deletion is not possible

*[jitni: Mina ~~kʰariːde-giː~~] [Ram utni:=hi: kita:bẽ
HOW.MANY.F Mina.F buy-FUT.3FSG Ram THAT.MANY.F-ONLY books.F
kʰariːde-gaː]
buy-FUT.3MSG

intended: 'Ram will buy as many books as Mina.'

Relative Deletion with equatives (and elsewhere) can leave behind multiple remnants.

(6) multiple remnants

Ram aːj utni:=hi: kita:bẽ kʰariːde-gaː [jitni: Mina
Ram today THAT.MANY.F-ONLY books.F buy-FUT.3MSG HOW.MANY.F Mina.F
kal ~~kʰariːde-giː~~]
tomorrow buy-FUT.3FSG

'Ram will buy as many books today as Mina will tomorrow.'

The elided phrase can include nominals.

(7) [utne=hi: log Ram=ko yeh kita:b dẽge] [jitne
THAT.MANY=ONLY people Ram=DAT this book give.FUT.3MSG HOW.MANY
Mina=ko ~~yeh kita:b dẽge~~]
Mina=DAT this book give.FUT.3MGS

'As many people will give this book to Ram as will to Mina.'

We noted earlier that an internal head is not possible with extraposed relative clauses. This is true for equatives too.

(8) *Ram utni:=hi: kita:bẽ k^hari:de-ga: [jitni: kita:bẽ
Ram THAT.MANY.F-ONLY books.F buy-FUT.3MSG HOW.MANY.F books.F
Mina ~~k^hari:de-gi:~~]
Mina.F buy-FUT.3FSG
intended: 'Ram will buy as many books as Mina.'

But there is an interesting exception. When the internal head is different from the external head, the internal head is possible and in fact obligatory. But the NP head and the relative pronoun appear discontinuously suggesting that the relative phrase moves to a higher location stranding the NP head quite generally, even when the internal and the external head is the same.

(9) sub-equative (different heads)
a. ok with stranding
Ram utni:=hi: kita:bẽ k^hari:de-ga: [jitne Mina
Ram THAT.MANY.F-ONLY books.F buy-FUT.3MSG HOW.MANY.M Mina.F
akhba:r ~~k^hari:de-gi:~~]
newspaper.M buy-FUT.3FSG
intended: 'Ram will buy as many books as Mina newspapers.'
b. bad without stranding
*Ram utni:=hi: kita:bẽ k^hari:de-ga: [jitne
Ram THAT.MANY.F-ONLY books.F buy-FUT.3MSG HOW.MANY.M
akhba:r Mina ~~k^hari:de-gi:~~]
newspaper.M Mina.F buy-FUT.3FSG
intended: 'Ram will buy as many books as Mina newspapers.'

When the heads are the same, however, stranding does not help. The stranding counterpart of (8) is still ungrammatical.

(10) *Ram utni:=hi: kita:bẽ k^hari:de-ga: [jitni: Mina
Ram THAT.MANY.F-ONLY books.F buy-FUT.3MSG HOW.MANY.F Mina.F
kita:bẽ ~~k^hari:de-gi:~~]
books.F buy-FUT.3FSG
intended: 'Ram will buy as many books as Mina.'

This pattern is similar to that found with comparative deletion in English: when the compared element is the same, it is obligatorily deleted in the comparative clause but when it is not, deletion is not required and is in fact not possible (Kennedy 2002; Lechner 2004).

2.2.2 Comparatives

Like equatives in Hindi-Urdu, comparatives can also be realized as a correlative.

(11) [jitni: (kita:bẽ) Mina k^hari:de-gi:] [us=se zya:da: (kita:bẽ)
HOW.MANY.F books.F Mina.F buy-FUT.3FSG that=THAN MORE books.F
Ram k^hari:de-ga:]
Ram.M buy-FUT.3MSG
'Ram will buy more books than Mina.' (literally: [how many books Mina will buy], Ram will buy more than that many books.)

Curiously there seems to be no way to construct a comparative as an extraposed headed relative in Hindi-Urdu.

(12) *[Ram us=se zya:da: kita:bẽ k^hari:de-ga:] [jitni: Mina
Ram.M that=THAN MORE books.F buy-FUT.3MSG HOW.MANY.F Mina.F
k^hari:de-gi:]
buy-FUT.3FSG
intended: 'Ram will buy more books than Mina.'

We don't know why this is the case but the unacceptability of the extraposed headed relative variant raises the expectation that Relative Deletion is not possible with comparatives. This prediction is borne out. Relative Deletion is blocked inside the correlative clause, perhaps because the ellipsis site precedes its antecedent and the extraposed headed relative variant where we might have expected the extraposed variant to be good is bad for independent reasons.

(13) a. correlative: relative deletion is bad.
*[jitni: (kita:bẽ) Mina ~~k^hari:de-gi:~~] [us=se zya:da: (kita:bẽ)
HOW.MANY.F books.F Mina.F buy-FUT.3FSG that=THAN MORE books.F
Ram k^hari:de-ga:]
Ram.M buy-FUT.3MSG
'Ram will buy more books than Mina.' (literally: [how many books Mina will buy], Ram will buy more than that many books.)

b. extraposed headed relative: independently bad

c. *[Ram us=se zya:da: kita:bẽ k^hari:de-ga:] [jitni: Mina
Ram.M that=THAN MORE books.F buy-FUT.3MSG HOW.MANY.F Mina.F
~~k^hari:de-gi:~~]
buy-FUT.3FSG
intended: 'Ram will buy more books than Mina.'

2.2.3 *when* and *where* clauses

When and *where* clauses can be realized both as correlatives and as extraposed headed relatives.

(14) *where* clauses:
a. correlative:

[jahã: Mina ja:-egi:] [Vina bhi: vahĩ: ja:-egi:]
where Mina.F go-FUT.3FSG Vina ALSO there.ONLY go-FUT.3FSG

'Vina will go where Mina goes.'

b. extraposed headed relative:

Vina: vahĩ: ja:-egi: [jahã: Mina ja:-egi:]
Vina.F there.ONLY go-FUT.3FSG where Mina.F go-FUT.3FSG

'Vina will go where Mina goes.'

(15) *when* clauses:

a. correlative:

[jab Mina a:-egi:] [tab(=hi:) Vina a:-egi:]
WHEN Mina.F come-FUT.3FSG THEN=ONLY Vina.F come-FUT.3FSG

'When/in case Mina comes, then Vina will come.'

b. extraposed headed relative:

[Vina tab=hi: a:-egi:] [jab Mina a:-egi:]
Vina.F THEN=ONLY come-FUT.3FSG WHEN Mina.F come-FUT.3FSG

'Vina will only come when/in case Mina comes.' (Vina will come in the very same circumstances in which Mina will come.)

Relative deletion is possible with *where* clauses but only when the *where* clause follows the main clause i.e. in the extraposed headed relative structure but not in the correlative structure.

(16) *where* clauses:

a. correlative: *relative deletion

*[jahã: Mina ~~ja:-egi:~~] [Vina bhi: vahĩ: ja:-egi:]
where Mina.F go-FUT.3FSG Vina ALSO there.ONLY go-FUT.3FSG

'Vina will go where Mina goes.'

b. extraposed headed relative: $\surd$ relative deletion

Vina: vahĩ: ja:-egi: [jahã: Mina ~~ja:-egi:~~]
Vina.F there.ONLY go-FUT.3FSG where Mina.F go-FUT.3FSG

'Vina will go where Mina goes.'

As with individual relatives, equatives and *where* clauses, relative deletion is impossible in correlative *when* structures (17a). However unlike individual relatives, equatives and *where* clauses, where relative deletion is full grammatical in extraposed headed relatives, relative deletion in *when* clauses is somewhat degraded but not ungrammatical (17b). We suspect there is also some speaker variability here with some speakers finding this structure fully ungrammatical and others reporting the in-between status that we are indicating.

(17) a. correlative: relative deletion is not possible

*[jab Mina ~~a:-egi:~~] [tab(=hi:) Vina a:-egi:]
WHEN Mina.F come-FUT.3FSG THEN=ONLY Vina.F come-FUT.3FSG
'When Mina comes, then Vina will come.'

b. extraposed headed relative: √ relative deletion
??[Vina tab=hi: a:-egi:] [jab Mina ~~a:-egi:~~]
Vina.F THEN=ONLY come-FUT.3FSG WHEN Mina.F come-FUT.3FSG
'Vina will only come when/in case Mina comes.' (Vina will come in the very same circumstances in which Mina will come.)

3 Conditions on Wellformedness of Relative Deletion

3.1 Case Marking on the Relative Pronoun and the Remnant

Our initial description of relative deletion notes that a relative pronoun and one or more XPs survive relative deletion. However not every combination of relative pronoun and XP yields an acceptable instance of relative deletion. Some of these restrictions follow from general principle of ellipsis. For example, the case on the relative pronoun needs to be the same as the case on the XP that the relative clause modifies.

(18) Mina=ne us=hi: laṛke=se ba:t ki: [jɪs=se/*jɪs=ko Tina=ne
Mina=ERG that=ONLY boy=INST talk.F do.PFV.F REL=INS/REL=Dat Tina=ERG
~~ba:t ki:~~]
talk.F do.PFV.F
'Mina talked to the same boy as Tina.'

The impossibility of dative on the relative pronoun follows directly from the fact that dative would not be licensed inside the elliptical clause, which we assume is structurally identical to the antecedent clause, modulo the remnant and the relative pronoun.

Does the elliptical clause need to be featurally identical to the antecedent? Consider the variants of (18) in (19), where the subject remnant triggers agreement. As a result the elided verb in (19a) has different features from the verb in the antecedent while elided verb has the same features as the verb in the antecedent in (19b).

(19) a. mismatch: ?
?Mina us=hi: laṛke=se ba:t kar-egi: [jɪs=se Ramesh ~~ba:t~~
Mina.F that=ONLY boy=INST talk.F do-FUT.3FSG REL=INS Ramesh.M talk.F
~~kar-ega:~~]
do-FUT.3MSG
'Mina talked to the same boy as Ramesh did.'

b. no mismatch: √

Mina us=hi: laṛke=se ba:t kar-egi: [jɪs=se Vina ~~ba:t~~
Mina.F that.ONLY boy=INST talk.F do-FUT.3FSG REL=INS Vina.F talk.F
~~kar-egi:~~]
do-FUT.3FSG

'Mina talked to the same boy as Vina did.'

If there was no ellipsis, the verb in the relative clause in (19a) would be *kar-ega:* 'do-Fut.3MSg', distinct from the verb in the main clause *kar-egi:* 'do-Fut.3FSg'. In (19b), the verbs in the relative clause and the main verb would have the same form *kar-egi:* 'do-Fut.3FSg'. The mismatch condition leads to a mild deviance compared to the case where there is feature identity. Moreover in the equative cases, feature mismatch does not produce even mild deviance (5). Since the deviance created by feature mismatch is mild and variable, we will not consider it further but to avoid potential interference from mismatches, we will check for well-formedness of relative deletion in environments where there is no mismatch.

Not all restrictions on wellformed combinations of the relative phrase and other remnant XPs in relative deletion contexts can be derived from ellipsis identity considerations. It seems that in a range of cases where the relative pronoun is the bare *jo* 'REL' or *jab* 'when' and the remnant XPs include a bare subject (i.e. not overtly case-marked), relative deletion is ungrammatical.

(20) a. *jo* + bare subject remnant: *Relative Deletion

*Tina vo=hi: kita:b k^{h}ari:de-gi: [jo=ki Mina ~~kita:b~~
Tina.F that=ONLY book.F buy-FUT.3FSG REL=that Mina.F book.F
~~k^{h}ari:de-gi:~~]
buy-FUT.3FSG

intended: 'I'll buy the same book as Mina.'

b. *jo* + case-marked subject remnant: √Relative Deletion

Tina=ne vo=hi: kita:b k^{h}ari:d-i: [jo=ki Mina=ne ~~kita:b~~
Tina=ERG that=ONLY book.F buy-PFV.MSG REL=that Mina=ERG book.F
~~k^{h}ari:d-i:~~]
buy-PFV.FSG

intend: 'Tina bought the same book as Mina.'

Multiple remnants do not fix the problem created by non-overtly case-marked remnants.

(21) *Ravi Mina=ko vo=hi: kita:b de-ga: [jo=ki Atul Tina=ko
Ravi.M Mina=DAT that=ONLY book.F give-FUT.3MSG REL=that Atul.M Tina-DAT
~~kita:b de-ga:~~]
book give-FUT.3MSG

'Ravi will give the same book to Mina which Atul will to Tina.'

We noted in the previous section that relative deletion is degraded with *when* clauses. A more nuanced picture emerges when we consider remnant XPs with case-marked nominals.

The cases considered so far all involved bare NP remnantas. If the NP remnant is in fact case-marked, things are quite different.

(22) Mina=ko duty=ke liye tab=hi: bula:-ya: ja:-ta:
Mina=DAT duty=GEN.OBL for then=ONLY call-PFV PASS-IMPFV.3MSG
hɛ [jab=ki Tina=ko ~~...~~]
BE.PRS.3SG when=that Tina=DAT
'Mina is only called for duty when Tina is.'

Relative deletion seems to be blocked in cases where both the relative pronoun and the subject remnant are not overtly case marked. We see in (22) that when the subject remnant is overtly case marked, relative deletion is ok. Relative deletion also becomes ok if the relative pronoun is overtly case marked.

(23) case-marked relative pronoun + bare subject

a. inidvidual
Mina us=hi: kita:b ko k^hari:de-gi: [jɪs=ko ki Tina ~~...~~]
Mina.F that=ONLY book DAT buy-FUT.3FSG REL=DAT that Tina.F
'Mina will buy the same book that Tina will.'

b. where
Sheela us=hi: sheher=mẽ maka:n k^hari:de-gi: [jɪs=mẽ Tina ~~...~~]
Sheela.F that=ONLY city=IN house buy-FUT.3FSG REL=IN Tina.F
'Sheila will buy a house in the very same city as Tina.'

(24) Relative Deletion generalization: either the relative pronoun or the subject remnant of the elliptical clause must be overtly case marked.

The situation is reminiscent of a pattern found with mutiple sluicing in English. Sluicing which would involve two DP remnants is ungrammatical while sluicing with one DP and one PP is acceptable.

(25) a. John said that he gave someone something
*but I don't remember who what.

b. John said that he gave something to someone
but I don't remember what to whom.

The relative deletion generalization needs to be qualified when we go beyond individual denoting relative pronouns (*jo* 'REL' and its variants). By the metric of permitting relative deletion with bare DP remnants, the locative relative pronoun *jahã:* 'where' and the degree relative pronoun *jitna:* 'how much', but not the temporal relative pronoun *jab* 'when', count as overtly case marked.

3.2 Location and Form of the Relative Pronoun

The initial position is the default position of a relative pronoun in a headed relative clause in Hindi-Urdu. But relative clause internal material can precede the relative pronoun and such relative clauses are judged as only mildly deviant.

(26) a. relative pronoun is initial in relative clause:
vo laṛki: acchi: hɛ [jo(=ki) Ram=ko pasand hɛ]
DEM girl.F good.F is REL=that Ram=DAT pleasing is
'The girl who Ram likes is good.'

b. relative pronoun is not initial in relative clause:
(?)vo laṛki: acchi: hɛ [Ram=ko jo(=ki) pasand hɛ]
DEM girl.F good.F is Ram=DAT REL=that pleasing is
'The girl who Ram likes is good.'

However in situations where relative deletion takes place, the relative pronoun must be in initial position in the relative clause.

(27) a. Relative Pronoun Initial: √ Relative Deletion
mɛ̃=ne Sita=ko vo=hi: film dɪkʰa-yi: tʰi: [jo=ki Rita=ko
I=ERG Sita=DAT that=ONLY film.F show-PFV.F BE.PST.F REL=that Rita=DAT
~~film dɪkʰa-yi: tʰi:~~]
film.F show-PFV.F BE.PST.F
'I had showed the same film to Sita as to Rita.'

b. Relative Pronoun Non-Initial: *Relative Deletion
*mɛ̃=ne Sita=ko vo=hi: film dɪkʰa-yi: tʰi: [Rita=ko
I=ERG Sita=DAT that=ONLY film.F show-PFV.F BE.PST.F Rita=DAT
jo=ki ~~film dɪkʰa-yi: tʰi:~~]
REL=that film.F show-PFV.F BE.PST.F
intended 'I had showed the same film to Sita as to Rita.'

The non-elliptical version of the ungrammatical (27b) is in fact acceptable, albeit mildly degraded. The contrast between (27b) and (28) is clear.

(28) ?mɛ̃=ne Sita=ko vo=hi film dɪkʰa-yi: tʰi: [Rita=ko jo=ki
I=ERG Sita=DAT that=ONLY film.F show-PFV.F BE.PST.F Rita=DAT REL=that
dɪkʰa-yi: tʰi:]
show-PFV.F BE.PST.F
'I showed the same film to Sita as to Rita.'

We conclude that the relative pronoun must be initial in the relative clause for relative deletion to be possible. Moreover as we saw through the contrast in (9) this initial element cannot be explicitly phrasal i.e. the relative pronoun in the initial position cannot be part of

a phrase. To sum up, the initial element in the relative clause must be a relative pronoun, which is discontinuous from its NP modifier if it has one. The relative pronouns may, however, be case-marked.

4 Sluicing versus Gapping

Mishra (2024) analyzes Relative deletion as an instance of non-wh sluicing. In this section we provide evidence that suggest that this categorization needs to be rethought. We concur with Mishra (2024) that Relative deletion is a form of 'big ellipsis', but we will show that it shares various features with the operation of gapping in Hindi-Urdu. Both sluicing and gapping involve elision but while sluicing typically elides an entire clause except for a wh-phrase (Ross 1969) (see 29) , gapping involves elision of the verb or other elements, leaving behind remnants in a coordinated structure (Johnson 2009) (see 30).

(29) Ram met someone, but I don't know who ~~Ram met [e]~~.

(30) Ram read three books, and Meena ~~read~~ four.

Hindi-Urdu Relative deletion, unlike sluicing, imposes stricter requirements on the placement of the antecedent and ellipsis site. As seen earlier, deletion is only possible when the relative clause follows the main clause, which contains the antecedent of the ellipsis i.e. deletion is only possible 'going forward'. In only allowing for deletion when the antecedent precedes the ellipsis site, Relative Deletion (31a) parts ways with sluicing in Hindi-Urdu (29), which can precede (or follow) its antecedent .

(31) a. muj^{h}e nahĩː pataː kɪs=ne ~~Mahesh=ko~~ ~~maːraː tʰaː~~ par
I.DAT NEG know who.OBL=ERG Mahesh-ACC hit-PFV.M.SG but
kɪsiː=ne Mahesh=ko maːraː tʰaː
someone=ERG Mahesh=ACC hit.PFV.M.SG be
'I don't know who hit Mahesh, but someone hit Mahesh.'

b. kisiː=ne Mahesh=ko maːraː tʰaː par muj^{h}e nahĩː pataː
someone=ERG Mahesh=ACC hit.PFV.M.SG be.PST but I.DAT NEG know
kɪs=ne ~~Mahesh=ko~~ ~~maːraː tʰa~~
who.OBL=ERG Mahesh=ACC hit.PFV.M.SG
'Someone hit Mahesh, but I don't know who.'

Regardless of whether the ellipsis site precedes or follows the antecedent, the sentence remains grammatical. This indicates that sluicing is not constrained by the directionality of ellipsis. However, as previously demonstrated, Relative deletion exhibits a different pattern (32), allowing only forward ellipsis.

(32) Equatives
a. Ungrammatical: *[CP….. < ... >] [IP ……..]

*[Jitni: kita:bẽ Meena=ko ~~kal~~ ~~paRHni:~~ ~~hɛ̃~~] muj^{h}e a:j
How.many.F books Meena=ACC yesterday read.INF.F be.PRS.PL I.DAT today
utni:=hi: paRHni: hɛ̃
that-many=ONLY read.INF.F be.PRS/PL

'I have to read as many books today as Meena has to tomorrow.'

b. Grammatical : [IP ..... [CP..... < ... >]]

muj^{h}e a:j utni:=hi: kita:bẽ paRHni: hɛ̃ [jitni: Meena=ko
I=DAT today that-many EMP books read.INF.F be.PRS.PL how.many.F
~~kal~~ ~~paRHni:~~ ~~hɛ̃~~]
Meena=ACC tomorrow read.INF.F be.PRS/.PL

'I have to read as many books today as Meena has to tomorrow.'

The fact that relative deletion cannot be 'backward' ellipsis means that it cannot apply inside correlative clauses: given the clause-initial position of correlative clauses, an ellipsis site within them would necessarily precede its antecedent. This constraint distinguishes relative deletion from sluicing and instead aligns it more closely with gapping, where similar restrictions are observed (Johnson 2009). Both processes require the ellipsis site to follow the antecedent (as in (32b) and (33)), and violations of this order result in ungrammaticality.

(33) Ram read three books, and Meena ~~read~~ four.

(34) *Ram ~~read~~ three books, and Meena read four.

Another aspect of relative deletion in Hindi-Urdu is that unlike sluicing, it cannot find its antecedent in a different utterance. In fact, even within the same utterance, the ellipsis antecedent in Relative deletion needs to be local to the ellipsis site. To be precise, the ellipsis antecedent must be in the clause to which the relative clause is attached. This can be seen from the ungrammaticality of the non-local ellipsis resolution in (35). Relative deletion is only grammatical if the antecedent is the local 'sell books', and not if it is the non-local 'buy books'.

(35) Mina utni:=hi: kita:bẽ khari:d-egi: jitni: Tina ~~kita:bẽ~~
Mina.NOM that-many=ONLY books buy.FUT.F.SG REL.F Tina.NOM books
~~khari:d-egi:~~ ɔr mɛ̃ utni: kita:bẽ bec-ũga: [jitni: Tina
buy-FUT.F.SG and I.NOM that-many-EMPH books sell-FUT.M.SG REL.F Tina
~~kita:bẽ~~ ~~bec-egi:~~/*~~kita:bẽ~~ ~~khari:d-egi:~~]
books sell-FUT.F.SG/books buy-FUT.F.SG]

'Mina will buy as many books as Tina will buy books and I will sell as many books as Tina sells/*buys.'

This is similar to gapping, where the ellipsis site and antecedent must be part of the same utterance, much like relative deletion, which similarly disallows cross-utterance ellipsis. Consider the following instance of gapping (36).

(36) A: Ram=ne Mina=se ba:t nahĩ: ki:
Ram=ERG Mina=WITH talk.F NEG do.PFV.F
intended: A: 'Ram didn't talk to Mina.'

B: *Mina=ne Vina=se [~~ba:t nahĩ: ki:~~]
Mina=ERG Vina=WITH talk.F NEG do.PFV.F
intended B:'*Mina didn't talk to Vina.'

Sluicing, by contrast, demonstrates far more flexibility. In (37), sluicing is permissible even though the antecedent occurs in a separate utterance.

(37) A: Ram kɪsi=se roz mɪl-ta hE
Ram someone=WITH daily meet-IMPF.M.SG be.PRES.3SG
'Ram meets someone everyday.'

B: muj^{h}e nahĩ: pata: kIs=se
I.DAT NEG know who=with
'I don't know who with.'

This ability leads us to think that sluicing seems to be governed by looser syntactic constraints than both Relative deletion and Gapping since the wh-phrase manages to provide sufficient information to recover the elided material, even when the ellipsis occurs in a different utterance. The evident structural dependency, and possibly a notion of locality, existing between the antecedent and the ellipsis site in Relative deletion in Hindi-Urdu implies an operation potentially akin to English gapping.

Understanding the structural similarities between Gapping and relative deletion becomes easier when we adopt the Conjunction Reduction Hypothesis (CR Hypothesis) posited by Lechner (2004). Such a framework argues for comparative clauses and coordinate structures sharing enough syntactic properties that reduction processes, such as gapping, can apply to both.

(38) This screen is wider than that screen is.
This screen is wider [$_{than-XP}$ than that screen ~~is~~].

The CR hypothesis also accounts for the strict requirements of intra-utterance ellipsis in both processes, as neither gapping nor relative deletion allows for cross-utterance dependencies. It gains further credibility from the observation that, aside from gapping, right node raising (or 'backward gapping') also occurs with right-adjoined equatives. In standard coordinate structures, right node raising refers to the phenomenon where a verb phrase or another element common to both clauses is elided from one of them, as illustrated in (39).

(39) a. Right Node Raising in a Coordinate Structure:

Ram=ne vaki:lõ=ko aur tum=ne doctrõ=ko rishvat de-ni:
Ram=ERG lawyers=DAT and you=ERG doctors=DAT bribe.F give-INF.F
ca:h-i:
want.PFV.F

'Ram wanted to bribe the lawyers and you the doctors.'

In this sentence, the verbal sequence 'want to give a bribe' is shared between the two conjuncts. (40) extends this argument by showing that right node raising is also possible in equative constructions, specifically when the equative clause is right-adjoined.

(40) Right Node Raising in a Right Adjoined Equative:

Ram=ne utne=hi: vakilõ=ko, jitne tum=ne doctrõ=ko,
Ram=ERG that.many=ONLY lawyers=DAT REL.MANY you=ERG doctors=DAT
rishvat de-ni: ca:hi:
bribe.F give-INF.F want.PFV.F

'Ram wanted to bribe as many lawyers as you wanted to bribe doctors.'

Here, the same elision pattern occurs - the verbal sequence 'want to give a bribe' is elided in the second clause, with the equative conjunction introducing an equation between the number of lawyers and the number of doctors. This equivalence suggests that structurally, certain right adjoined equative (relative) clauses are similar to coordinate structures, permitting the same kinds of ellipsis operations, such as right-node raising.

Therefore, the CR hypothesis, when applied to both standard coordinate structures and equative clauses, demonstrates that the mechanism of Relative deletion can be assimilated into a Gapping mechanism under the broader CR framework. But what the exact nature of this gapping mechanism is remains to be seen.

4.1 The Gapping Mechanism

At this point, we have established three key properties of Relative deletion in our analysis:

1. The antecedent must precede the ellipsis.
2. The antecedent must be in the clause to which the relative clause is attached (and hence, there can be no utterance boundary between the antecedent and the ellipsis).
3. Ellipsis is only permissible in restricted syntactic environments (namely coordination-like structures (CR Hypothesis)).

We previously argued that Relative deletion can be conceptually assimilated with Gapping, specifically when the relative or equative clause is treated similarly to a coordinate clause, as observed in English (Lechner 2004). Johnson's (2009) treatment of Gapping, in particular, offers a compelling explanation for the three properties outlined above. In

English, gapped clauses are typically minimal vPs that share a single Tense head with their antecedent clause. This observation is part of the Small Conjunct Analysis of Gapping (see Johnson 2009; Coppock 2001; Lin 2002), where Gapping involves the ellipsis of material from a conjunct, leaving behind a remnant that must be interpreted with reference to the antecedent. Under this analysis, Gapping in English operates via the movement of a VP remnant across both conjuncts, resulting in a shared tense interpretation across the two clauses. Johnson (2009) conceptualizes Gapping as across-the-board movement of a VP remnant and so his approach directly derives the requirement for the antecedent to precede the ellipsis (Property 1), the restriction against an utterance boundary between the antecedent and ellipsis (Property 2), and the limitation of ellipsis to specific environments such as coordination (Property 3). These three properties naturally fall-out from the ATB movement mechanism, which provides the perfect account for English Gapping.

While this would provide a straightforward solution to our puzzle here, we will see that Hindi-Urdu Gapping (HUG) is different from English Gapping, particularly with respect to the size of the conjuncts involved. Unlike English, where gapped clauses are typically vPs, Hindi-Urdu Gapping involves larger or differently structured constituents that are at least clause-sized (Kush 2016). Hindi-Urdu gapping conjuncts are suggested to contain a larger syntactic structure, possibly extending up to the TP or even including some layer of CP, as proposed by the Large Conjunct Analysis of Gapping (Ross 1969; Sag 1976; Jackendoff 1971; Jayaseelan 1990; Lin 2002). We repeat two of Kush's (2016) arguments for the claim, below.

4.1.1 Absence of Wide Scope Readings

A distinctive property of gapping in Hindi-Urdu is the absence of wide-scope readings, as illustrated by (41):

(41) Manu=ko tila:pia: kha:-na: ca:hiye ya: Tanu=ko bi:f.
Manu=DAT tilapia eat-INF.M.SG must or Tanu=DAT beef.
'Manu must eat tilapia or Tanu must eat beef.'
not available: must (manu-eat-tilapia OR tanu-eat-beef)

In this sentence, the wide-scope interpretation, where the obligation applies to the disjunction (i.e., 'Manu must eat tilapia or Tanu must eat beef'), is not available. Instead, only a narrow-scope reading is possible, where the obligation is specific to each individual conjunct (i.e., 'Manu must eat tilapia' or 'Tanu must eat beef'). The missing reading is easily accessible in the English counterpart of (41) supporting a small conjunct analysis for English. The absence of this reading in Hindi-Urdu argues for a big conjunct analysis where the two conjuncts are syntactically independent clauses, each containing its own modal structure rather than sharing a single operator.

4.1.2 Lack of Tense Sharing Across Conjuncts

Gapping in Hindi-Urdu does not allow tense sharing between conjuncts – see (42), where eliding the auxiliary leads to ungrammaticality. If T-sharing across conjuncts was allowed, we would expect this to be an option. The corresponding structure is grammatical in English ('Manu was making chai and Tanu drinking water').

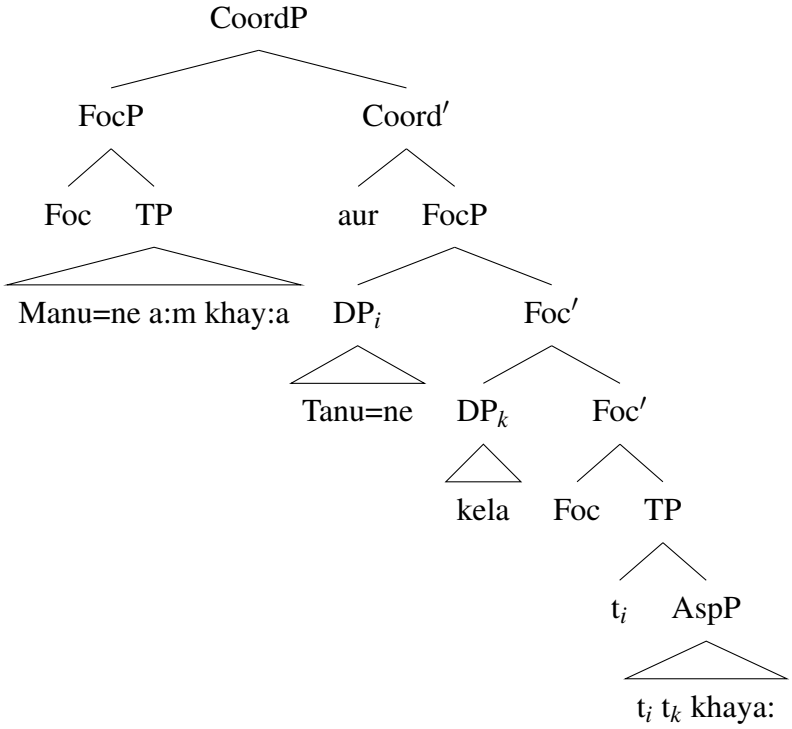
(42) *Manu cai bana:-ta: t^ha: aur Tanu pa:ni: ~~pi:-ta:~~
Manu cai make-IMPF.M.SG aux.PST.M.SG and Tanu water drink-IMPF.M.SG
t^ha:.
aux.PST.M.SG
'Manu was making chai and Tanu was drinking water.'

4.1.3 Kush's Big Ellipsis structure

These arguments lead to the conclusion that each conjunct in Hindi-Urdu gapping is treated as a fully-fledged clause, as opposed to smaller structures in languages like English. Thus, Hindi-Urdu Gapping (43) has the structure in (44), as per Kush (2016).

(43) Manu=ne a:m kha:-ya: aur Tanu=ne kela: ~~kha:-ya:~~
Manu=ERG mango eat-PFV.M.SG and Tanu=ERG banana eat-PFV.M.SG
'Manu ate the mango and Tanu ate the banana.'

(44)

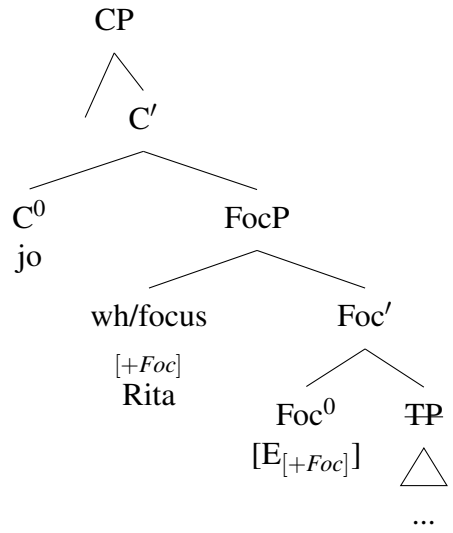


(Kush, 2016, 22)

The aforementioned gapping structure in the language, where conjuncts are clause-sized and exhibit structural independence, bears a striking resemblance to the derivation of Hindi-Urdu Relative deletion as proposed by Mishra (2024) (45), where the focus operator with its E[uFoc*] feature deletes the complement of the Foc head, leaving behind the (relative) complementizer (jo-ki) and the focus element (Rita=ko). A clipped version of the derivation is presented in (46).

(45) mɛ̃=ne Seema=ko vo=hi film dɪk^{h}a-yi t^{h}i [jo=(ki) Rita=ko]
I=ERG Seema=DAT that=ONLY film show-PFV.F be.PST.F REL=that Rita=DAT
'I showed Seema the movie which I showed to Rita.

(46)



(Mishra, 2024, 87)

What we see here is a reinforcement of the argument for the Large Conjunct Analysis in gapping in Hindi-Urdu, as the language's ellipsis operations seem to consistently involve the preservation of higher syntactic structures, whether in gapping or sluicing. But now we find ourselves in a difficult place as the Large Conjunct Analysis of gapping does not derive the three core properties of Relative deletion that we listed earlier— namely, the antecedent precedes the ellipsis, lack of utterance boundary between antecedent and ellipsis and stringent locality effects, and the restricted environments for ellipsis. We will have to leave development of an adequate account for future work.

5 Non-Local Relative deletion

We now turn to two cases of Relative deletion, where the antecedent of the ellipsis is not as local as the cases we have discussed earlier. The first case involves fragment answers and the ellipsis seems to find its antecedent in a prior utterance (47).

(47) A: tum=ne kɪs=ko bula:ya:?
2p=ERG who=ACC invite.PFV.M.SG
A: 'Who did you invite?'

B: Us=hi:=ko jɪs=ko Ramesh=ne ~~bula:ya:~~
Dem=FOC=ACC REL.OBL=ACC Ramesh=ERG invite.PFV.M.SG
intended: 'The person.FOC who Ramesh did.'

However, here it is plausible that there is an elided clause along the lines of 'I invited __' and that the ellipsis nevertheless finds a local antecedent within the elided clause. The second case, (48), highlighted in Mishra (2024), is more challenging. The antecedent of the ellipsis is within the same utterance but it is not local.

(48) mɛ̃ us=ko nahĩ: ja:nti jɪs-ko Rita=ne bula:ya par
I DEM=ACC NEG know.F REL.OBL=ACC Rita=ERG invite.PFV.M.SG but
us=ko zəro:r ja:nti hũ jɪs-ko Rama=ne
DEM=ACC definitely know.F PRES.F REL.OBL=ACC Rama=ERG
'I do not know the person who Rita invited, but I know the person who Rama did.'

We do not understand what distinguishes it from the cases like (35) where the ellipsis antecedent has to be local. It is worth noting, however, that attempts to alter the subject in the second clause - see (49) —result in ungrammaticality.

(49) *mɛ̃ us=ko nahĩ: janti jɪs=ko Ram=ne bula:ya:, par tum
I DEM=ACC NEG know.F REL.OBL=ACC Ram=ERG invite.PFV.M.SG but 2p
us=ko zəro:r ja:nti: ho-gi jɪs=ko Sita=ne
DEM=ACC definitely know.F PRES=FUT REL.OBL=ACC Sita=ERG
'I do not know the person whom Ram invited, but you definitely know the person whom Sita did.'

Along with the proper analysis of Relative Deletion in Hindi-Urdu, we leave the challenge posed by (49) for the future.

6 Conclusion

The current study explores the phenomenon of Relative Deletion (RD) in Hindi-Urdu (HU), focusing on its interaction with various relativization structures. We start with Mishra (2024)'s observation that there are striking similarities with Focus-Based TP/IP Deletion seen in other languages. A detailed examination of Relative deletion reveals distinctive syntactic properties pertaining to the locality of the ellipsis antecedent and the location of the relative pronoun. The antecedent of the ellipsis must precede the ellipsis and must be located in the clause to which the relative clause is attached. These restrictions on the antecedent in particular place Relative deletion closer to gapping than sluicing. Constraining the analysis space is the fact that both gapping and sluicing seem to be instances of 'big' ellipsis in Hindi-Urdu.

7 Acknowledgements

We would like to thank the FASAL 14 audience at Stony Brook University in April 2024. In particular, we would like to thank Ashwini Deo, Parul Upadhyay, Sreekar Raghotham, Thomas Graf, and Troy Messick. Anushree would also like to thank the FASAL 14 committee for their generous travel grant and to Daniel Greeson and Gustavo Freire for their unmatched hospitality and support throughout the trip. Rajesh would like to thank Anikó Lipták, for a very helpful meeting in the Old Botanical Garden in Göttingen and a meeting of the Spring 2024 UMass Syntax Workshop. The authors names are in alphabetical order.

References

Bajaj, Vandana. 2016. *Scaling up Exclusive -hii*. New Brunswick, New Jersey: Rutgers University dissertation.

Coppock, Elizabeth. 2001. Gapping: in Defense of Deletion. In Mary Andronis, Christopher Ball, Heidi Elston & Sylvain Neuvel (eds.), *Papers from the 37th Meeting of the Chicago Linguistic Society*, 133–148. Chicago, IL: Chicago Linguistics Society.

van Craenenbroeck, Jeroen & Anikó Lipták. 2006. The Crosslinguistic Syntax of Sluicing: Evidence from Hungarian Relatives. *Syntax* 9. 248–274.

Jackendoff, Ray. 1971. Gapping and Related Rules. *Linguistic Inquiry* 2.

Jayaseelan, K. A. 1990. Incomplete VP Deletion and Gapping. *Linguistic Analysis* 20. 64–81.

Johnson, Kyle. 2009. Gapping is not (VP-) Ellipsis. *Linguistic Inquiry* 40. 289–328.

Kennedy, Christopher. 2002. Comparative Deletion and Optimality in Syntax. *Natural Language Linguistic Theory* 20. 553–621.

Kush, Dave. 2016. Notes on Gapping in Hindi-Urdu: Conjunct Size and Focus Parallelism. *Linguistic Analysis* 40. 255–296.

Lechner, Winfried. 2004. *Ellipsis in Comparatives*. Berlin: Mouton de Gruyter.

Lin, Vivian. 2002. *Coordination and Sharing at the Interfaces*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

Lobeck, Anne. 1995. Ellipsis: Functional heads, Licensing, and Identification. *Language* 72. 634.

Merchant, Jason. 2001. *The Syntax of Silence: Sluicing, Islands, and the Theory of Ellipsis* Oxford studies in Theoretical Linguistics. Oxford: Oxford University Press.

Mishra, Anushree. 2024. *Sluicing in Hindi-Urdu and Bangla: Syntactic and Theoretical Implications*: The English and Foreign Languages University dissertation. Unpublished Thesis.

Ross, John. 1969. Guess who? In Robert Binnick, Alice Davison, Georgia Green & Jerry Morgan (eds.), *Papers from the 5th Regional Meeting of the Chicago Linguistic Society*, 252–286. Chicago, IL: Chicago Linguistics Society.

Sag, Ivan A. 1976. *Deletion and Logical Form*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

It's about time!: Relating structure, the brain, and comparative syntax

DUSTIN A. CHACÓN, *University of California: Santa Cruz*

ABSTRACT

Studying language in the brain is hard. We've identified a left-lateralized 'language network' that supports language comprehension across languages, individuals, and ages. However, it's proven difficult to relate the parts of this language network to specific representations or computations. Why is it so hard to get better insight into the functions of the pieces of the language network? One reason is that careful, cross-linguistic comparison across languages is still in its infancy in neurolinguistics. Another reason is that our theories of language comprehension are largely informed by results from serial, slow, word-by-word reading tasks. To understand how the brain processes and represents grammatical knowledge, we need to carefully vary and contrast languages and modalities – our theories of language should not be over-fit to one language or one kind of task. Here, I show how different reading paradigms in Bengali (Bangla), Hindi/Urdu, Nepali, and English can refine our understanding of the brain bases of language.

1 Introduction

The theme of this year'*(f)ASAL* conference is locality. 'Locality' in linguistic theory usually refers to the domain of application of grammatical principles, e.g., the relations between an anaphor and its antecedent. One reason why locality is of interest is that languages differ in how locality operates – English anaphors cannot be bound in possessor positions (**Ram read himself's book*), whereas the Hindi/Urdu[1] anaphor *apnā* can, for instance (Cohen 1973). Factoring out the universals and perspicuously describing the remaining variation is a key goal of linguistic theory. In psycholinguistic research, more attention has been paid to the how long the cognitive processes needed to comprehend and produce language take, and how the grammatical principles identified in linguistic research might (or might not) be implemented (e.g., Apurva & Husain 2021 on Hindi verb prediction; Chacón et al. 2016 on Bangla scrambling). Because of this, studies on locality and variation are also important for characterizing the relationship of grammatical structure and the mind and brain.

Of course, 'locality' most literally refers to the position in space. From the perspective of where in the brain language-related computations and representations reside, what can we learn by comparing and contrasting different languages? At larger grain-sizes, activity in the 'language network' – a set of regions in the left temporal and frontal lobes – appear to be uniform across individuals and language groups, even across typologically and historically unrelated languages (Lipkin et al. 2022; Malik-Moraleda et al. 2022). But, what can we say about the differences between languages? Does the uniform language network

[1] I follow standard practice in using 'Hindi/Urdu' when making generalizations that apply to both standardized languages, and 'Hindi' and 'Urdu' in the context of specific experiments where one orthography/standard must be adhered to.

'do' things differently in the brains of users of different languages? At some level of abstraction, the answer to this question must be 'yes', but how? To date, explicit cross-linguistic comparisons have contributed prominently to understanding the neural responses to unexpected vs. expected words in sentence processing (Bornkessel-Schlesewsky et al. 2011, 2019; see also Gulati et al. 2024). Despite this, we are still in the early days of parceling which aspects of brain activity correspond to shared universal features and which correspond to the kinds of variation observed in language descriptions.

This lag can be attributed to a few reasons. The first reason is methodological. It is difficult to know in advance which languages or phenomena to investigate without careful language descriptions. We need linguists and psycholinguists to tell us what languages look like and how they are processed before we can ask about their neural bases. The second reason is practical. The equipment needed for experimental research historically has been expensive and difficult to use and maintain, and therefore not always accessible to scientists or participant populations who speak these languages. The relative time and money to establish a data point in the cognitive neuroscience of language is much greater than traditional methods used in theoretical linguistics and language description, which further contributes to the lag between theoretical and experimental work in syntax. This also results in biases towards languages that are spoken in wealthy and industrialized countries (Anand et al. 2011; Collart 2024), which affects both the data that we can collect and the questions that are considered important. The third reason is conceptual. What is the relationship between the results of cognitive neuroscience experiments and our theoretical primitives? Cognitive neuroscience experiments usually require participants to perform a language processing task, and the results are typically a difference in recorded brain activity in time and space. By contrast, the constructs of our representational theories (Agree, *v*P, [CORONAL]) are amodal, time-independent, abstract. Cognitive neuroscience experiments usually require some kind of language processing task, and so recorded brain responses index the cognitive processes deployed moment-by-moment (prediction, cue-based retrieval, reanalysis), indirectly reflecting the representations that linguists traditionally are interested in. More generally, the expectations that we have for an explanatory theory in cognitive neuroscience are evolving (Poeppel 2012, Embick & Poeppel 2015), especially as our techniques and understanding of the brain become more sophisticated over time.

Here, I hope to make a humble contribution. I will not draw any strong conclusions about any particular brain area or function in this paper, targeted towards an audience with specialization in South Asian languages and theoretical linguistics. Instead, I hope to show that cross-language comparisons *per se*, drawing specifically from South Asian languages, are valuable for understanding the brain bases of language. This point is probably uncontroversial, but in practice it is not obvious how to (systematically) compare and contrast languages. I also hope to demonstrate that rethinking standard methodological assumptions in language processing research – that sentences are necessarily comprehended 'word-by-word' (Snell & Grainger 2017; Wen et al. 2019; Flower & Pylkkänen 2024) – can introduce a new perspective on connecting the brain and language description by manipulating the relation between abstract grammatical structure and temporally-bound psycholinguistic processes. Overall, my argument is that our theories of language in the brain should not be over-fit to any particular language, modality, or task and that a theory of *language* in the

brain must allow for a theory of *languages* in the brain.

2 Morphosyntactic Processing in Bengali in the Temporal and Orbitofrontal Lobes

Before jumping into more complex syntactic phenomena, let us start with something (deceptively) simpler – words. Here, I hope to demonstrate that careful cross-language comparisons of a well-understood brain response can clarify which aspects are universal and which aspects may reflect specific properties of individual languages. In short, recent work on the processing of morphologically complex words in Bengali (Bangla) demonstrates a right-lateralized brain response that largely mirrors a left-lateralized response that otherwise appeared universal (Moitra et al. 2024). This demonstrates that some aspects of the cortical organization and time course of these processes are largely uniform, but which hemisphere supports these computations can vary.

Early brain responses to words show distinct patterns of activation near the relevant sensory cortex, i.e., auditory word recognition shows early patterns of activity near auditory cortex and visual word recognition shows early patterns of activation near visual cortex (Marinkovic et al. 2003). These earlier brain responses to visual stimuli have been of great interest for understanding 'morphological decomposition', or identifying the constituent morphemes of a word, e.g., *kicked* consists of the stem *kick* and the suffix *-ed*. In reading, morphemes are identified on their orthographic form (Taft & Forster 1975; Rastle & Davis 2008 for review). In magnetoencephalography (MEG) recordings, morphologically complex words (*refill*) exhibit different activity than orthographically-similar, monomorphemic controls (*reckon*) ~170ms post-word onset (the M170). This M170 response localizes to the left fusiform gyrus, known as the 'visual word form area' (VWFA), an area showing specialization to written words (Cohen et al. 2000; Tarkiainen et al. 1999; Gwilliams et al. 2016). The amplitude of the M170 response correlates with stem-to-whole word transition probability, i.e., the ratio of the frequency of the whole word (*refill*) and the frequency of the stem in all its uses (*fill*), further demonstrating early morphological decomposition on the basis of the word form (Solomyak & Marantz 2010; Lewis et al. 2011; Wray et al. 2022).

Subsequent to morphological decomposition, there are (at least) three separate identifiable brain responses (Schreuder & Baayen 1995). The first stage is lexeme look-up, in which the properties of the constituent morphemes are accessed (what lexical item does <*fill*> correspond to?). MEG evidence from Greek (Neophytou et al. 2018) and English (Stockall et al. 2019) show that stem frequency correlates with activity in left temporal lobe in grammatical words ~200–300ms post-word onset, consistent with this stage[2]. The two other stages are category licensing, in which the syntactic category of the morphemes are identified (e.g., *fill* is a verb), and composition, in which the interpretation of the entire structure (*refill* means 'to fill something up a second time'). These stages can be identified by exploiting grammatical affixes with category and semantic selectional restrictions to generate non-words, and comparing which neural response shows sensitivity to which selectional violation. For instance, *re-* requires a verb stem (category restriction) that has a

[2] This neural response likely correlates with the N400/M350 (see Lau et al. 2008).

patient/internal argument (semantic restriction). Thus, **rehat* and **relaugh* are category and semantic violations respectively. MEG evidence shows greater activity for word-internal category violations (**rehat*) ~200–300ms post-word onset in left posterior temporal lobe, a candidate brain area for syntactic information (Matar & Marantz 2021; Matchin & Hickok 2022). This is then followed by greater activity ~300–400ms post word-onset for word-internal semantic violations (**relaugh*) in orbitofrontal cortex, an area often implicated in semantic violations (Brenan & Pylkkänen 2008; Pylkkänen et al. 2009a, Pylkkänen et al. 2009b). These results provide strong empirical support for psycholinguistic models that involve these distinct computations following morphological decomposition (Schreuder & Baayen 1995, Gwilliams & Stockall 2022), and align with the posterior-to-anterior flow of information from occipito-temporal to anterior brain areas (Marinkovic et al. 2003). This is summarized in Figure 1.

This model depends on fast mapping between visual orthographic features and abstract morphological structure. But, orthography and morphology vary dramatically across languages. Does this model hold up when investigating languages that use abugidas or abjads instead of alphabets, or morphological operations like infixation (as seen in Tagalog) or root-and-pattern morphology (as seen in Arabic)? Linnaea Stockall's group has sought to 'stress test' this model of (ortho)morphographic processing by contrasting Bengali, Arabic, Tagalog, Serbian, and Slovenian (Stockall 2021; see Wray et al. 2022; Cayado et al. 2024; Moitra et al. 2024a for Tagalog and Bengali findings).

For Bengali, the study needed to have a different design. Unlike English and Greek, Northern Indo-Aryan languages do not feature productive derivational verbal morphology like *re-* with clearly identifiable syntactic/semantic selectional criteria. Instead, Moitra et al. (2024b) extended the basic design to noun morphology. In a corpus study, we observed that the nominal prefixes *prôti-* and *duḥ-* overwhelmingly attach to independent nominal stems (category restriction) that describe processes, events, or otherwise abstract or non-concrete referents (semantic restriction). The prefix *prôti-* typically describes a reversal or mutual action (*hiṃsa* 'violence', *prôti-hiṃsa* 'revenge'; compare English ***counter**-argument*), and *duḥ-* imparts negative affect towards its stem's referent (*ghôṭôna* 'event', *dur-ghôṭôna* 'accident'; compare English ***mis**-fortune*). Thus, we constructed non-words by attaching these prefixes to adjectival stems to generate category violations (**prôti-lômba* PRÔTI-long; **dus-kalo* DUḤ-black), and to concrete noun stems to generate semantic violations (**prôti-rôktô* PRÔTI-blood; **dus-nak* DUḤ-nose).

The Bengali MEG data revealed a familiar pattern as to prior studies. We found the expected M170 response, ~170ms post-word onset, followed by greater activity for the category violations ~200–300ms and greater activity for semantic violations ~300–400ms, corresponding to category licensing and composition respectively. However, the M170 response localized to the right fusiform gyrus instead of the left fusiform gyrus, and the greater activity for category violations localized to posterior portions of the right middle temporal lobe instead of the left posterior temporal lobe. Thus, the pattern observed in the left hemisphere in English and Greek resurfaced in the right hemisphere in Bengali. Finally, in a *post-hoc* exploratory analysis, we found that the greater activity elicited by semantic violations in orbitofrontal cortex likely began earlier than we expected, also in the ~200–300ms time window. These results are summarized in Figure 1.

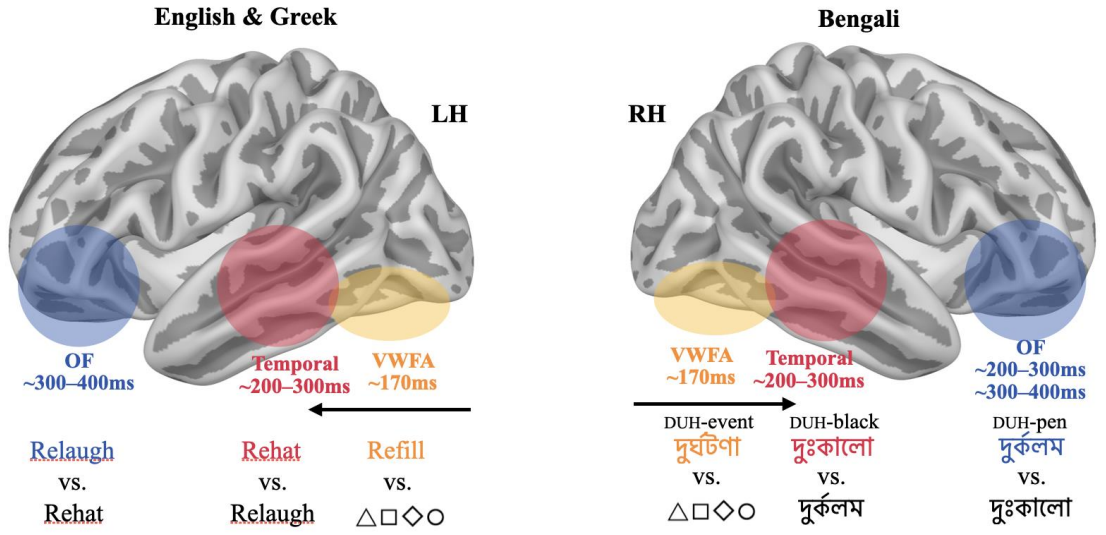


Figure 1. (Left) Schematic of English and Greek results (Neophytou et al. 2018; Stockall et al. 2019). (Right) Schematic of Bengali results (Moitra et al. 2024). Identification of word form begins at 170ms in visual word form area (VWFA), followed by greater activity for category violations in temporal lobe (200–300ms) and semantic violations in orbitofrontal cortex (OF). English and Greek results are plotted on the left hemisphere; Bengali results on the right. Bengali OF activity surfaces at 200–300ms and 300–400ms.

What conclusions can we draw from this? We did not conduct this study testing the hypothesis that these processes would localize in the right hemisphere in Bengali. Although the left hemisphere has largely been the focus of language research, activation in the right temporal lobe is language processing tasks is not unusual (e.g., Kircher et al. 2001, Stowe et al. 2005), including some studies on morphosyntax (Zweig & Pylkkänen 2009). But, this finding raises the question of why previous studies on morphological processing found left temporal activity, whereas ours found right temporal activity. Is this a feature of the denominal morphemes selected in our study, the writing system, or something else that we didn't anticipate? In Moitra et al. (2024a), we found that word length effects in Bengali modulated activity ~130ms in the VWFA in the left hemisphere, so it seems unlikely that Bengali readers' visual word recognition processes are wholly right lateralized. Secondly, our exploratory results suggest a concurrent activation of the category licensing stage with the semantic composition stage, given the patterns of activity identified in right temporal cortex and orbitofrontal cortex ~200–300ms – another surprise. If this coincident right temporal and orbitofrontal cortex activity replicates in other cases, then this places constraints on the architecture of the theory. It cannot be the case that category licensing necessarily *precedes* semantic composition. We now need to explain why some semantic violation

responses occur around the same time as category violation responses, and others are delayed[3]. At this point, no strong inferences can be made about the nature of the hemisphere localization in Bengali vs. English and Greek, nor the different time-courses. However, we can now formulate newer hypotheses that can further refine our understanding of earlier stages of lexical access in the brain.

3 Case/Agreement Hindi/Urdu and Nepali in Left Temporoparietal Juncture

The Bengali morphosyntactic processing data demonstrated the necessity of testing (putatively) universal models in new languages. Similarly, investigations into new languages may reveal questions that are otherwise unaskable in more well-studied languages like English. Moreover, comparisons between similar languages can suggest new hypotheses for the functions of brain regions that may not be obvious otherwise (Chacón et al. 2024a; Khokhar et al. 2024). In this section, I review three experiments on the interaction of split-ergativity and agreement in Hindi/Urdu and Nepali. In two MEG studies contrasting Hindi/Urdu and Nepali, I show that the left temporoparietal junction (LTPJ)[4] may support the processes deployed in processing argument-verb agreement in Hindi/Urdu. More tendentiously, I suggest that this brain area may be activated in these studies because it plays a key role in the (amodal) *representation* of aspect or agreement, not because of the temporal dynamics and memory operations of language processing. This is because we see a similar response in the LTPJ in word-by-word reading tasks, in which participants must process the sentence slowly and incrementally, and in parallel reading tasks, in which participants see the sentence 'at-a-glance', which may favor parallel reading vs. serial word-by-word reading (see Wen et al. 2019, Dunagan et al. 2024; Flower & Pylkkänen 2024).

The necessity of examining familiar processing questions in new languages was recently demonstrated by Bhatia & Dillon (2022) in their investigations into the processing of argument-verb agreement in Hindi. The processing of argument-verb agreement has been a useful window into the kinds of processes that support language comprehension generally. This is usually done by examining cases like (1), which exhibit the 'agreement attraction' phenomenon. Comprehenders rarely notice the ungrammatical plural verb *are*, which should agree with the singular verb *key*, due to interference of the plural NP *cabinets*. This faultiness of agreement has been leveraged as a window into how long-distance dependencies are formed and represented in the mind generally (Eberhard et al. 2005; Wagers et al. 2009; Chacón 2022).

(1) [$_{NP}$ The key$_{[SG]}$ [$_{PP}$ to the cabinets$_{[PL]}$]] { is$_{[SG]}$ / *are$_{[PL]}$ } on the table

Bhatia & Dillon (2022) explore agreement attraction phenomena in Hindi/Urdu, a language

[3] One possibility may be that concrete vs. abstract concepts in general are distinguished earlier in lexical access, potentially during the lexeme lookup stage. Abstract vs. concrete nouns elicit different N400 effects (Kounios & Holcomb 1994), the EEG correlate of the M350. Thus, the concrete / abstract distinction might 'come on-line' earlier than the kind of semantic features in previous studies on verb morphology.

[4] I use the left temporo-parietal junction to refer to the angular gyrus (Brodmann Area 39) and supramarginal gyrus (Brodmann Area 40) and posterior portions of the superior temporal lobe. This overlaps with the traditional 'Wernicke's Area' and the inferior parietal lobule.

in which the verb does not necessarily agree with the subject NP. In Hindi/Urdu, the verb agrees with the highest NP in the structure that does not bear a case suffix, a 'bare' NP (Pandharipande & Kachru 1977). This may be the subject NP, the object NP, or neither. The subject NP may be bare or ergative, because Hindi/Urdu uses an aspect-based split-ergative system. If the verb is perfective, then the subject NP is marked with the ergative suffix *-ne*, and cannot control agreement. The object NP may be bare or not because Hindi/Urdu uses differential object marking. Animate object NPs must always be marked with the dative/accusative suffix *-ko*, and therefore can never control agreement. Inanimate object NPs may surface as bare, or they may take the *-ko* case ending to mark a definite interpretation. If the subject NP is bare (2, 4), then the verb agrees with the subject NP in person, number, and gender, since it is the highest bare argument NP. If the subject NP is ergative and the object NP is bare (3), then the verb agrees with the object NP in person, number, and gender, since the object NP is the highest bare argument. Finally, if both subject and object NP are marked with an overt case suffix, (5), then a default 3rd person singular masculine form surfaces on the verb.

(2) **laṛkā**[3, M, SG] ek kitāb[3, F, SG] paṛhtā[3, M, SG] hai
boy a book read AUX
'A boy reads a book' – subject NP agreement

(3) laṛke-ne[3, M, SG] **ek kitāb**[3, F, SG] paṛhī [3, F, SG] hai
boy-ERG a book read AUX
'A boy read a book' – object NP agreement

(4) **laṛkā**[3, M, SG] ek kitāb-ko[3, F, SG] paṛhtā[3, M, SG] hai
boy a book-DAT read AUX
'A boy read a book' – subject NP agreement

(5) laṛke-ne[3, M, SG] ek kitāb-ko[3, F, SG] paṛhā [3, M, SG] hai
boy-ERG a book read AUX
'A boy read a book' – default agreement

In a series of behavioral studies, Bhatia & Dillon (2022) found that Hindi readers are susceptible to agreement attraction. But, across their studies, they find that only NPs that control agreement of one verb serve as attractors for other verbs. There is no evidence that Hindi readers attempt to retrieve an argument NP by its grammatical function or its morphosyntax. In other words, Hindi users do not systematically seek to relate verbs to subject NPs, object NPs, or even morphologically bare NPs necessarily. Rather, Hindi readers attempt to retrieve an 'agreer' NP, i.e., only the bolded NPs in (2–5) could tamper with the processing of other agreement relations in multiclausal structures, but not the unbolded NPs. This suggests that the grammatical details of the language guides comprehenders to represent a particular NP as relevant for agreement processes, which then guides processing of agreement.

This raises a new question – what does the brain do when it encounters an agreement-controlling NP, like *ek kitāb* 'a book' in (3), and how does this brain response compare to the same NP that does not control agreement, as in (2)? Chacón et al. (2024a) sought

to answer this question using a phrase-by-phrase reading paradigm using MEG[5]. In this study, we used simple, subject-object-verb (SOV) grammatical sentences, like those in (2)–(5). During the processing of the object NPs, we found a pattern of activity in the LTPJ which showed different patterns depending on the case of the subject NP. Specifically, we saw that there was greater activity during the object in the object NP-agreement NP-ERG–NP sequences, (3), vs. the subject NP-agreement NP–NP sequences, (4). This is sketched in Figure 2. We also found spatially and temporally distinct responses to the NP's case assignments, which had main effects in left inferior frontal cortex ('Broca's area') and left anterior temporal lobe. We interpreted this LTPJ response as reflecting an attention control process necessary for identifying the object NP as the agreement controller. In SOV structures, readers must shift from attending to the subject NP and its morphosyntactic features to those of the object NP's. In a language like Hindi with object-agreement structures, this may also require suppressing the number, gender and person features of the subject NP in favor of those of the object NP's. Thus, we interpreted the LTPJ finding as reflecting a necessary shift in which NP's features must be in the focus of attention for the purposes of processing argument-verb relations. This is consistent with other findings which suggest a role of LTPJ in reorienting attention (e.g., Doricchi et al. 2010; Silvetti et al. 2015).

Is this convincing? In the critical comparison, the object NPs are the same words with the same morphology, and they bind the same thematic role. Thus, these features could not be driving the difference in the MEG response. On the other hand, the object NP occurred immediately after an ergative subject NP in one condition, and after a bare subject NP in the other. There could be variables that we failed to control for, such as the frequencies of the noun-case-noun trigrams. Thus, we conducted a (near-)identical study in Nepali, a language with a largely similar aspect-based split-ergative case alignment (Li 2007) and differential object-marking system, but with no object agreement. Unlike Hindi/Urdu, Nepali exhibits subject agreement with person, number and gender regardless of whether the subject NP is bare or marked with the ergative suffix *-le,* (6–7).

(6) **keṭā**[3, M, SG] euṭā kitāb[3, SG] paṛhcha [3, M, SG]
boy a book read.PROG
'A boy reads a book' – subject NP agreement

(7) **keṭā-le**[3, M, SG] euṭā kitāb[3, SG] paṛhyo[3, M, SG]
boy-ERG a book read. PERF
'A boy read a book' – subject NP agreement

In the Nepali MEG study, we replicated the main effects of NP case in left inferior frontal cortex and left anterior temporal lobe, demonstrating processing the (correlates of) case morphology. But, we failed to identify any distinct patterns of case interactions in the LTPJ. Instead, the patterns of activation in the LTPJ during the processing of object NPs were

[5] We were not the first to investigate the brain bases of agreement in split-ergative Indo-Aryan languages. Previous results found P600 responses for unlicensed agreement (Nevins et al. 2007), and distinct N400/P600 complexes for agreement violations versus subject case marking-verb aspect mismatches (Choudhary et al. 2009). See also Sauppe *et al.* (2021) for neural bases of speech planning of split-ergative structures in Hindi.

similar, regardless of whether the previous subject was bare or ergative. Thus, Hindi users' brains respond to NP-ERG–NP sequences differently than NP–NP sequences, even with the same thematic relations and lexical material, whereas Nepali users' brains do not differentiate these sequences in the same way. I take this as more compelling evidence that this response reflects a unique processing adaptation that Hindi comprehenders' brains deploy in mapping agreement relations to argument NP case morphology. The results of the Hindi MEG results could only be suggestive without the contrast with Nepali, and leveraging the similarities (case assignment) and differences (argument-verb agreement).

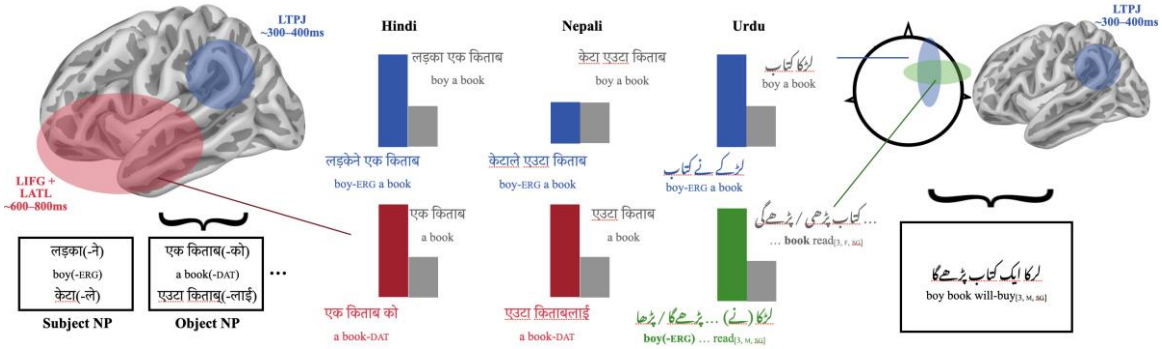


Figure 2. Summarized results from split-ergative agreement studies in Hindi, Nepali and Urdu. (Left) In a word-by-word reading task, responses in the LTPJ show greater activation in Hindi readers for NP-ERG–NP sequences compared to NP–NP sequences, but not in Nepali readers, whereas left inferior frontal gyrus and left anterior temporal lobe (LIFG+LATL) activity showed similar increased activity for case marking across structure types and languages. (Right) In a parallel reading task, Urdu readers showed greater activation in the right midline+anterior sensors and in the LTPJ for ergative subject NPs/perfective verbs compared to bare subject NP/imperfective verb sentences, and different activation for right parietal/lateral sensors for subject NP-verb agreement sentences compared to object NP-verb agreement sentences.

I could conclude the story here: The Hindi-Nepali comparison demonstrates the utility of comparing and contrasting languages to fine-tune the understanding of neural responses to case and agreement in the LTPJ, the left inferior frontal cortex, and the left anterior temporal lobe. But, are there any compelling alternatives? In these studies, we followed standard methodological practice with reading studies in sentence processing research. The interpretation we assigned to these results crucially depended on the attention and memory processes that Hindi and Nepali readers must undertake for the task, in which each phrase appears independently and sequentially. Our methods and theory both favor a view in which there are discrete processing stages at each word/phrase, i.e., something 'happens' at the subject NP, then at the object NP. But, this is not standard practice in morphological processing experiments. Participants are not asked to read morphemes one-by-one. Instead, experimenters allow participants' minds (and brains) to process structurally complex words 'all at once' by revealing the entire word, and permitting participants

to process the stimuli on their own accord. What could be learned by adopting this methodology and conducting studies in which participants read entire (short) sentences, displayed all-at-once?

With Liina Pylkkänen's lab, we have started systematically exploring what computations and representations are involved in this kind of reading (Pylkkänen & Chacón 2024; Fallon & Pylkkänen 2024). Previous results show distinct neural responses in both EEG and MEG recordings for grammatical sentences (*the man can run*) vs. scrambled non-sentences (*can man run the*), displayed at-once for 200ms (Wen et al. 2019; Flower & Pylkkänen 2024; Dufau et al. 2024). These findings may demonstrate some degree of rapid and parallel processing of grammatical information, and a useful new tool for investigating the brain's language network divorced from the methodological assumptions of careful word-by-word reading. In both EEG and MEG recordings in English, we replicated the distinction between sentences and scrambled non-sentences. However, in two independent EEG and MEG studies, we failed to find sensitivity to argument-verb agreement (*the man runs* vs. *the man run*) (Fallon & Pylkkänen 2024; Dunagan et al. 2024).

Do these findings reveal something crucial about the processing of sentences read 'at-a-glance'? Or, do they just show that an English *-s*, placed on a subject NP or a verb, can be easily missed in complex stimuli displayed for 200ms? Follow-up research in Urdu suggests the former (Khokhar et al. 2024). In a 'high-density' (HD-) EEG study, we contrasted grammatical subject NP-agreement and object NP-agreement sentences like (2) and (3) with ungrammatical counterparts, in which the verb was marked with the grammatically-unlicensed gender marking, (8–9). Urdu readers' neural responses diverged for bare subject/imperfective sentence structures (2, 8) and ergative subject/perfective sentence structures (3, 9) around 300ms, in right parietal/lateral sensors. Urdu readers' neural responses also diverged for sentence structures in which the verb agreed with the subject NP (2, 9) and for structures in which the verb agreed with the object NP (3, 8), also around 300ms, in midline/right anterior sensors. Crucially, this demonstrates that Urdu readers' brain responses are sensitive to the two 'ingredients' of agreement in Urdu – which argument NP the verb shares features with, and which aspect the verb carries (and subject NP case assignment). However, we failed to find evidence that Urdu readers distinguish whether the possible combinations of case morphology, verb aspect, and verb feature specifications are *grammatical* given the context of the entire sentence. In other words, Urdu readers 'noticed' the relevant morphosyntactic properties implicated in argument-verb relations, but these neural responses did not 'notice' whether the agreement relation is licensed by the grammar.

(8) ***laṛkā**$_{[3, M, SG]}$ ek kitāb$_{[3, F, SG]}$ paṛhtī $_{[3, F, SG]}$ hai
boy a book read AUX
'A boy reads a book' – subject NP agreement

(9) *laṛke-ne$_{[3, M, SG]}$ **ek kitāb**$_{[3, F, SG]}$ paṛhā $_{[3, M, SG]}$ hai
boy-ERG a book read AUX
'A boy read a book' – object NP agreement

How does this relate to our previous MEG findings in Hindi and Nepali? In an exploratory source reconstruction analysis with the Urdu study[6], we found that the neural response to verb aspect/subject NP case localized to the LTPJ. In other words, the same brain region showing distinct patterns of activity for subject- and object-agreement in our previous Hindi MEG study was also sensitive to subject NP case/verb aspect alignment in our Urdu EEG study. This provides more support for the view that the LTPJ is relevant for these processes. However, our Urdu readers did not seem to care about the well-formedness of the agreement relation, just as we found in English. Furthermore, the proposed explanation that we provided for the Hindi vs. Nepali contrast may not apply in the Urdu study. In Chacón et al. (2024a), our explanation assumed distinct stages of processing at the subject NP and the object NP necessitated the serial presentation paradigm. But, it is not clear that our Urdu participants needed to attend to the subject NP *and then* the to object NP in the parallel presentation study.

So, what function could the LTPJ serve in the brains of our Hindi and Urdu participants? In Khokhar et al. (2024), we suggested that LTPJ may instead serve a key *representational* function that is a precursor to evaluating agreement in Hindi/Urdu split-ergative structures, not necessarily a *processing* function as we suggested in Chacón et al. (2024a). The LTPJ supports processing and representing events and relations, although its precise function is still controversial (Bedny et al. 2013, Meltzer-Asscher et al. 2013; Williams et al. 2017; Matchin et al. 2019). Connecting this to the Hindi/Urdu and Nepali findings across the three experiments is similarly still murky. It may be tempting to suggest that the LTPJ activity we reported in Hindi and Urdu readers reflected construction of perfective vs. imperfective event representations, with no necessary connection to agreement processing. However, this approach is unlikely to succeed as well, given the insensitivity of LTPJ in otherwise identical structures in Nepali readers' brains. How to best theorize the role of the LTPJ and its relation to attention, event/argument structure interpretation, and argument-verb agreement is still ongoing. However, these three findings can place strong constraints on what kinds of theories are viable, and may also suggest that the LTPJ serves a function linking between grammatical structure and interpretation (e.g., Meltzer-Asscher et al. 2013).

4 Conclusion

It is a encouraging that we've identified a uniform left-lateralized language network that supports language. This also aligns with the theoretical linguistic goal of identifying universal linguistic representations and computations, and building explanatory models of language in the mind and brain. But, we must also characterize how differences in grammatical structure across languages mold moment-by-moment processing dynamics, and how these differences correspond to and are reflected in the language network. This is hard both for theoretical and logistical reasons, but I believe it's a necessary step and offers

[6] EEG data affords less spatial resolution than MEG data, and for this reason EEG data are usually only presented in 'sensor-space'. However, higher sensor density and more sophisticated analysis pipelines can provide source localization results similar to MEG; although this is still not standard practice, and there are still limitations of source localization with EEG (see Asadzadeh et al. 2020 for overview)

many exciting opportunities for collaborations between theoretical linguistics and psycho/neurolinguists. Here, I showed two cases in which careful comparison between languages, informed by linguistic theory, language descriptions, and sophisticated psycholinguistic models, can guide and refine our understanding of the neural bases of language.

In this paper, I focused on South Asian languages, for the obvious reason that it is the theme of this conference and my personal interest. However, much of the theoretical and descriptive work in South Asian languages has emphasized comparison. The findings I sketch here raise many challenging questions about other kinds of morphological and syntactic phenomena that are similar to, but not identical to, the Bengali, Hindi/Urdu, and Nepali ones described here. For instance, many questions remain about how the mind and brain processes and represents split-ergative agreement patterns in Gujarati, Punjabi, or Kashmiri, all of which have comparable patterns to Hindi/Urdu and Nepali, but with notable variations that challenge the basic models we proposed here.

Acknowledgements

This research was supported in part by ESRC grant ES/V000012/1to Linnaea Stockall, NSF grant 2335767 to Liina Pylkkänen and Dustin A. Chacón, NSF Grant 1903783 to John Hale, NYUAD grant G1001 to Liina Pylkkänen and Alec Marantz, and a UGA Office of Research Seed Grant to Dustin A. Chacón. I would like to thank the co-authors of the papers reported here (Diogo Almeida, Rajesh Bhatt, Brian Dillon, Donald Dunagan, John Hale, Zahin Hoque, Tyson Jordan, Hareem Khokhar, Alec Marantz, Liina Pylkkänen, Jill McLendon, Subhekshya Shrestha). I would also like to thank Dave Kenneth Tayao Cayado, Nigel Flower, Christina Manouilidou, Bojana Ristić, Samantha Wray, and the audiences of *Society for Neurobiology of Language*, *Human Sentence Processing*, and *Architectures and Mechanisms in Language Processing*, and the UCSC Linguistics colloquium for discussion and comments at various stages of these projects.

References

Anand, Pranav, Sandra Chung and Matt Wagers. 2011. *Widening the net: Challenges for gathering linguistic data in the digital age*. White paper published in NSF project SBE 2020: Future research in the Social, Behavioral and Economic Sciences.

Apurva and Samar Husain. 2021. Revisiting anti-locality effects: Evidence against prediction-based accounts. *Journal of Memory and Language* 121, 104280. https://doi.org/10.1016/j.jml.2021.104280

Bedny, Marina, Swethasri Davida, and Rebecca Saxe. 2013. Shindigs, brunches, and rodeos: The neural basis of event words. *Cognitive, Affective, and Behavioral Neuroscience* 14, 891–901. https://doi.org/10.3758/s13415-013-0217-z

Bhatia, Sakshi and Brian W. Dillon. 2022. Processing agreement in Hindi: When agreement feeds attraction. *Journal of Memory and Language* 125, 104322. https://doi.org/10.1016/j.jml.2022.104322

Bornkessel-Schlesewsky, Ina, Franziska Kretzschmar, Sarah Tune, Luming Wang, Safiye Genç, Markus Philipp, Dietmar Rohm, and Matthias Schlesewsky. 2011. Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain & Language* 117(2011), 133–152. https://doi.org/10.1016/j.bandl.2010.09.010

Bornkessel-Schlesewsky, Ina and Matthias Schlesewsky. 2019. Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology* 10, 298. https://doi.org/10.3389/fpsyg.2019.00298

Asadzadeh, Shiva, Tohid Yousefi Rezali, Soosan Beheshti, Azra Delpak, and Saeed Meshgini. 2020. A systematic review of EEG source localization techniques and their applications on diagnosis of brain abnormalities. *Journal of Neuroscience Methods* 339, 108740. https://doi.org/10.1016/j.jneumeth.2020.108740

Brennan, Jonathan and Liina Pylkkänen. 2008. Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Language* 106(2), 132–143. https://doi.org/10.1016/j.bandl.2008.04.003

Cayado, Dave Kenneth Tayao, Samantha Wray, Dustin Alfonso Chacón, Marco Chia-Ho Lai, Suhail Matar and Linnaea Stockall. 2024. MEG evidence for left temporal and orbitofrontal involvement in breaking down inflected words and putting the pieces back together. Accepted for publication in *Cortex*.

Chacón, Dustin A. 2022. Default is different: Relations and representations in agreement processing. *Language, Cognition and Neuroscience* 37(6), 785–804. https://doi.org/10.1080/23273798.2021.2022172

Chacón, Dustin A., Subhekshya Shrestha, Brian W. Dillon, Rajesh Bhatt, Diogo Almeida, Alec Marantz. 2024a. Same Sentences, Different Grammars, Different Brain Responses: An MEG study on Case and Agreement Encoding in Hindi and Nepali Split-Ergative Structures. Under review.

Chacón, Dustin A., Donald Dunagan, Jill McLendon, Hareem Khokhar, and Zahin Hoque. 2024b. Quick, don't move!: Wh-movement and wh-in-situ structures in rapid parallel reading – EEG studies in English, Urdu, and Mandarin Chinese. *bioRXiv*. https://doi.org/10.1101/2024.05.06.592830

Chacón, Dustin A., Mashrur Imtiaz, Shirsho Dasgupta, Sikder M. Murshed, Mina Dan, and Colin Phillips. 2016. Locality and word order in active dependency formation in Bangla. *Frontiers in Psychology* 7, 1235. https://10.3389/fpsyg.2016.01235

Choudhary, Kamal Kumar, Matthias Schlesewsky, Dietmar Rohm, and Ina Bornkessel-Schlesewsky. 2009. The N400 as a correlate of interpretatively relevant linguistic rules: Evidence from Hindi. *Neuropsychologia* 47(13), 3012–3022. https://doi.org/10.1016/j.neuropsychologia.2009.05.009

Cohen, David. 1973. Hindi *apnaa*: A problem in reference assignment. *Foundation of Language* 10, 399–408. https://www.jstor.org/stable/25000727

Cohen, Laurent, Stanislas Dehaene, Lionel Naccache, Stéphane Lehéricy, Ghislaine Dehaene-Lamertz, Marie-Anne Hénaff and François Michel. 2000. The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123, 291–307. https://doi.org/10.1093/brain/123.2.291

Collart, Aymeric. 2024. A decade of language processing research: Which place for linguistic diversity? *Glossa: Psycholinguistics* 3(1), 1–37. https://doi.org/10.5070/G60111432

Doricchi, Fabrizio, Enrica Macci, Massimo Silvetti, and Emiliano Macaluso. 2009. Neural correlates of the spatial and expectancy components of endogenous and stimulus-driven orienting of attention in the Posner task. *Cerebral Cortex* 20(7), 1574–1585. https://doi.org/10.1093/cercor/bhp215

Dufau, Stéphane, Jeremy Yeaton, Jean-Michel Badier, Sophie Chen, Phillip J. Holcomb, Jonathan Grainger. 2024. Sentence superiority in the reading brain. *Neuropsychologia* 198, 10885. https://doi.org/10.1016/j.neuropsychologia.2024.108885

Dunagan, Donald G., Tyson Jordan, John T. Hale, Liina Pylkkänen, and Dustin A. Chacón. 2024. Rapid visual form-based processing of (some) grammatical features in parallel reading: An EEG study in English. *bioRXiv.* https://doi.org/10.1101/2024.04.10.588861

Eberhard, Kathleen M., J. Cooper Cutting, and Kathryn Bock. 2005. Making syntax of sense: Number agreement in sentence production. *Psychological Review* 112(3), 531–559. https://doi.org/https://doi.org/10.1037/0033-295X.112.3.531

Embick, David and David Poeppel. 2015. Towards a computational(ist) neurobiology of language: Correlational, integrated, and explanatory neurolinguitics. *Linguistics, Cognition, and Neuroscience* 30(4), 357–366. https://doi.org/10.1080/23273798.2014.980750

Fallon, Jacqui and Liina Pylkkänen. 2024. Language at a glance: How our brains grasp linguistic structure from parallel visual input. Submitted manuscript, New York University.

Flower, Nigel and Liina Pylkkänen. The spatiotemporal dynamics of bottom-up and top-down processing during at-a-glance reading. To appear in *Journal of Neuroscience.*

Gulati, Mahima, R. Muralikrishnan, and Kamal Kumar Choudhary. 2024. An ERP study on the processing of subject-verb and object-verb gender agreement in Punjabi. *Journal of Psycholinguistic Research* 53, 59. https://doi.org/10.1007/s10936-024-10095-4

Gwilliams, L., G.A. Lewis and A. Marantz. 2016. Functional characterisation of letter-specific responses in time, space and current polarity using magnetoencephalography. *NeuroImage* 132, 320–333. https://doi.org/10.1016/j.neuroimage.2016.02.057

Gwilliams, Laura and Linnaea Stockall. 2022. Distributed Morphology and Neurolinguistics. To appear in *The Cambridge Handbook of Distributed Morphology*. Available on http://lingbuzz.net/lingbuzz/006928

Khokhar, Hareem, Zahin Hoque and Dustin A. Chacón. 2024. Still not in agreement?: A rapid parallel reading EEG study in Urdu. In preparation.

Kircher, Tilo T.J., Mick Brammer, Nuria Tous Andreu, Steven C.R. Williams and Philip K. McGuire. 2001. Engagement of right temporal cortex during processing of linguistic context. *Neuropsychologia* 39(8), 798–809. https://doi.org/10.1016/S0028-3932(01)00014-8

Kounios, John and Phillip J. Holcomb. 1994. Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology* 20(4), 804–823. https://doi.org/10.1037/0278-7393.20.4.804

Krogh, Simone and Liina Pylkkänen. 2024. Manipulating syntax without taxing working memory: MEG correlates of syntactic dependencies in a Verb-Second language. *bioRXiv*. https://doi.org/10.1101/2024.02.20.581245

Lau, Ellen, Colin Phillips, and David Poeppel. 2008. A cortical network for semantics: (de)constructing the N400. *Nature Reviews: Neuroscience* 9, 920–933. https://doi.org/10.1038/nrn2532

Lewis, Gwyneth, Olla Solomyak Olla and Alec Marantz. 2011. The neural basis of obligatory decomposition of suffixed words. *Brain and Language* 118(3), 118–127. https://doi.org/10.1016/j.bandl.2011.04.004

Li, Chao. 2007. Split ergativity and split intransitivity in Nepali. *Lingua* 117(8), 1462–1482. https://doi.org/10.1016/j.lingua.2006.09.002

Lipkin, Benjamin, Greta Tuckute, Josef Affourtit, Hannah Small, Zachary Mineroff, Hope Kean, Olessia Jouravlev, Lara Rakocevic, Brianna Pritchett, Matthew Siegelman, Caitlyn Hoeflin, Alvincé Pongos, Idan A. Blank, Melissa Kline Struhl, Anna Ivanova, Steven Shannon, Aalok Sathe, Malte Hoffman, Alfonso Nieto-Castañón, and Evelina Fedorenko. 2022. Probabilistic atlas for the language network based on precision fMRI data from >800 individuals. *Scientific Data* 9, 529. https://doi.org/10.1038/s41597-022-01645-3

Malik-Moraleda, Saima, Dima Ayyash, Jeanne Gallée, Josef Affourtit, Malte Hoffmann, Zachary Mineroff, Olessia Jouravlev, and Evelina Fedorenko. 2022. An investigation across 45 languages and 12 language families reveals a universal language network. *Nature: Neuroscience* 25, 1014–1019. https://doi.org/10.1038/s41593-022-01114-5

Marinkovic, Ksenija, Rupali D. Dhond, Anders M. Dale, Maureen Glessner, Valerie Carr and Eric Halgren. 2003. Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* 38(3), 487–497. https://doi.org/10.1016/S0896-6273(03)00197-1

Matar, Suhail, Julien Dirani, and Alec Marantz. 2021. Left posterior temporal cortex is sensitive to syntax within conceptually matched Arabic expressions. Nature: Scientific R ports 11, 7181. https://doi.org/10.1038/s41598-021-86474-x

Matchin, William and Gregory Hickok. The cortical organization of syntax. *Cerebral Cortex* 30, 1481–1498. https://doi.org/10.1093/cercor/bhz180

Matchin, William, Chia-Hsuan Liao, Phoebe Gaston, and Ellen Lau. 2019. Same words, different structures: An fMRI investigation of argument relations and the angular gyrus. *Neuropsychologia* 125(2019), 116–128. https://doi.org/10.1016/j.neuropsychologia.2019.01.019

Meltzer-Asscher, Aya, Julia Schuchard, Dirk-Bart den Ouden and Cynthia K. Thompson. 2012. The neural substrates of complex argument structure representations: Processing 'alternating transitivity' verbs. *Language and Cognitive Processes* 28(3), 1154–1168. https://doi.org/10.1080/01690965.2012.672754

Moitra, Swarnendu, Dustin A. Chacón, and Linnaea Stockall. 2024a. How long is long?: Word length effects in reading correspond to minimal graphemic units: An MEG study in Bangla. *PLoS One* 19(4), e0292979. https://doi.org/10.1371/journal.pone.0292979

Moitra, Swarnendu, Dustin A. Chacón, and Linnaea Stokcall. 2024b, ms. Beyond the left hemisphere: MEG evidence for right temporal lobe recruitment in Bangla morphosyntax processing. Queen Mary University of London.

Neophytou, Kyriaki, Christina Manouilidou, Linnaea Stockall, and Alec Marantz. 2018. Syntactic and semantic restrictions on morphological recomposition: MEG evidence from Greek. *Brain & Language* 183, 11–20. https://doi.org/10.1016/j.bandl.2018.05.003

Nevins, Andrew, Brian Dillon, Shiti Malhotra and Colin Phillips. 2007. The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research* 1164, 81–94. https://doi.org/j.brainres.2007.05.058

Pandharipande, Rajeshwari and Yamuna Kachru. 1977. Relational grammar, ergativity, and Hindi-Urdu. *Linua* 41(3–4), 217–238. https://doi.org/10.1016/0024-3841(77)90080-8

Poeppel, David. 2012. The *maps problem* and the *mapping problem*: Two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology* 29(1–2), 34–55. https://doi.org/10.1080/02643294.2012.710600

Pylkkänen, Liina, Dustin A. Chacón. 2024. *Examining syntactic and semantic computations when no order is imposed from the input*. NSF grant 2335767 awarded to New York University. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2335767

Pylkkänen, Liina, Andrea E. Martin, Brian McElree, and Andrew Smart. 2009. The anterior midline field: Coercion or decision making?. *Brain and Language* 108(3), 184-190. https://doi.org/10.1016/j.bandl.2008.06.006

Pylkkänen, Liina, Bridget Oliveri, B. and Ander Smart 2009. Semantics vs. World knowledge in prefrontal cortex. *Language and Cognitive Processes* 24, 1313–1334. https://doi.org/10.1080/01690960903120176

Rastle, Kathleen and Matthew H. Davis. 2008. Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes* 23(7–8), 942–971. https://doi.org/10.1080/01690960802069730

Sauppe, Sebastian, Kamal K. Choudhary, Nathalie Giroud, Damián E. Blasi, Elisabeth Norcliffe, Shikha Bhattamishra, Mahima Gulati, Aitor Egurtzegi, Ina Bornkessel-Schlesewsky, Martin Meyer, Balthasar Bickel. 2021. Neural signatures of syntactic variation in speech planning. *PLoS Biology* 19(1), e3001038. https://doi.org/10.1371/journal.pbio.3001038

Schreuder, Robert and R. Harald Baayen. 1995. Modeling morphology processing. In L.B. Feldman, ed., *Morphological Aspects of Language Processing*, 131–154. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Silvetti, Massimo, Stefano Lasaponara, Francesca Lecce, Alessio Dragone, Emiliano Macaluso, and Fabrizio Doricchi. 2016. The response of the left ventral attentional system to invalid targets and implication for the spatial neglect syndrome: A multivariate fMRI investigation. *Cerebral Cortex* 26(12), 4551–4562. https://doi.org/10.1093/cercor/bhv208

Snell, Joshua and Jonathan Grainger. 2017. The sentence superiority effect revisited. *Cognition* 168, 217–221. https://doi.org/10.1016/j.cognition.2017.07.003

Solomyak, Olla and Alec Marantz. 2010. MEG evidence for early morphological decomposition in visual word recognition: A single-trial correlational MEG study. *Journal of Cognitive Neuroscience* 22(9), 2042–2057. https://doi.org/10.1162/jocn.2009.21296

Stockall, Linnaea. 2021. *Systematicity and Variation in Word Structure Processing Across Languages: A Neuro-Typology Approach*. ESRC grant ES/V000012/1 awarded to Queen Mary University of London. https://gtr.ukri.org/projects?ref=ES%2FV000012%2F1

Stockall, Linnaea, Christina Manouilidou, Laura Gwilliams, Kyriaki Neophytou and Alec Marantz. 2019. Prefix stripping re-re-revisited: MEG investigations of morphological decomposition and recomposition. *Frontiers in Psychology* 10, 1964. https://doi.org/10.3389/fpsyg.2019.01964

Stowe, Laurie, Marco Haverkort and Frans Zwarts. 2005. Rethinking the neurological basis of language. *Lingua* 115(7), 997–1042. https://doi.org/10.1016/j.lingua.2004.01.013

Taft, Marcus and Kenneth I. Forster. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14(6), 638–647. https://doi.org/10.1016/S0022-5371(75)80051-X

Tarkiainen, A., P. Helenius, P.C. Hansen, P.L. Cornellisen and R. Salmelin. 1999. Dynamics of letter string perception in the human occipitotemporal cortex. *Brain* 122, 2119–2131.

Wagers, Matthew W., Ellen Lau, and Colin Phillips. 2009. Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61(2), 206–237. https://doi.org/10.1016/j.jml.2009.04.002

Wen, Yun, Joshua Snell, and Jonathan Grainger. 2019. Parallel, cascaded, interactive processing of words during sentence reading. *Cognition* 189, 221–226. https://doi.org/10.1016/j.cognition.2019.04.013

Williams, Adina, Samir Reddigari and Liina Pylkkänen. 2017. Early sensitivity to left perisylvian cortex to relationality in nouns and verbs. *Neuropsychologia* 100, 131–143. https://doi.org/10.1016/j.neuropsychologia.2017.04.029

Wray, Samantha, Linnaea Stockall, and Alec Marantz. 2022. Early form-based morphological decomposition in Tagalog: MEG evidence from reduplication, infixation, and circumfixation. *Neurobiology of Language* 3(2), 235–255. https://doi.org/10.1162/nol_a_00062

Verbalization as Re-categorization of Lexical Categories in Santali

BISWANATH DASH, *Indian Institute of Technology Delhi*

ABSTRACT

A well-known lexicon-syntax debate in the generative tradition concerns whether word formation occurs in the lexicon or in syntax (Bruening, 2018; Embick & Noyer, 2007). This paper builds on the idea of word formation/ categorization as a syntactic process, focusing on verbalization. In the literature that takes categorization as a syntactic process, verbalization is considered either idiosyncratic or compositional (Arad, 2003). Typological literature (Rijkhoff & van Lier, 2013; Peterson, 2011, 2010; Rau, 2013) indicates that Austro-Asiatic (AA) languages such as Santali and Kharia possess flexible verbal categorization, where a root x can behave like both a noun (N) and a verb (V), defying the N-V distinction that is found in most languages. However, I show empirical support from Santali, an AA language spoken in the Indian states of Odisha and Jharkhand, to argue that verbal categorization is a compositional syntactic process in Santali, where any root must go through a categorization process forming an N or adjective (A) before getting verbalized.

Santali displays high semantic transparency in verbalization, where the verbalized items have a predictable meaning of an N or A. This paper analyzes Santali fluid verbalization and compares it with the kinds of verbalization seen in English. It also questions how re-categorization (verbalization of lexical categories, not roots) incurs a predictable meaning in the verbalized structures and which head of the structure takes care of the semantic transparency or compositionality in Santali.

1 Introduction

Categorization is the most rudimentary trait of human cognition (Harnad, 2017). The paper focuses on lexical categories, which are, according to Baker (2003), the fundamental concepts humans learn, providing special emphasis on verbalization [1] in Santali.

The distinction among lexical categories (LC), more specifically, between N and V, is considered to be the most robust categorial distinction. If any language has a distinction among its lexical categories at all, it is between N and V (Sapir et al., 1944; Whaley, 1996, p. 32; Evans, 2000, p. 103; Croft, 2002, p. 183). The universality of the distinction between N and V is attested in both typological and generative literature. Also, Baker (2003) considers A to be one of the universally present LCs.

A conflicting claim that has long been reported in the literature states that some languages do not have dedicated categories for basic communicative functions like reference,

[1] Verbalization refers to changing any non-verbal category, like N or A, to a V.

predication, and modification. These languages, instead, have fluid word classes, the members of which can carry out more than one of these communicative functions (Rijkhoff & van Lier, 2013). Table 1 presents a list of languages with fluid categorial distribution. These are few of the languages that are reported as the ones without a clear distinction among the LCs.

Language Family	**Languages**	**Reference**
Malayo-Polynesian	Tongan	Churchward, 1953; Broschart, 1997
	Samoan	Churchward, 1951
	Tagalog	Foley, 1998
	Riau Indonesian	Gil, 2013
	Sri Lanka Malay	Nordhoff, 2012
Wakashan family	Nootka	Hockett, 1958; Mithun, 1999
Turkic languages Australian languages Salishan languages		Van Lier & Rijkhoff, 2013
Austroasiatic languages	Santali	McPhail, 1953; Rau, 2013
	Kharia	Peterson, 2011, 2013

Table 1: Languages with a categorial fluidity

It is important to note that not having the distinction among the categories does not mean lacking distinction among the functions like reference, predication, and modification (Gil, 2005). Rather, when we assume that there is no N-V distinction, it means that there is no distinction in the mental lexicon. The subsequent step is to investigate the extent of fluidity among LCs in these flexible languages by using empirical data to understand the absence of categorial distinction.

The paper is organized as follows: section 2 presents empirical support from Santali, an AA language that is reported to have a flexible word class (Table 1), to show that even if there is no N-V distinction in the lexicon, syntax can differentiate between the two. It also uses some diagnostic tests to establish that the construction types under study are instances of verbalization. Section 3 compares Santali verbalization with similar constructions in English-type languages to differentiate between compositional and idiosyncratic verbalization. Subsequently, in section 4, I differentiate Santali verbalization as re-categorization

of already categorized elements from root-derived categorization. This section also discusses the functional head in the derivation responsible for the extreme fluid verbalization in Santali. Section 5 concludes the paper.

2 Fluid categorization in Santali

2.1 Santali

Santali is one of the widely spoken languages from the North Munda sub-group spoken dominantly by the Santal community, resulting in it being the third most spoken AA language. It is majorly spoken in central and eastern parts of India, including the states of Odisha, Jharkhand, Chhattisgarh, Madhya Pradesh, Uttar Pradesh, West Bengal, and Maharastra, etc. (Anderson, 2015) and also in eastern Nepal and western Bangladesh (Peterson, 2015). The specific empirical support for this study is provided by the data from Santali spoken in Odisha and the border regions of Odisha and Jharkhand (Figure 1).

Figure 1: Santali spoken regions in India

The highlighted regions in Figure 1 include districts like Saraikela, Purbi Singhbhum,

Paschimi Singhbhum, Gumla, and Simdega in the state of Jharkhand and Mayurbhunj, Sundargharh, and Jharsuguda in Odisha where Santali is predominantly spoken.

2.2 Verbalization as Fluid Categorization

As mentioned earlier, Santali, along with many other AA languages, is well known in the literature for its weak distinction between N and V. The broader classification substituting the rudimentary N-V classification is the contrast between an argument and a predicate. Any morpheme can attach with a case marker, number marker, and sometimes a definite marker and behave like an argument/N(P). Similarly, a concept can merge with tense, aspect, mood (TAM) markers, phi, and voice markers to behave as a predicate/V(P). This entails that there is no N-V distinction in the lexicon, since the same concept can behave as an N or a V in syntax, depending on the grammatical markers it attaches to. I use the terms N and V to denote 'argument' and 'predicate', respectively. In this paper, since we focus majorly on the verbalization, we will see how the TAM, phi and the voice marker verbalize any morpheme in Santali (1). In (1), the tense/voice marker *-en* is attached to the verb *sen* 'go' (1a). The same marker *-en* is also attached to the noun *raajaa* 'king' (1b) stamping out the N-V distinction in the lexicon and motivating a syntactic categorization assumption.

(1) Santali

a. uni sen-en-a-e
He go-PST.MID-FIN-3SG
'He left.'

b. uni raajaa-en-a-e
He king-PST.MID-FIN-3SG
'He king-ed.' (He became a king)

Peterson (2003, 2010, 2011, 2015) also shows similar evidence from Kharia that supports a syntactic account of categorization (2).

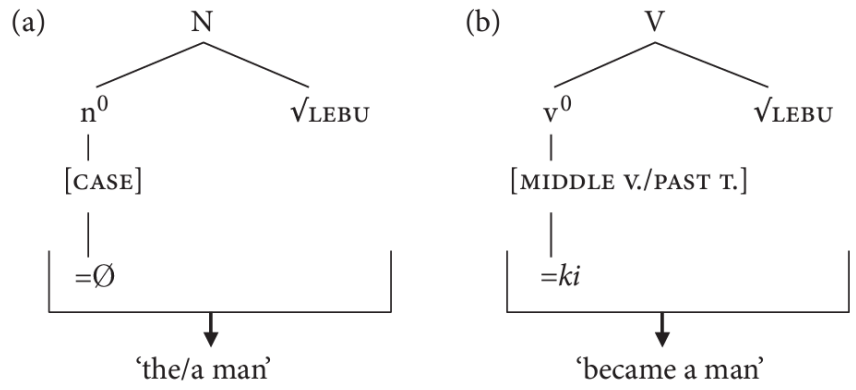
(2) Kharia

a. lebu Del-ki
man come-PST.MID
'The man came.'

b. bhagwan lebu-ki
God man-PST.MID
'God mann-ed.' (God became a man)

According to Peterson, the categorial status of a morpheme is decided depending on whether a root merges with a nominal (n) or verbal (v) categorizer (3).

(3) Kharia



In this paper, however, while analysing such a fluid categorial phenomenon in Santali, I claim that even if categorization in the AA languages like Santali and Kharia take place in the syntax, there is more to the process of categorization than what has been established in the literature (Peterson, 2003, 2011, 2015) (see Section 4 for more).

Before giving an analysis of how verbalization of non verbal entities happens in Santali and comparing it with similar looking phenomenon in other languages, we need to establish if the constructions (1) are really instances of verbalization. Looking at the constructions in (1), we can have two possibilities on the surface. First is that the non-verbal element like *raajaa* 'king' (1b) merges with the verbal markers and behaves as a verb. Secondly, there is a possibility of the presence of a null copula [2] that takes the verbal clitics which results in a verbalized kind of predicate on the surface. We now use some diagnostics to check if the verbalized looking structure is really verbalization or attachment of verbal markers on a null copula.

2.3 The Verbalization Tests

We perform three tests to see if the verbal clitics attach to the non-verbal entities like N or A, resulting in denominal/deadjectival verbalization or the verbal markers attach to the null copulas presenting the verbalization-like illusion in a structure like (1b) repeated as (4).

(4) uni raajaa-en-a-e
He king-PST.MID-FIN-3SG
'He king-ed.' (He became a king)

[2]The motivation for the null copula assumption comes from one of the comments during the FASAL (14) presentation. The idea was that since Santali verbalization structures give a regular become meaning, there could be a null copula present and the TAM PF forms, that look like cliticizing to the N/A, are actually markers of the null (become) copula. The tests in section 2.3 proves the verbalization claim, contrasting with the null copula assumption.

2.3.1 NP Scrambling

The first test concerns sentences with NP scrambling.
Test 1: NP1 NP2 -TAM.phi = NP2 NP1 -TAM.phi
Let us assume the possibility of presence of a null copula that takes the verbal markers in the predicate in (4). Therefore, we assume that in (4), *uni* 'he' is NP1 and *raajaa* 'king' is NP2, not a denominal verbal. There is a null copula after NP2, which hosts the verbal markers. If this assumption is true, Santali should behave like any free word order language where constructions like (4) are possible using a become copula resulting change of state semantics. Also, since Santali is a free word order language, we should be able to scramble both the NPs of a structure like (4) and get the same meaning. We can see that in Hindi, a free word order language that shows change of state semantics using a become copula, changing the order of NP1 and NP2 (5b) doesn't result ungrammaticality. However, when we apply the NP scrambling test to Santali, changing the order of the NPs in (6b) doesn't give the identical semantics as (6a).

(5) Hindi
   a. vah raajaa ban-a
      He king become-3SG.PST
      'He became a king.'
   b. raajaa vah ban-a
      King he become-3SG.PST
      'He became a king.'

(6) Santali
   a. uni raajaa-en-a-e
      He king-PST.MID-FIN-3SG
      'He became a king.'
   b. *raajaa uni-en-a-e
      King he-PST.MID-FIN-3SG
      'He became a king.'[3]

The null copula assumption is invalid according to the NP scrambling test.

2.3.2 Displacing verbal clitics

Verbal clitics are displaced in the second test.
Test 2: -TAM.phi NP1 NP2 = -TAM.phi NP1 NP2
Again, considering the free word order nature of Santali, if the null copula assumption is

[3]The sentence itself is not ungrammatical. When we scramble the NPs leaving the verbal clitics *in situ*, assuming there is a copula hosting it, the meaning of the sentence changes. The meaning of the sentence changes to 'The king became him', which is different from the expected meaning, 'He became a/the king'.

true, we should be able to displace the verbal clitics assuming they are hosted by the null copula, not the NP2. If the structure after displacement is grammatical, it will prove that constructions like (4) are not instances of verbalization. In Hindi, such a displacement is possible (7), in contrast to Santali (8).

(7) Hindi

a. vah raajaa ban-a
He king become-3SG.PST
'He became a king.'

b. ban-a vah raajaa
become-3SG.PST he king
'He became a king.'

(8) Santali

a. uni raajaa-en-a-e
He king-PST.MID-FIN-3SG
'He became a king.'

b. *-en-a-e uni raajaa
-PST.MID-FIN-3SG he king
'He became a king.'

(8b) shows that fronting a verbal clitic (with the assumed null copula) would yield an ungrammatical structure. This implies that there is no null copula and only *raajaa* can host the verbal clitic.

2.3.3 Inserting an intervener

In the third and the final test for verbalization we can check if the TAM and phi markers are attached to the NP or a null copula by inserting an interviner between the NP2 and the verbal clitics in (10a) and the resulting construction (10b) will be acceptable. Again, comparison with Hindi in (9), that has a become copula, shows that insertion of any grammatical marker after the NP2 doesn't incur ungrammaticality.

(9) Hindi

a. vah raajaa ban-a
He king become-3SG.PST
'He became a king.'

b. vah raajaa hi ban-a
He king FOC become-3SG.PST
'He became only a king.'

(10) Santali

a. uni raajaa-en-a-e
He king-PST.MID-FIN-3SG
'He became a king.'

b. *uni raajaa da-en-a-e
he king FOC-PST.MID-FIN-3SG
'He became only a king.'

If a null copula is the host to the verbal clitics, inserting any marker after the NP2 shouldn't create any issue, considering we are not inserting anything between the hypothetical null copula and the verbal clitics. Still, the ungrammaticality in (10b) shows that when we insert the focus marker *da*, it breaks the verbal structure of the denominal verbal and results in ungrammaticality.

These three tests show that the cliticization on the NPs in Santali take place because of verbalization where any category like N or A turn into verbs. Although a phenomenon like verbalization is not so unique across languages, as languages like English also has instances of verbalization (Clark & Clark, 1979), Santali verbalization structure are quite different from English-type languages and show much more productivity and regularity than other languages.

3 Comparing Santali Verbalization with English

Constructions like (1b) repeated as (4) are certainly not unique to Santali, as we see such instances of verbalization even in English (11). The sentences in (11) seem similar to to the Santali verbalized structure in (1b) and (4), since 'hammer', 'chair', 'water', 'tape', 'chain', etc. are prototypically used as nouns in English.

(11) English

a. He hammered the metal.
b. The professor chaired the meeting.
c. Peter taped the box.
d. The policeman chained the criminal.

While it may look like English and Santali are doing the same thing, we shall soon see that it is not the case. One of the major differences between English-type and Santali-type verbalization is that English verbalization is both idiosyncratic and compositional. In (11) two distinct types of verbalization can be noticed. Arad (2003) differentiates them as idiosyncratic and compositional[4] verbalization. The differences between idiosyncratic and compositional verbalization, on the surface, arises from the kinds of meaning each type

[4]Panagiotidis (2015) use the terms hammer-type and tape-type verbalization for idiosyncratic and compositional, respectively.

carries. 'Hammer' and 'chair' in (11a) and (11b) do not have any compositional correspondence with their nominal counterparts. On the other hand, 'tape' and 'box' in (11c) and (11d) have direct semantic correspondence with their nominal counterparts. 'hammer(v)' doesn't mean hit with a hammer, but 'tape(v)' mean seal with a tape. Arad shows example like (12) and (13) to argue that verbs in (12) are idiosyncratic and the meanings of the verbs are not dependent on a corresponding noun. The ungrammaticality in (13), on the other hand, shows that the verbs are compositional, and verbs like 'tape', 'chain', or 'button' cannot exist without the exact physical objects 'a tape', 'a chain', or 'a button'.

(12) English (Arad, 2003)
   a. He hammered the nail with a rock. (Kiparsky, 1982)
   b. String him up with a rope!
   c. She anchored the ship with a rock.

(13) English (Arad, 2003)
   a. *She taped the picture to the wall with pushpins.
   b. *They chained the prisoner with a rope.
   c. *Jim buttoned up his pants with a zipper.

Thus, English has both compositional as well as idiosyncratic verbalisation. Santali, however, displays only one type of verbalization, i.e., the compositional verbalization (14) [5]. The past tense middle voice marker *-en* in (14a) and (14b) produce intransitive structures and the past tense active voice clitic *-kidi* in (14c) and (14d) give a transitive structure of the intransitive counterparts.

(14) Santali
   a. uni dhiri-en-a-e
      He stone-PST.MID-FIN-3SG
      'He stoned.' (He became a stone)
   b. merhed martul-en-a-e
      metal hammer-PST.MID-FIN-3SG
      'The metal hammered.' (The metal became a hammer)
   c. jon uni-ke dhiri-kidi-a-e
      John he-ACC stone-PST.ACT-FIN-3SG
      'John stoned him.' (John changed him to become a stone)
   d. uni merhed-ke martul-kidi-a-e
      He metal-ACC hammer-PST.ACT-FIN-3SG
      'He hammered the metal.' (He changed the metal to become a hammer)

[5]Even if it shows tape-type compositional verbalization, Santali verbalization is much more regular and predictable than English, since, it shows only change of state semantics.

In Santali verbalization, there is no idiosyncrasy like some verbalized structures in English (11a) and (11b). The verbs in (14) show extreme predictability as they have direct semantic correspondence with their nominal counterparts. The Santali verbalized elements also display change of state (become) semantics (see Section 4.3 for more). Santali seems to have an extreme case of categorial fluidity in the case of verbalization; not only referent words like nouns (14), but any word can inflect for a verbal marker like tense and voice and get verbalized in syntax (15). Adjectives like *maaraang* 'big' (15a), demonstratives like *noa* 'this' (15b), kinship DPs like *ini bohya* 'my sister' (15c), animal names like *seta* 'dog' (15d), and even the most restricted kind of noun, i.e., proper names like *binit* 'Vineet' (15e) also get verbalized by taking the tense and voice clitics.

(15) Santali

a. daare maaraang-en-a
Tree big-PST.MID-FIN
'The tree bigged.' (The tree became big)

b. hana noa-en-a
That this-PST.MID-FIN
'that this-ed.' (That became this)

c. uni ini bohya-en-a-e
She my sister-PST.MID-FIN-3SG
'She my sister-ed.'[6] (She became my sister)

d. uni seta-en-a-e
He dog-PST.MID-FIN-3SG
'He dogg-ed.' (He became a dog)

e. uni binit-en-a-e
He Vineet-PST.MID-FIN-3SG
'He Vineet-ed.' (He became Vineet)

In comparison to the fluidity of verbalization in (15), English-type languages have some restrictions on verbalization.

The extreme compositionality seen in Santali verbalization (15) shows that the verbalized elements do not display any idiosyncrasy in meaning and carry the semantics of categories like, N, A, etc. There is always a change of state meaning in the formation of verbs from some category. This entails that verbalization in Santali is, in fact, re-categorization of an already categorized element. Now since, this is established from (15) that verbalization is re-categorization in Santali due to the lack of idiosyncrasy, the next step would require a detailed analysis of the idiosyncratic and compositional verbalization to show how the Santali verbalization takes place in Syntax.

[6]Context: My father married someone who had a daughter.

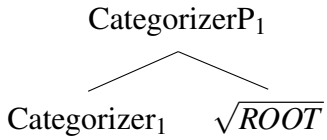
4 Verbalization is Re-categorization

It is evident from the extreme fluidity of the verbalization that there is a significant amount of overlap among categories in Santali. Words are underspecified with the categorial value in the lexicon. We employ a syntactic approach of word formation to explain Santali categorization. The fundamental assumption is that roots enter into the syntactic space without any categorial value. It is in syntax by the categorizer heads like *n* or *v*, the roots get their nominal or verbal categorial value. Since, we noticed both idiosyncratic and compositional verbalization in English (11) and extremely compositional verbalization in Santali, the subsequent step is to understand how the both types are generated in the syntax, considering syntax is the single universal derivational engine for word formation (Marantz, 1997, 2000).

4.1 Idiosyncratic and Compositional Verbalization on the Structure

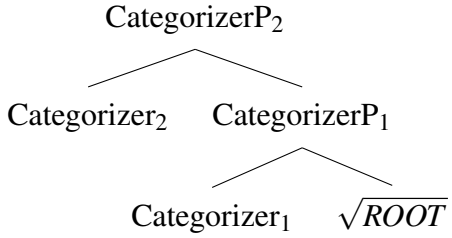
Arad (2003) discusses the differences between root-derived and word-derived verbalization to distinguish between idiosyncratic and compositional verbalization, respectively. According to the locality constraint on interpretation of roots (LCIR) (Arad, 2003), the interpretation of the root is restricted to the first categorizer[7] position in the derivational domain. In (16), $categorizer_1$ is the first categorizer, and the semantically underspecified root gets the interpretation when it merges with $categorizer_1$. The domain of the first categorizerP forms a closed interpretation domain (CID), and anything outside the CID does not have access to any atomic unit inside the domain, such as the root. The CID is the idiosyncratic/ non-productive domain.

(16)



Once the root merges with a categorizer, the meaning of the root is fixed in the first categorizerP. Any categorial derivation outside the CID, doesn't categorize the root, but recategorizes the first categorizer. In (17), the $categorizer_2$, that merges with the $categorizerP_1$, can only access the fixed interpretation of the root on the $categorizer_1$. It creates the compositional/ regular domain since it carries the semantics of the first categorizer head.

(17)



[7]The first categorizer position is described as the first phase position by Panagiotidis (2015).

From (14, 15), it is evident that Santali verbalization is extremely regular, displaying compositional semantics of an already categorized element, and never carries any idiosyncratic meaning. Therefore, the position of the Santali verbalizer head in the structure is the $Categorizer_2$ position in (17).

4.2 Against the Root-derived Approach

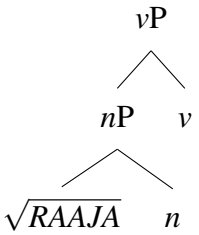
Since, according to Arad (2003), the idiosyncratic meaning of the root is restricted to the first merge position, the denominal and deadjectival kinds of verbalization doesn't take place by merely merging a verbalizer to a root. Based on this argument, we contradict the structure for the formation of N and V (3) in AA languages given for Kharia (Peterson, 2003, 2011, 2015) following the skeletal structure in (17).

In (1b) or (4), repeated here as (18), the V entails the interpretation of an N *raajaa* 'king'. When the V in the process of verbalization carries the meaning of an N, there must be an intervening nominal projection in the structure (Arad, 2003, p.759).

(18) uni raajaa-en-a-e
He king-PST.MID-FIN-3SG
'He king-ed.' (He became a king)

The *n* in (18) works as an intervener between the *v* and $\sqrt{RAAJAA}$, forming an N in the CID. According to the LCIR, $\sqrt{RAAJAA}$ gets nominalized in the CID and the meaning of it is fixed as a noun for the further derivations. When the *v* merges in the derivation, it accesses the nominal semantics of the $\sqrt{RAAJAA}$ from the *n* and verbalizes the noun *raajaa* 'king', resulting in the re-categorization of the N 19.

(19)



The intervening nominalizer brings nominal compositional meaning to the verbalized structure. The following section further elaborates on the compositionality dealing with the productive 'become' semantics in a verbalized structure.

4.3 Compositionality on the Structure

Santali carries a uniform *become* (change of state) semantics in the verbalized structure (15). At this stage, a pertinent question to ask is which head takes care of the *become* semantics in the structure. We adopt the analysis of Embick (2004) for resultative secondary

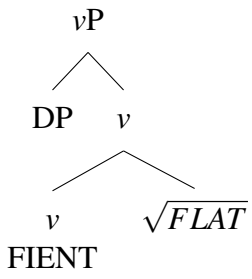
predicates (RSP) in English to see the position of become semantics on a verbalized structure.

Embick (2004) implements the analysis of Hale & Keyser (1993) for deadjectival verbals (20) where the *v*, that merges with the root to verbalize it, carries the *become* semantics as a feature [FIENT].

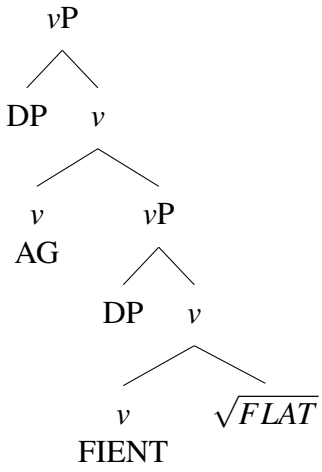
(20) a. The metal flatt-en-ed.
b. The smith flatt-en-ed the metal.

According to Embick, the [FIENT] feature is the become operator that denotes change of state or the transition event. In the process of verbalization, [FIENT] on the verbalizer (*v*), assigns the 'became flat' meaning to 'flat' (21). To explain the transitive structure in (20b), Embick puts the [FIENT] feature on the lower *v* in (22) that provides the 'changed something (the metal) to flat' meaning. The [AG] feature on the upper *v* in (22) is an agentive feature that licenses the agent on the external argument DP.

(21) Intransitive



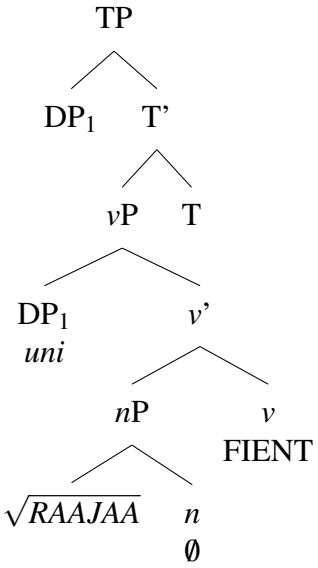
(22) Transitive



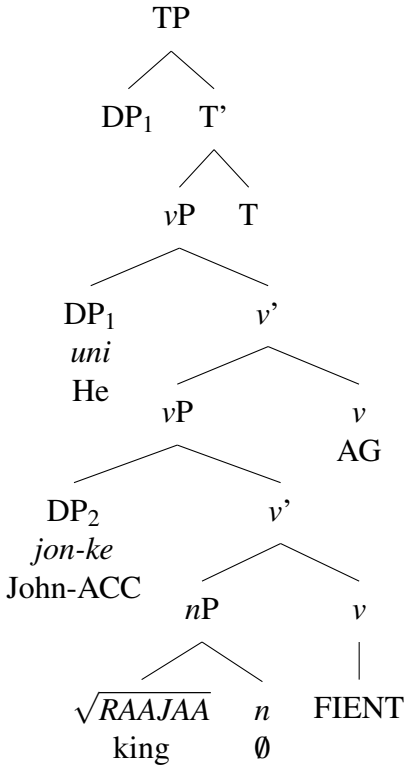
Since a productive *become* semantics is seen in Santali, unlike a restricted set of deadjectival verbalization in English, we adopt Embick's analysis to propose a [FIENT] feature on the verbalizer head in any word-derived verbalization (15) in Santali. Structures like (24) and (25) can explain the intransitive and transitive verbalized constructions like (23a) and (23b), respectively in Santali.

(23) Santali
a. jon raajaa-en-a-e
John king-PST.MID-FIN-3SG
'John king-ed.' (John became a king)
b. uni jon-ke raajaa-kidi-a-e
He John-ACC king-PST.ACT-FIN-3SG
'He king-ed John.' (He changed John into a king)

(24) Intransitive



(25) Transitive



One of the major differences between Embick's analysis for deadjectival verbalization in English and our current analysis for word-derived (denominal, deadjectival, etc.) verbalization in Santali is that we emphasize on the re-categorization of already categorized elements by showing an intermediate categorizer.

5 Conclusion

One of the major findings of the current paper is that Santali verbalization is word-derived (re-categorized from a categorized element), not root-derived. Categorization in Santali-type languages is much more fluid than English-type languages in terms of productivity and compositionality. Secondly, unlike English-type languages, any concept or lexical category can be verbalized in syntax in Santali. The presence of the 'become' operator as a FIENT feature on the verbalizer head is the locus of the extreme verbal productivity in Santali. Even if syntax is the single derivational engine for word formation, there are language-specific variations of categorization depending on the productivity of word formation.

References

Anderson, Gregory DS. 2015. *The Munda languages*. Routledge.

Arad, Maya. 2003. Locality constraints on the interpretation of roots: The case of Hebrew denominal verbs. *Natural Language & Linguistic Theory* 21(4). 737–778.

Baker, Mark C. 2003. *Lexical categories: Verbs, nouns and adjectives*, vol. 102. Cambridge University Press.

Bruening, Benjamin. 2018. The lexicalist hypothesis: Both wrong and superfluous. *Language* 1–42.

Clark, Eve V & Herbert H Clark. 1979. When nouns surface as verbs. *Language* 767–811.

Croft, William. 2002. *Typology and universals*. Cambridge University Press.

Embick, David. 2004. On the structure of resultative participles in English. *Linguistic Inquiry* 35(3). 355–392.

Embick, David & Rolf Noyer. 2007. Distributed morphology and the syntax—morphology interface .

Evans, Nicholas. 2000. Kinship verbs. *Empirical Approaches to Language Typology* 103–172.

Gil, David. 2005. From repetition to reduplication in Riau Indonesian. In Bernhard Hurch (ed.), *Studies on reduplication*, 31–64. Berlin, Boston: De Gruyter Mouton.

Hale, Ken & Samuel Keyser. 1993. On argument structure and the lexical expression of syntactic relations. *An Annotated Syntax Reader* 312–327.

Harnad, Stevan. 2017. To cognize is to categorize: cognition is categorization. In *Handbook of categorization in cognitive science*, 21–54. Elsevier.

Kiparsky, Paul. 1982. Word-formation and the lexicon, Mid-America Linguistics Conference.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In *Proceedings of the 21st Annual Penn Linguistics Colloquium*, vol. 4 2, 201–225.

Marantz, Alec. 2000. Roots: the universality of root and pattern morphology. In *Conference on Afro-Asiatic languages, University of Paris vii*, vol. 3, 14.

Panagiotidis, Phoevos. 2015. *Categorial features* 145. Cambridge University Press.

Peterson, John. 2003. Languages without nouns and verbs? an alternative to lexical classes in Kharia. In *Old and new perspectives on South Asian languages. Grammar and semantics. Papers growing out of the Fifth International Conference on South Asian linguistics (ICOSAL-5), held at Moscow, Russia in July*, .

Peterson, John. 2010. Language contact in Jharkhand: Linguistic convergence between Munda and Indo-Aryan in eastern-central India. *Himalayan Linguistics* 9(2).

Peterson, John. 2011. A grammar of Kharia: A south Munda language. In *A grammar of Kharia*, Brill.

Peterson, John. 2015. Kharia. In *The Munda languages*, 434–507. Routledge.

Rau, Felix. 2013. Proper names, predicates, and the parts-of-speech system of Santali. *Rijkhoff & van Lier (eds.)* 169–184.

Rijkhoff, Jan & Eva van Lier. 2013. *Flexible word classes: Typological studies of underspecified parts of speech*. OUP Oxford.

Sapir, Edward et al. 1944. Grading, a study in semantics. *Philosophy of Science* 11(2). 93–116.

Whaley, Lindsay J. 1996. *Introduction to typology: The unity and diversity of language*. SAGE publications.

Modal Debris: Threefold Ambiguities between Permission, Weak Necessity, & Strong Necessity in Bengali

Shrayana Haldar, *Massachusetts Institute of Technology*

Abstract

In this paper, I explore the possibility of X-marking (in the sense of von Fintel & Iatridou 2023) used in weak necessity modals being null. I argue that this is exactly what seems to be the case in a hitherto undiscussed phenomenon in Bengali, in which the modal that's canonically described as the strong necessity modal of the language shows a systematic ambiguity between strong and weak necessity in upward entailing environments, and between strong necessity and permission in non-upward-entailing environments. The behavior in upward entailing contexts can be understood if X-marking (that is known to turn strong necessity modals into weak necessity ones) can have null exponence, and the behavior in non-upward-entailing contexts can be explained if Staniszewski's (2022) account of weak necessity and X-marking is espoused, which involves strengthening an underlyingly existential meaning into a universal one. Crucially, the QR approach to neg-raising in weak necessity modals fails to explain the facts. I also address an independent problem of alternatives faced by Staniszewski's account and propose a solution for that.

Keywords: modality, weak necessity, neg-raising, exhaustification, scalar implicatures

1 Introduction: X-marking

It has been well-known since von Fintel & Iatridou (2008) that there's a very robust cross-linguistic trend of deriving the Weak Necessity (henceforth, "WN") modal by putting a special marker on top of the Strong Necessity (henceforth, "SN") modal. In von Fintel & Iatridou (2023), this marker has been dubbed *the X-marker*. This is easily observed in Romance, among other languages that von Fintel and Iatridou survey. The French examples that illustrate this are in (1).

(1) a. Tu **devrais** faire la vaisselle, mais tu n'es pas obligé.
you **must-X** do the dishes but you NEG-be-2s NEG obliged
"You ought to do the dishes, but you're not obliged to do them."

b. #Tu **dois** faire la vaisselle, mais tu n'es pas obligé.
you **must** do the dishes but you NEG-be-2s NEG obliged
The only possible reading, which is contradictory:
"You must do the dishes, but you're not obliged to do them."

von Fintel & Iatridou (2008, 2023) illustrate that language after language follows this pattern. See their papers for more data to the same effect. What I will explore in this paper is a mostly empirical question: what if the X-marker is null? von Fintel and Iatridou's observations suggest that, if a language has a null X-marker, then a systematic ambiguity will arise between SN and WN. In the rest of this paper, I will make some observations that will lead us to believe that Bengali is such a language, unlike what has been assumed before.

The following is how the rest of the paper is organized. In section 2, I elaborate on the empirical evidence for the SN-WN ambiguity in Bengali and point out that there's a possibility of analyzing this as a consequence of having a null X-marker. In section 3, I show various ways of detecting a permission reading in the modal, which is expected under Staniszewski's account, and thus, provides cross-linguistic support for it. In section 4, I summarize in prose how this whole state of affairs can be accounted for under that account. Section 5 points out an independent problem for the process of exhaustification to lead to the desired meaning, and I propose a solution to that, which I call *Logical Parallelism*, a contrast on the available alternatives of an LF. Section 6 concludes the paper.

2 Bengali: ambiguity between SN and WN

Bengali, like many other Indo-Aryan languages such as Hindi, doesn't have dedicated lexical elements for modals. (See Bhatt 1999 for Hindi.) For instance, the Bengali copula *[ɦɪɒ̰ɔa]* has been reported to express universal modal force, as shown in (2) (Bjorkman & Cowper 2016, Lahiri 2022).[1] The subject of this construction is always marked with a dative. The predicate of the prejacent of the modal always shows an invariant third person agreement, which is the default agreement. Here, only the obligation reading ("$\Box_{SN}$") is available, the permission reading ("$\Diamond$") isn't (but of course, it's entailed.) (All Bengali examples are given in the IPA.)

(2) to̱-ke e-ʈa kor-te̱ ɦɪɒ-ḛ.
2.SG.INFRML-DAT this-CLF do-INF COP-PRES.3
One of the possible readings (see below for more):
"You have to do this." (✓$\Box_{SN}$, ✗$\Diamond$)

However, what has gone unmentioned in the literature on Bengali modals to the best of my knowledge is that *[ɦɪɒ̰ɔa]* can also have a WN interpretation in upward-entailing (henceforth, "UE") environments. We can appreciate the WN reading of *[ɦɪɒ̰ḛ]* by looking at (3)-(4). In (3a-3b), with different adverbs in the conjuncts, contradiction arises if the modal isn't changed appropriately. (3c) shows that when the modal is changed appropriately, contradiction doesn't arise. That is, the two modals in (4) must be different in exactly the same way the modals in (3c) are.

1. I will restrict attention to the present tense form of *[ɦɪɒ̰ɔa]* in this paper, which is *[ɦɪɒ̰ḛ]*.

(3) a. #You should always do this, but right now, you shouldn't do this.

b. #You always have to do this, but right now, you don't have to do this.

c. You should always do this, but right now, you don't have to do it.

(4) ṯo-ke e-ʈa ʃɒb.somɒe̯-i kor-ṯe fiɒ-e̯, kiṉṯu ækʰon
2.SG.INFRML-DAT this-CLF all.time-FOC do-INF COP-PRS.3 but now
ṯo-ke e-ʈa kor-ṯe fiɒ-e̯ n-a.
2.SG.INFRML-DAT this-CLF do-INF COP-PRS.3 NEG-IMPFV

The only possible non-contradictory reading:
"You should always do this, but right now, you don't have to do it."

This is a challenge for von Fintel & Iatridou (2008, 2023) because they provide extensive cross-linguistic evidence to the effect that asserting the SN modal in a language and then subsequently negating it leads to contradiction, (1b) being an example of that. That means that the two modals in (4) — one affirmed and the other negated — must actually be two different things. What are these different things and how do they come to be different? An answer to this question becomes apparent when we pay attention to several recently observed interesting facts about WN modals which leave their intricate traces in the behavior of the ambiguous Bengali modal *[hɒe̯]*. What these facts bring out is an underlying permission meaning in WN modals. These were observed in Staniszewski (2022) for English. I will show below that this permission reading is also detectable in the WN *[fiɒe̯]*. This will suggest that the way the two modals in (4) are different in exactly the way a core component of Staniszewski's account of WN predicts they would be.

3 The permission reading

Staniszewski (2022) has made the intriguing observation that, when embedded under *no longer*, modals like *should* and *supposed to* trigger a presupposition that has the meaning of an existential, *i.e.,* a permission modal. This is shown in (5)-(6). The way (5)-(6) show this is the following: assuming p stands for the proposition *Students enter through the main lobby*, and q, for the proposition *Students enter through the cafeteria*, the (a) examples has the schema of $\Box(p \vee q)$, which gives rise to the distributive inferences $\Diamond p$ and $\Diamond q$; that is, *Students used to be **allowed** to enter through the main lobby* and *Students used to be **allowed** to enter through the cafeteria*. Because of these inferences arising from the (a) examples, the presuppositions of the (b) examples are satisfied in both discourses. Crucially, if *should* and *supposed to* only had a necessity meaning, then the presuppositions would have necessity meanings in them; that is, we would predict the presuppositions *Students used to be **required** to enter through the main lobby* and *Students used to be **required** to enter through the cafeteria*. Therefore, these two modals can't be given a simple analysis under which they just have necessity meanings. (7) shows that the presupposition triggered by *no longer* is indeed that its prejacent used to be true.

(5) a. It used to be the case that students should either enter through the main lobby or the cafeteria.

b. . . . but now they no longer should enter through the cafeteria.

(6) a. It used to be the case that students were supposed to either enter through the main lobby or the cafeteria.

b. . . . but now they are no longer supposed to enter through the cafeteria.

[Staniszewski (2022), (74), (75): 204]

(7) a. #I don't know whether John used to smoke, but he no longer smokes.

b. #John never used to smoke, but he no longer smokes.

Iatridou & Zeijlstra (2013), Homer (2015), and Zeijlstra (2017) have accounted for the ability of modals like *should* and *supposed to* to scope over sentential negation by proposing that they QR over negation. Zeijlstra (2017) should be especially noted since he gives a compositional account of neg-raising. There are several problems with this QR approach that have been pointed out in Staniszewski (2022), Jeretič (2021), Jeretič & Thoms (2023). The reader is pointed to this body of literature for evidence to against a QR approach in general, but, for our purposes here, there's no need to go over them, because it's easy to see why a QR approach won't work for (5)-(6): in the (b) examples, QR will result in LFs schematized as in $\Box >$ *no longer* $> p$, where p stands for the prejacents *Students enter through the main lobby* and *Students enter through the cafeteria*. In this LF, the prejacent of *no longer* is not MODAL $> p$, but just p. That is, the QR approach predicts the presupposition that p. But, as can be easily checked, these presuppositions are unattested because the following discourses are felicitous.[2]

(8) a. It used to be the case that students should either enter through the main lobby or the cafeteria. But {they never used to enter through the cafeteria/I don't know whether they ever used to enter through the cafeteria.} Now, the rules have changed and they no longer should enter through the cafeteria.

b. It used to be the case that students were supposed to either enter through the main lobby or the cafeteria. But {they never used to enter through the cafeteria/I don't know whether they ever used to enter through the cafeteria.} Now, the rules have changed and they are no longer supposed to enter through the cafeteria.

2. There's also the very recent paper by Mirrazi & Zeijlstra (2023), who propose an exhaustification-based account that I became aware of only after developing this paper. I leave a consideration of their paper to a future occasion. For what it's worth, they don't predict an ambiguity between strong and weak necessity.

Exactly same effects are found in Bengali. In the sentence in (9), we can see, thanks to the presupposition triggered by *[aɾ]* "no longer", that *[ɦiɒ̯e̯]* can have a permission reading, just as in (5)-(6), precisely because (9) is felicitous in both of the given contexts. Say, the prejacent of the modal is *p*, *i.e.,* the addressee wandering around all day. Felicity in the context in (9b) shows that the presupposition can be $\Diamond p$; otherwise, the sentence wouldn't be felicitous in this context. In fact, in the context in (9b), even if one assumes that for the addressee to have been wandering around all day before she got engaged is frowned upon, given the conservatism, **therefore, definitely not obligatory,** the sentence is still felicitous. Therefore, we can conclude that the permission reading is available. Bengali examples with disjunction, parallel to (5)-(6), show the same behavior. Moreover, the fact that the sentence is fine in the context in (9b) as well confirms that there's an actual ambiguity.

(9) ʈo-ke aɾ ʃara-ɖin ʈo-ʈo koɾ-e gʱuɾ-ʈe ɦɒ-e̯
2.SG.INFRML-DAT any.longer whole-day ONOMAT do-GER travel-INF COP-PRES.3
n-a.
NEG-IMPFV
"You {are no longer {supposed/allowed} to/no longer have to} wander around all day. "

a. ***Context:*** A person whose job involved a lot of wandering around throughout the day has found a new job and no longer has to do all the wandering around they once had to. The speaker says this to them. (✓$\Box_{SN}$)

b. ***Context:*** In a certain conservative society, until a woman is engaged to be married, she has the permission to wander around wherever she wants. But once she is engaged to be married, she is no longer allowed to. In such a situation, a mother says this to her daughter who has been engaged to be married. The daughter never used to wander around before she got engaged. (✓$\Diamond$)

This is part of a whole bunch of non-UE environments where WN modals seem to allow a permission reading, as well as an SN reading, for instance, under *only*, as shown in (10), and in polar questions, as shown in (9). In (10), the effect arises because of the negation in the assertive DE component of *only* (von Fintel 1999, von Fintel & Iatridou 2007). As for (11), in a response to (11a) in the given context, (11b), expressing the existence of a permission, is okay, but (11c), only understandable as expressing an obligation, is not okay. Therefore, the question must be asking about whether being down in the area is okay, not whether there's an obligation to do so. The off-limits nature of the area ensures the naturalness of this permission reading, because then, there would definitely not be an obligation to be there. Again, the SN reading is also available in polar questions when a suitable context is provided, as shown in (12), which again confirms the ambiguity.

(10) t̪o-ke æk^{h}on ʃud̪^{ɦ}u d̪in-eɾ bæla-i̯ beɾ-o-t̪e
2.SG.INFRML-DAT now only day-GEN half.of.the.day-FOC leave-CAUS-INF
ɦiɒ-e̯.
COP-PRS.3
"Now, you are only {supposed/allowed} to go out during the day."

a. ***Context:*** A person whose job involved a lot of wandering around throughout the day has found a new job and no longer has to do all the wandering around they once had to. The speaker says this to them. (✓□$_{SN}$)

b. ***Context:*** In a certain conservative society, until a woman is engaged to be married, she may go out during the day or during the night. But once she's engaged to be married, she no longer may go out during the night, although she retains the permission to go out during the day. The following is said by a mother to her daughter who has been engaged to be married in such a society. The daughter never used to go out either during the day or during the night before she got engaged. (✓◇)

(11) a. ***Context:*** Inside a possibly off-limits area.
A: t̪o-ke ki ek^{h}ane aʃ-t̪e ɦiɒ-e̯?
2.SG.INFRML-DAT POL here come-INF COP-PRS.3
"Are you supposed to come here?" (✓◇)

b. B: ɦiæ̃, ʈ^{h}ik at͡ʃ^{h}e.
yes right exist.PRS.3
"Yes, it's okay."

c. B′: #ɦiæ̃, ɦiɒ-e̯.
yes, COP-PRS.3
"Yes, I am."

(12) a. ***Context:*** The speaker has never seen the addressee's workplace where the latter has to be present for work. The speaker is now there for the first time and they really don't like the place. They ask the addressee this question, with the implication that they're hoping the addressee would say that that they don't have to come here because maybe they can work remotely.
A: t̪o-ke ki ek^{h}ane aʃ-t̪e ɦiɒ-e̯?
2.SG.INFRML-DAT POL here come-INF COP-PRES.3
"Do you have to come here?" (✓□$_{SN}$)

b. B: #ɦiæ̃, ʈ^{h}ik at͡ʃ^{h}e.
yes right exist.PRES.3
"Yes, it's okay."

c. B′: ɦiæ̃, ɦiɒ-e̯.
yes, COP-PRES.3
"Yes, I am."

To sum up, we've seen that a single Bengali modal is ambiguous between SN and WN. This, when understood against the backdrop of von Fintel & Iatridou (2008, 2023), leads us to believe that Bengali is a language where the X-marker can have null exponence. Moreover, although Bengali differs from English in this respect, *i.e.,* the nullness of the X-marker, both of these languages show a particularly intriguing behavior under *no longer* and other non-UE environments, especially, those that trigger presuppositions. The presuppositions triggered in these cases allow a permission reading, which is unexpected if the basic meaning of WN modals involves universal quantification.

4 What this teaches us

There are two main take-aways from this discussion of the SN-WN ambiguity in Bengali. First, since there's an ambiguity between SN and WN in UE environments, it poses a challenge for the robust cross-linguistic picture von Fintel & Iatridou (2023) paint. Second, since there's an ambiguity between SN and a permission reading in non-UE environments, especially in polar questions with no negation in them, it shows that the negated permission reading observed in cases of neg-raising can't be attributable to the modal QRing above negation, as already explained above.

This state of affairs provides striking support for the account of WN given in Staniszewski (2022). His is an account exceptionally complex, explaining which will prevent us from appreciating the paradigm-shifting insight he brings. Therefore, I will schematize the principal components of his analysis in prose and refrain from giving any formalisms. I refer the interested reader to the dissertation for the explicit semantics.

Staniszewski's account seeks to tie the cross-linguistic tendency of X-markers to derive WN from SN with the emergence of the permission reading of WN modals. He proposes that WN modals are not underlyingly WN modals. They begin their lives merged into the tree as a modal whose meaning is indistinguishable from the meaning of SN. For instance, in English, *should* starts off as a modal whose basic meaning is the same as that of, for instance, *have to*. The X-marker is then put on top of it. The X-marker is a generalized existential quantifier over ordering source sequences. This leads to the second step in the life of *should* — the existential force of the permission reading. At the next step, this permission reading is strengthened into the WN meaning through the process of computing scalar implicatures in the grammar (Fox 2007; Chierchia, Fox & Spector 2012; *inter alia*), for instance, by the EXH operator of Bar-Lev & Fox (2020), in UE contexts, and such exhaustification is vacuous in non-UE contexts because the unexhaustified meanings are already the strongest alternatives in their respective sets of alternatives. That is, WN is related to SN derivationally and the meaning of WN is reached through three steps: SN to permission to WN. This is part of a broader line of work that proposes underlyingly existential readings getting strengthened into universal ones via exhaustification (Bowler 2014; Bar-Lev 2018, 2021, Bar-Lev & Margulis 2014; Oikonomou 2022; Singh, Wexler, Astle-Rahim, Kamawar & Fox 2016; Jeretič 2021; *inter multa alia*).

The middle step, where the permission reading is reached, plays the central role in deriv-

ing the permission readings in non-UE contexts. For instance, when the X-marker is merged into the tree and the sister of *no longer* is this permission modal, the permission reading straightforwardly becomes the presupposition in the case of *no longer* > *should*. When the X-marker isn't merged, the permission meaning is simply never generated; therefore, the sister of *no longer* contains the basic SN meaning, which becomes the presupposition in the case of *no longer* > *have to*. In English, there's no optionality in whether the X-marker is merged in the case of *should*. Bengali is special in that there's this extra dimension of optionality in whether or not the X-marker is merged. This optionality of merging X-marking, then, can in principle, account for the ambiguity in Bengali.[3, 4]

So, Bengali, like English, is consistent with Staniszewski's account. But, I would like to argue that the implication is much stronger, in that Bengali bears out a typological prediction that Staniszewski makes. The prediction is that, under *no longer*, **an ambiguity between SN and WN should arise in languages where it's possible to have a null X-marker.** Crucially, this is not something that's predicted by QR accounts of neg-raising in WN, and therefore, pushes us precisely towards an account like Staniszewski's. Once we entertain the possibility of a null X-marker, **there's no longer any puzzle for von Fintel & Iatridou (2008, 2023), because, then, the presence/absence of the null X-marker is how the two modals in (4) differ, and no contradiction arises because the first is *should* and the second is *must*.**

There's also the matter of how the permission reading arises in polar questions. This is a more complicated issue, which involves going into the specifics of the account. But, in short, the question LFs that lead to the permission reading lack the EXH operator, which is why the strengthening doesn't take place. Schematically, the LF looks like (13), where EVEN is an operator that asserts its prejacent and presupposes that its prejacent is the unlikeliest of its alternatives (Karttunen & Peters 1979, Staniszewski 2022). See Iatridou & Tatevosov (2016) for more on this kind of use of EVEN.[5]

(13) ***The ◇ LF:***
[EVEN [whether$_1$ [Q [t_1 [have-to-X p]]]]]

a. *Yes:* [have-to-X p]

b. *No:* [¬ [have-to-X p]]

3. Another crucial aspect of Staniszewski's account involves deriving the weak nature of WN, in the sense of von Fintel & Iatridou (2008, 2023). He implements this via pruning of alternatives in a way that is sensitive to the Question Under Discussion (QUD) of the form "Which preferences do I care about in this situation?" This is a direct translation of Rubinstein's (2014) notion of *negotiable* and *non-negotiable* priorities. See Staniszewski (2022) for more.

4. Recently, Weingartz & Hohaus (2023) have also presented a similar ambiguity between SN and WN in Afrikaans and Samoan. However, they don't consider issues of relative scope between modals and negation. So, their account is not relevant to my concerns here.

5. Staniszewski follows Guerzoni (2004) for how the meanings of polar questions are derived.

As shown above, the positive answer to this question has a permission reading. Therefore, if this question is answered with a positive *[ɦiɒ̰ḛ]* in a declarative sentence, then, because declarative LFs are always exhaustified, we'll get WN if the X-marker is present in the LF and we'll get SN if it isn't. Either way, this declarative answer with *[ɦiɒ̰ḛ]* ends up having a necessity reading, which is infelicitous (recall (11c)), because the question was about whether a permission exists. And, as usual, a negative answer to this question would have the meaning of a negated permission modal. This doesn't rule out (12), because, in that LF, there isn't any X-marker; therefore the SN reading is generated. **Again, the optionality of X-marking is crucial in deriving the ambiguity.**[6]

5 Absence of alternatives

5.1 The problem

Staniszewski derives the WN meaning by exhaustifying the permission reading into a stronger meaning. But this strengthening is only possible when the SN LF is unavaialble to WN as an alternative. Otherwise, the SN alternative would be negated by EXH and further strengthening into the universal reading will no longer be possible. But, given what I've said above

6. For the sake of completeness, I should mention that there's another dimension of optionality, which is whether or not EXH is merged in the structure. This predicts that, if EXH is merged in the question LF, then the WN reading should arise in questions as well. This is borne out, as shown in (1). That is, in questions, the ambiguity is threefold.

(i) a. ***Context:*** The addressee is a boss at an office. They and a friend of theirs are present at an initiation event for new employees. From their own experience at their own office, the friend knows that the boss doesn't have to be present at these events. That is, there's no hard-and-fast rule regarding this. What they do not know is whether, at this office, there's an unwritten, collectively understood desideratum that the boss be preferably present at such events. To get an answer to this question, the speaker asks this question to the boss, the addressee.

A: t̪o-ke ki ekʰane aʃ-t̪e ɦiɒ̰-ḛ?
2.SG.INFRML-DAT POL here come-INF COP-PRES.3
"Are you supposed to come here?" (✓□$_{WN}$)

b. B: #ɦiæ̃, ʈʰik atɟʰe.
yes right exist.PRES.3
"Yes, it's okay."

c. B′: ɦiæ̃, ɦiɒ̰-ḛ.
yes, COP-PRES.3
"Yes, I am."

The LF below accounts for this possibility. Both of the possible answers to this question LF are the strongest alternatives in their respective sets of alternatives. That is, answering with a permission meaning ("Yes, it's okay") is not an option, which explain the infelicity of (1b).

(ii) ***The*** □$_{WN}$ ***LF***
[whether$_1$ [Q [EVEN [EXH [t_1 [have-to-X p]]]]]]

a. *Yes:* [EVEN [EXH [have-to-X p]]]

b. *No:* [EVEN [EXH [¬ [have-to-X p]]]]

schematically, we can identify a problem. Suppose *p* stands for the prejacent of the modal, $\Box_{SN}$ stands for the SN modal, and X stands for the X-marker in (14). Then, (14a) schematizes the LF that leads to the WN meaning, and the SN alternative is in (14b). As I said above, it's this SN alternative that must somehow be unavailable in order for STaniszewski's account to go through.

(14) a. [EXH [$\Box_{SN}$-X *p*]]

b. ***The SN alternative:*** $\Box_{SN}$ *p*

The same problem exists for English, which didn't go unacknowledged by Staniszewski (2022). His final reasoning was that the use of the X-marked structure is governed by the status of priorities in the contexts, unlike in the case of the ⟨*allowed*, *required*⟩ scale, where the only difference is in the quantificational force, not in the ordering source. **Therefore, a X-marked structure and a non-X-marked structure can't compete for the purposes of implicature calculation (Staniszewski 2022: 291).** But the non-X-marked structure is still a possible deletion alternative, in the Katzirian sense (Katzir 2007, Fox & Katzir 2011). **Then, how do we reconcile the unavailability of the SN alternative with the notion of structural alternatives?** Staniszewski leaves this as an open question. Therefore, it seems we haven't found an answer to this question.

5.1.1 Logical Parallelism

I would like to propose an overarching solution to this narrow problem, which, as far as I can see, solves a very specific kind of problem manifesting in myriad ways in the literature. I propose (15).

(15) **Logical Parallelism (LP)**
If an LF has the schema [$_X$ *O* [$_Y$ Z]], then [$_Y$ Z] can't be an alternative of this LF, if *O* is a projection of a logical word (in the sense of Gajewski 2002, Chierchia 2021), unless the logical word at that node is what EXH associates with.

As a simple consequence of LP, the X-marker can't be simply deleted to get the SN structure. Henceforth, I would call such inappropriately derived deletion alternatives *logically non-parallel alternatives*. Also notice that this doesn't prevent the generation of disjunct alternatives from a disjunction, which would require the removal of a logical word *or*, because, whenever that array of alternatives is to be generated, *or* is the associate of EXH. However, in the case of a WN structure — consisting of the SN modal and the X-marker — the X-marker isn't the associate of EXH, the SN modal is. Because, indeed, the associate of EXH is what generates the alternatives, and the modal is what triggers the generation of the subdomain alternatives. This is why the X-marker can't simply be deleted from the structure.

6 Discussion and conclusion

I've argued in this paper that Bengali seems to be a challenge for the cross-linguistic generalization in von Fintel & Iatridou (2008, 2023) that the SN modal of a language can't be both affirmed and negated as part of the same discourse. It no longer remains a challenge if X-marking can be null in the language. However, on a cautionary note, Bengali has a separate X-marking that appears in counterfactual conditionals. For instance, in (16), both antecedent and consequent X-marking is expressed with past habitual morphology. That is, in this respect, Bengali is similar to Hungarian, which also expones antecedent and consequent X-marking with the same morphology (von Fintel & Iatridou 2023: 1491-1492).

(16) o ʤodi ʤe-ṭ-o, ṭa-fiol-e ami-o ʤe-ṭ-am.
3.SG.NOM if go-HAB-PST.3 that-COP-GER 1.SG.NOM-too go-HAB-PST.1
"If (s)he {went/had gone}, then I {would have gone/would go} too."

This past habitual morpheme that serves the purpose of X-marking doesn't appear in (4), unlike, for instance, in Spanish, where the combination of SN and the consequent X-marker yields the WN reading (von Fintel & Iatridou 2023, (61): 1492). So, although we seem to have discovered a null X-marker in Bengali, this appears to be a less than ideal way to understand the WN meaning of *[fiɒ̃e]* because neither antecedent X-marking or consequent X-marking is null in the language. In Haldar (to appear) and ongoing work, I explore this in more detail and my current understanding is that this is, indeed, not a null X-marker, but something else. I refer the interested reader to Haldar (to appear) for some further interesting aspects of this Bengali modal that help us understand how this WN reading might be arising.

To conclude, this paper had two purposes: to provide empirical evidence from Bengali for the possibility of null X-marking, and to point out this cluster of data also provides cross-linguistic support for Staniszewski's (2022) account of WN. I've achieved these two goals by showing that, in Bengali, a single modal is ambiguous between SN and WN, which falls out of Staniszewski's account of X-marking, combined with the possibility of a null X-marker, and that *modulo* this ambiguity, the modal behaves under presupposition triggers like *no longer* exactly as expected from Staniszewski's account. There's also an independent problem of alternatives for exhaustification, *i.e.,* how to prevent the SN LF from being an alternative of the WN LF, which would jeopardize the account. For this, I proposed what I dubbed Logical Parallelism, which prevents the generation of the SN alternative from the WN alternative.

Acknowledgements

I'm grateful to Danny Fox, Kai von Fintel, Sabine Iatridou, and Amir Anvari for the discussions I had with them on this material. Also thanks to the people I consulted for their judgements and the audience for their comments at the LF Reading Group at MIT on November 22, 2023, and FASAL 14, where this was presented. All errors are mine.

References

Bar-Lev, Moshe E. 2018. *Free choice, homogeneity, and innocent inclusion*: The Hebrew University of Jerusalem dissertation.

Bar-Lev, Moshe E. 2021. An implicature account of homogeneity and non-maximality. *Linguistics and Philosophy* 44. 1045–1097.

Bar-Lev, Moshe E. & Danny Fox. 2020. Free choice, simplification, and innocent inclusion. *Natural Language Semantics* 28. 175–223.

Bar-Lev, Moshe E. & Daniel Margulis. 2014. Hebrew *kol*: a universal quantifier as an undercover existential. In Urtzi Etxeberria, Anamaria Fălăuş, Aritz Irurtzun & Bryan Leferman (eds.), *Proceedings of Sinn und Bedeutung 18*, 60–76.

Bhatt, Rajesh. 1999. *Covert modality in non-finite contexts*: University of Pennsylvania dissertation.

Bjorkman, Bronwyn & Elizabeth Cowper. 2016. Possession and necessity: From individuals to worlds. *Lingua* 182. 30–48.

Bowler, Margit. 2014. Conjunction and disjunction in a language without 'and'. In Todd Snider, Sarah D'Antonio & Mia Weigand (eds.), *Proceedings of SALT 24*, 137–155. LSA and CLC Publications.

Chierchia, Gennaro. 2021. On being trivial: grammar vs. logic. In Gil Sagi & Jack Woods (eds.), *The semantic conception of logic: Essays on consequence, invariance, and meaning*, 227–248. Cambridge, United Kingdom: Cambridge University Press.

Chierchia, Gennaro, Danny Fox & Benjamin Spector. 2012. Scalar implicature as a grammatical phenomenon. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: an international handbook of natural language meaning*, vol. 3, 2297–2331. Berlin, Boston: de Gruyter.

von Fintel, Kai. 1999. NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics* 16. 97–148.

von Fintel, Kai & Sabine Iatridou. 2007. Anatomy of a modal construction. *Linguistic Inquiry* 38(3). 445–483.

von Fintel, Kai & Sabine Iatridou. 2008. How to say ought in Foreign: The composition of weak necessity modals. In J. Guéron & J. Lecarme (eds.), *Time and modality*, vol. Studies in Natural Language and Linguistic Theory 75, 115–141. Berlin, Germany: Springer.

von Fintel, Kai & Sabine Iatridou. 2023. Prolegomena to a theory of X-marking. *Linguistics and Philosophy* 46. 1467–1510.

Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presupposition and implicature in compositional semantics*, 71–120. London, United Kingdom: Palgrave Macmillan.

Fox, Danny & Roni Katzir. 2011. On the characterization of alternatives. *Natural Language Semantics* 19(1). 87–107.

Gajewski, Jon. 2002. L-analycity in natural language. Unpublished manuscript. MIT.

Guerzoni, Elena. 2004. *Even*-NPIs in Yes/No questions. *Natural Language Semantics* 12(4). 319–343.

Haldar, Shrayana. to appear. Towards an etiology of weak necessity in a Bengali modal based on homogeneity. In *Proceedings of CLS 60*, .

Homer, Vincent. 2015. Neg-raising and positive polarity: The view from modals. *Semantics & Pragmatics* 8(4). 1–88.

Iatridou, Sabine & Sergei Tatevosov. 2016. Our *even*. *Linguistics and Philosophy* 39(4). 295–331.

Iatridou, Sabine & Hedde Zeijlstra. 2013. Negation, polarity, and deontic modals. *Linguistic Inquiry* 44. 529–568.

Jeretič, Paloma. 2021. *Neg-raising modals and scaleless implicatures*: New York University dissertation.

Jeretič, Paloma & Gary Thoms. 2023. Modals, negation and movement: a reassessment. *Glossa* 8(1).

Karttunen, Lauri & Stanley Peters. 1979. Conventional implicature. *Syntax and Semantics* 11. 1–56.

Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistics and Philosophy* 30(6). 669–690.

Lahiri, Utpal. 2022. Modal-aspect interactions in Bangla. In Samir Alam, Yash Sinha & Sadhwi Srinivas (eds.), *Proceedings of (F)ASAL 11*, .

Mirrazi, Zahra & Hedde Zeijlstra. 2023. Missing words and missing worlds. In Maria Onoeva, Anna Staňkova & Radek Šimík (eds.), *Proceedings of Sinn und Bedeutung 27*, 435–452. Praha: Charles University.

Oikonomou, Despina. 2022. Detecting variable force in imperatives: A modalized minimal approach. *Natural Language and Linguistic Theory* 41. 1201–1256.

Rubinstein, Aynat. 2014. On necessity and comparison. *Pacific Philosophical Quarterly* 95. 512–554.

Singh, Raj, Ken Wexler, Andrea Astle-Rahim, Deepthi Kamawar & Danny Fox. 2016. Children interpret disjunction as conjunction: Consequences for theories of implicature and child development. *Natural Language Semantics* 24. 305–352.

Staniszewski, Frank. 2022. *Modality and time in logical context*: Massachusetts Institute of Technology dissertation.

Weingartz, Siena & Vera Hohaus. 2023. Variable modal strength in Afrikaans and Samoan: Deriving strong necessity from weak necessity. In Mira Grubic, Jeanne Lecavelier, Prarthanaa Manjunath Bharadwaj & Malte Zimmerman (eds.), *Proceedings of TripleA 10*, .

Zeijlstra, Hedde. 2017. Universal quantifier PPIs. *Glossa* 2(1). 1–25.

Borrowing and disappearance of light verbs: Loan-verb integration in Indian languages

AADITYA KULKARNI, *Independent Researcher*

ABSTRACT

In this paper, I discuss patterns of loan-verb integration attested in Indian languages and show that certain English verbal borrowings in languages like Hindi and Marathi can either be accommodated into the host language using a supporting light verb or be directly integrated to carry the host language's verb morphology without needing to undergo any means of verbalization. I argue that syntactic analyses which assume the existence of a common (or identical) verbal functional structure between the donor and recipient (or host) languages to be a prerequisite for direct integration of loan verbs fail to adequately explain this optionality of direct integration. Instead, I show that it is the degree of bilingualism of speakers which makes direct integration of loan verbs into the target language possible; and propose that verbal borrowings are borrowed with an understanding that they are verbs – irrespective of whether they are accommodated using a light verb construction or not.

1 Borrowing and bilingual compound verbs

English verbs, when borrowed into Indian languages, cannot directly take the host language's inflection and need to be accompanied by a light verb which carries the necessary verbal inflections – thus forming what has been termed as a bilingual compound verb (Romaine 1995, Bhatia & Ritchie 2016, Muysken 2016)[1]. This is consistent with observations made in the typological literature (Moravcsik 1975, Muysken 2000, Wichmann & Wohlgemuth 2008) which demonstrates the accommodation of loan verbs using a supporting light verb to be a widely attested strategy of loan-verb integration.

Patterns of English borrowings in Kannada observed by Amritavalli (2017: 9) also fall into this pattern, where a borrowed English verb needs the support of a light verb to appear felicitously in a Kannada sentence (1-2). [2]

(1)	**nanu*	*i:*	*post-ige*	***apply**-idd-i:ni*		
	I	this	post-DAT	apply-AUX.PST-1.SG		
	(intended) 'I have applied to this post.'					
(2)	*nanu*	*i:*	*post-ige*	***apply***	*maḍ*	*-idd-i:ni*
	I	this	post-DAT	apply	make/do	-AUX.PST-1.SG
	'I have applied to this post.'					

Amritavalli (2017) also brings to our attention some interesting patterns of borrowings observed elsewhere in the literature. In Bangla, it seems that borrowed English verbs need a

[1] A bilingual compound verb consists of a loan verb and a 'supporting' light verb from the host language.

[2] For the sake of uniformity, minor changes have been made in the transcription of data cited from previous literature. In the transcribed data from Indian languages, [t] and [d] refer to dental plosives, and [y] refers to the palatal approximant. The original glossing has been largely retained for cited examples.

supporting light verb (3) – just like Kannada; however, Hindi verbs can appear directly in Bangla without needing the support of a light verb (4). Rather, adding a light verb into the mix yields an ungrammatical structure (5).

(3) ami ***celebrate*** *kor-l-am /* ****celebrate****-l-am*
I celebrate do-PST-1 / celebrate-PST-1
'I celebrated.' (Amritavalli 2017: 12)

(4) ***qʰund****-e-che*
find-PRF-AUX
'has found' (Amritavalli 2017: 12)

(5) ****qʰund*** *kor-be*
find do-FUT
(intended) 'will find' (Amritavalli 2017: 12)

Further, like the Bangla-Hindi borrowings, English verbs can also be integrated directly into American Norwegian (6).

(6) jeg ***celebrate****-a*
I celebrate-PST
'I celebrated.' (Amritavalli 2017: 12)

Given that the direct borrowing of verbs is grammatical only in certain language pairs but not in others, this begs the question as to what dictates the grammaticality of these borrowings. Amritavalli (2017: 11) makes the following claim about such cases: "a verb – borrowed or otherwise – needs licensing (i.e., verbalizing) by the verbal functional structure of the particular language". This analysis further assumes the verbal functional structures to be non-identical between languages and thus to be the locus of parametric variation. Therefore, it follows that there must be a verbal functional structure common between the languages where direct integration of a loan-verb (sans the light verb that is) is possible. On the other hand, whenever there is a mismatch between the verbal functional structures of the donor and recipient languages, the need for licensing (or verbalizing) is satisfied by introducing the light verb. From Amritavalli's analysis, it follows that language pairs Bangla & Hindi and English & American Norwegian must have identical verbal functional structures, whereas there must be a mismatch between the verbal functional structures of English and Indian languages such as Bangla, Hindi, and Kannada. This analysis explains as to why a light verb is needed to accommodate the borrowed English verbs in these Indian languages, but borrowing of English verbs in American Norwegian is felicitous without one.

2 Borrowing and disappearance of light verbs

Amritavalli (2017) shows that English verbal borrowings in Indian languages cannot occur with the host language's morphology given the mismatch between their verbal functional structures and thus need to be verbalized using a supporting light verb. In this section, I present some intriguing cases which pose a challenge to this account – where verbs bor-

rowed into Indian languages can appear straightforwardly with the host language's morphology, and hence exhibit optionality in terms of the appearance of a supporting light verb which has been deemed a necessity for borrowing under Amritavalli's licensing analysis.

2.1 Optional direct integration of English verbs in Indian languages

Like in other Indian languages such as Bangla, Hindi, and Kannada, English verbal borrowings in Marathi also appear in a bilingual compound verb (7-8). Under Amritavalli's analysis, this entails a mismatch between the verbal functional structures of Marathi and English.

(7)	*utpadəne*	***tag*** *ke-li*	*gelya-nəntər*	*deal*	*pah-ta*	*yeil*[3]
	products	tag do-PRF	went-after	deal	see-IMPRF	come.FUT

'The deal can be seen after the products are tagged.'

(8)	*social media-wər*	*sərwə kahi*	***post*** *kərɳe*	*t͡saŋgle ahe*		*ka*[4]
	social media-on	everything	post do.INF	good	be.AUX	Q

'Is it okay to post everything on social media?'

However, I will now present cases where English verbs can be borrowed and directly integrated in Indian languages such as Marathi (9-10) and Hindi (11-13). Such cases, though attested largely in the writings in Indian languages on various social media platforms and blogs, are not entirely uncommon in other domains (see 11). That English loan-verbs can appear in these languages without needing a licensing light verb and can take the host language's verb morphology provides clear empirical evidence against Amritavalli's (2017) analysis which rules out this possibility on the grounds of apparent mismatch between the verbal functional structures of English and Hindi & Marathi.

(9)	*mi*	*ʃodʰ-ət*	*hoto*	*kuɳi*	*mə-la*	***tag****-le*	*ahe*	*ka*[5]
	I	find-IMPRF	was	anyone	1-dat	tag-PRF	be.PRS	Q

'I was trying to check if anyone had tagged me.'

(10)	*kəwite-la*	*prətisad*	*mʰəɳun*	*dusri kəwita*	***post****-u*	*nəye*[6]
	poem-ACC	response as		second poem	post-INF	NEG.AUX

'One shouldn't post a(nother) poem in response to one.'

(11)	*any time mood-wa ko*	***upset****-ao*	*nəhĩ*	*mu:ra*[7]	
	any time mood-CLF DAT	upset-IMP	NEG.AUX	naïve/innocent	

'O innocent one, don't get upset.' (lit. Don't let your mood get upset)

(12)	*məntri*	*d͡ʒi*	*d͡zəra*	*aram se*	*itna*	*kahe*	***frəst-****iya*
	minister	HON	little	easy with	this much	why	frustrate-INF
	rəhe hɛ[8]						
	be.prog be.prs						

[3] https://support.google.com/youtube/answer/13202713?hl=mr

[4] https://www-quora-com.translate.goog/Is-it-good-to-post-everything-on-social-media?_x_tr_sl=en&_x_tr_tl=mr&_x_tr_hl=mr&_x_tr_pto=tc

[5] https://manaatale.wordpress.com/2009/12/25/टॅगला/

[6] https://www.maayboli.com/node/35244

[7] From Hindi movie Gangs of Wasseypur 2

[8] https://twitter.com/kislay_official/status/1651923507641749504

'Take it easy, Minister sir. Why are you getting so frustrated?'

(13) *ɛsa* *hi* *kəl* *həm* *bʰi* ***speak****-enge*
like this EMPH tomorrow 1.HON too speak-will
'I will also speak like that tomorrow.' (Poonam 2020: 140)

See also (14), where a Hindi verb *cun(ə)na* 'choose' can appear in English with English verb morphology, and without needing any overt verbalization or licensing.

(14) my wife will be ***cun****-ing some Sarees.*
choose
'My wife will be choosing some Sarees.' (Bhatia & Ritchie 2016: 11)

Given that English verbs can be directly accommodated in Hindi (11-13) and Hindi verbs can be accommodated directly in Indian English (14), it establishes that direct integration of loan-verbs can occur in either or both languages that are in a contact situation.

2.2 Direct integration of verbal borrowings within Indian languages

In this subsection, I present some cases of direct integration of verbal borrowings within some Indian language pairs.

In a variety of Hindi spoken in Mumbai often referred to as Bambaiya Hindi, certain Marathi verbs like *wapərɳe* 'use' and *pərwəḍɳe* 'afford' seem to have been borrowed directly (15-16).

(15) *zyada kər ke* *log* *yahi* ***wapər-****te* *hɛ*
most people this use-IMPRF be.PRS
'Most people also prefer to use this one.' (from 1982 Hindi film '*Angoor*')[9]

(16) *company ke* *bosses* *ko* ***pərwəḍ****-ega* *nəhĩ*[10]
company of bosses DAT afford-FUT NEG.AUX
'The company leadership won't be able to afford that.'

Miranda (1977: 262) points out that Konkani has borrowed numerous verbs from Dravidian languages such as Kannada and Tulu:

(17) Konkani borrowings from Dravidian
a) Konkani *kuʈʈi* from Kannada *kuʈʈu* 'knock'
b) Konkani *təḷḷi* from Kannada/Tulu *təḷḷu* 'push away'
c) Konkani *ɔppa* from Kannada *oppu* 'agree'

[9] The full dialogue and its translation for context:
A: ye rassi kitne ki hai? [how much is this rope for?]
B: kya karni hai? [what do you want it for?]
A: khudkushi karni hai [suicide]
B: thehro, dusri deta hun [wait, I'll give you another one]
mazboot bhi hai, sasti bhi hai [it is strong and cheap]
zyada kar ke log yahi ***wapar****-te hai* [and most people also prefer to use this one]

[10] https://twitter.com/JayshreePT/status/1459941152660996100

Note that the changes seen in the Konkani verb forms in (17a-c) are to accommodate the Konkani pattern where intransitive verbs are '-a' ending and transitive verbs are '-i' ending. That is, the changes seen Kannada or Tulu verbs borrowed into Kannada are not motivated by morphosyntax but rather are phonological in nature.

2.3 Some more cases of direct integration of loan-verbs

In this subsection, I present some more instances of direct integration of loan-verbs.

Certain Farsi verbs have made their way into the lexicon of many Indo-Aryan languages and can occur in these languages with the host language's verb morphology, e.g. Farsi *xəridœn* 'buy' or have been borrowed into Hindi as *kʰəri:d(ə)na*. Madrikar (1954: 207) points out that Marathi verbs such as *nawad͡z(ə)ɳe* 'be famous', *fərmaw(ə)ɳe* 'order', and *bəd͡ʒaw(ə)ɳe* 'warn' have been borrowed from their Farsi counterparts. There are also numerous other cases of Farsi verbal borrowings in these Indian languages where it can be slightly tricky to argue for direct integration to have taken place since both – the Persian verb and their borrowed counterparts – almost obligatorily occur with a light verb in the form of a N+V construction, e.g. *xærd͡ʒ kərdan* as *kʰərt͡sə kərɳe* 'spend' (lit. spend do), *xərid kərdan* as *kʰəredi kərɳe* 'buy' (lit. buy do) in Marathi.[11,12]

Several other cases of direct integration of loan-verbs have also been attested in the broader typological literature on loan-verb integration:

(18) Tukano (Tukanoan) loan-verb *yuu* 'wait' in Hip (Maku)
ʔam-ăn *ʔãh* ***yu**-té-h*
you-ABS I wait-FUT-DECL
'I will wait for you.' (Wichmann & Wohlgemuth 2008: 100)

(19) French loan-verb *gonfler* 'swell' in Figuig Berber (Berber)
*i-**gõfla***
3.SG.M-be.swollen
'He is swollen up.' (Wichmann & Wohlgemuth 2008: 100)

This shows that direct integration of loan-verbs is common even outside of languages and language families of India and is found across languages of the world. That is, direct integration of borrowed verbs is an accommodation strategy used far and wide in the languages of the world and is not a marginal phenomenon in any sense.

2.4 A note on the usage of loan-verbs

Before delving into the issues that arise from data presented in §2.1-2.3 for a licensing analysis, I would like to provide a brief note about the usage of directly integrated English loan-verbs in Indian languages, focusing largely on Marathi.

Direct integration of English loan-verbs in Marathi as an accommodation strategy is

[11] Wichmann & Wohglemuth (2008: 107) also point out the same and affirm that both direct insertion and light verb strategy are common when it comes to Farsi borrowings in (Hindi-)Urdu.

[12] For discussion regarding the influence of Farsi N+V constructions on Hindi and Marathi and the similarities between them, see Hook & Pardeshi (2009) and chapter 2 of Kulkarni (2017).

both well-attested and well-accepted within its sphere of influence. That is, though extended to a small set of verbs[13] and attested only in certain domains on a regular basis – largely in Marathi writings on social media, blogs, and such related platforms – direct integration of English loan-verbs in Marathi is omnipresent in these domains. Contrary to Poonam (2020: 119) who opines the usage of English loan-verbs in the Hindi Twitter discourse to be for humorous or amusement purposes, usage of directly integrated English loan-verbs in the domain of Marathi blogging is rather communicative in nature and does not come across as 'marked'.

Dharurkar's (2019: 251) observations regarding the pragmatics and aesthetics of English borrowings in Indian languages are worth pointing out in this regard. He notes that '… the recent large-scale English borrowings that happen in Indian languages are a result of the changing native-sensibility or native-understanding of the usage labels that reflect attitudes of the speakers', and points out that '… English words stand for an informal, associating, communicative attitude' rather than '… erudition, literacy, and education' (Dharurkar 2019: 249) as it was the case for speakers of previous generation(s). [14]

2.5 Issues with a licensing analysis

Despite Amritavalli's (2017) licensing analysis being seemingly adequate to explain the data presented in §1, the patterns of borrowings presented in §2.1-2.3 present some puzzles for the same. This section outlines why that is the case.

2.5.1 Multiple strategies of loan-verb integration

That English loan-verbs can either appear with the help of a supporting light verb or can be directly integrated in the host language (9-13) indicates that certain languages can employ more than one strategy to accommodate verbal borrowings. A number of other languages also permit multiple strategies of loan-verb integration. For example, Nepali (Indo-Aryan) loan-verbs like *hai* 'call' and *bolai* 'call' in Manange (Tibeto-Burman) can appear either with a supporting light verb (20a) and with a verbal suffix (20b) respectively.

(20) Nepali loan-verbs in Manange

a) ***[1]hai*** *[1]la-pa*
yawn do-NOM
'to yawn' (Wichmann & Wohlgemuth 2008: 93)

b) ***bolai****-ti* *[1]mi* *ro*
call-SUFFIX EVID REP
'He called (for the frog)' (Wichmann & Wohlgemuth 2008: 97)

In such cases, there is not necessarily any correlation between a particular loan-verb and

[13] Most notable among these being 'tag', 'post', 'paste', 'type', 'google', among others.

[14] However, it is important to keep in mind that such usages can have different acceptability status and usage frequencies in different varieties of the same language. For example, Sakshi Bhatia (p.c.) points out that direct integration English loan verbs in Hindi (such as in examples 11-13) is much more frequent in eastern varieties of Hindi than its standard counterpart.

the strategy being used for its accommodation, and a single loan-verb can be accommodated using more than one strategy: e.g. English verbs like 'click' can appear in Spanish either as *clicar* or as *hacer clic*. Such patterns of loan-verb integration are incompatible with Amritavalli's (2017) licensing analysis.

2.5.2 Locus of borrowability of loan-verbs

Amritavalli (2017: 11) assumes the verbal functional structure to be 'the locus of parametric variation'. This idea, however, is not without its problems. First, it is unclear what the nature of the 'verbal functional structure' that Amritavalli refers to is, and whether it is a genealogical feature pertaining to a specific language family or sub-family, an areal phenomenon, or a wider typological feature. Secondly, there is no clear evidence in the broader typological literature (Moravcsik 1975, Wichmann & Wohlgemuth 2008) that could support either of these alternatives. As Moravcsik (1975) points out:

> 'The set of languages whose verb borrowing patterns have been illustrated above is clearly not characterizable as a genetic group or as a group of spatially adjacent languages; there is similarly no obvious typological property, either, that would define this group.' (Moravcsik 1975: 16)

Decades later, the typological accounts of loan-verb integration still concur with this opinion:

> '… the choice in a given language of one of the four major loan-verb-accommodation patterns cannot be predicted absolutely from structural properties of the languages involved.' (Wichmann & Wohlgemuth 2008: 108)

Given that patterns of direct integration of verbs transcend boundaries of language families or sub-families, geographical or linguistic areas, as well as properties of donor and recipient languages, it is imperative to seek an explanation of this phenomenon elsewhere. I return to this issue in §3.

2.5.3 Issue with optionality of direct integration of loan-verbs

The final issue faced by Amritavalli's (2017) analysis too is concerned with the verbal functional structure of languages. Under her analysis, the ability of a loan-verb to take the host language's inflection indicates the presence of a common verbal functional structure for the concerned languages. However, extending this analysis to the data from §2.1-2.3 leads to some contradictory results. At the end of §1, we already established that language pairs such as Hindi & Bangla and English & American Norwegian must have identical verbal functional structures given that borrowing of a verb within the given pairs result in direct borrowing without the need for verbal licensing. However, as seen in §2.1-2.2, at least some English verbs can directly appear in languages such as Hindi, and Marathi without needing a light verb (9-13); whereas some Marathi verbs have been borrowed into (Bambaiya) Hindi (15-16). Given the licensing analysis, we must then assume that there is a verbal functional structure common to these languages. However, if that is indeed the case, then why does the 'default' strategy of accommodating the English verbal borrowings

in Indian languages involve introducing a supporting light verb? It is also noteworthy that for all cases where direct integration of English verbs into Indian languages is possible, the direct integration of loan-verbs is entirely optional; and that each case of direct integration of a loan-verb can be substituted by a corresponding bilingual compound verb (9-10). However, Amritavalli's (2017) analysis does not allow us to entertain this duality where a licensing light verb is needed for some cases of borrowing but is rather optional for others. This raises the following question: what dictates the need, and more importantly the optionality of verbal licensing? I will address this issue in §3.

3 Accounting for optionality of direct integration of English loan-verbs

It turns out that key to the issues of the locus of borrowability and optionality of direct integration lies in the intensity of language contact and the degree of bilingualism of the speakers.

3.1 Borrowability of verbs and typology of loan-verb integration

Given that certain word classes or morpheme types can be more easily borrowed than others, several works have tried to capture this ease of borrowability leading to postulations of hierarchies of borrowing (Moravcsik 1978) and borrowing scale (Thomason & Kaufmann 1988), among others. According to Matras (2011: 204-205), that a particular word class or morpheme type is easier to borrow simply means that its '… borrowing will occur earlier in the history of contact and hence that it will require less intensive contact'. Works as early as Moravcsik (1975) have discussed the general difficulty of borrowing verbs where she shows that borrowed verbs are always accommodated in the host language by undergoing at least some mechanism of verbal derivation native to the host language. This begs the question as to why is borrowing of verbs difficult and why do verbs, once borrowed, require so many efforts to be integrated into the host language. Matras (2007: 47) thinks that their borrowing is '… made cumbersome in some languages due to the widespread tendency of verbs to be morphologically complex'. According to him:

> '… the difficulty lies in the conceptual complexity of the verb, and the fact that when borrowed and integrated, the verb is expected to perform two operations: the first is to serve as a referential lexical item – a context word, not dissimilar to a noun, adjective, or descriptive adverb. The second is to initiate the predication and so to serve as the principal anchor point for the entire proposition of the utterance. This latter function constitutes its verbness.' (Matras 2007: 48-49)

Under this view, a borrowed verb can serve its referential function immediately upon entering the host language's system. This, however,

> '… is not always sufficient in order to assume the role of predication-initiator. A great number of languages therefore require this additional, crucial function to be explicitly marked out in the verbal expression: in other words, they need to transform the strictly "lexical" depiction of an action/event into

a predicate.' (Matras 2007: 49)

Establishing why verbs are comparatively harder to borrow still leaves us with the following question: how do we make sense of the variation in the morphosyntactic means employed by languages (or by a single language in some cases) to accommodate a borrowed verb? Wichmann & Wohlgemuth (2008: 108) advance an explanation based on intensity of contact and degree of bilingualism of the speakers; and propose that '… if a language has different patterns, these could correlate with the degrees to which speakers of the target language are exposed to the source language(s)'. Taking this idea one step further, they propose the following:

> 'The degree to which a loan verb is integrated into the target language may be considered inversely proportional to the amount of formal mechanics expended by the target language on accommodating the loan verb.' (Wichmann & Wohlgemuth 2008: 109)

That is, the higher the proficiency of a speaker is in the concerned languages, lesser would be the morphosyntactic means that they need to employ in accommodating a loan-verb. In turn, usage of a particular accommodation strategy is also indicative of the speaker's proficiency in the language's concerned. This allowed Wichmann & Wohlgemuth (2008: 109) to place the strategies of loan-verb integration on a 'loan-verb integration hierarchy':

(21) Light verb strategy < indirect insertion < direct insertion < paradigm transfer

Such a hierarchy – proposed 'as an idea to be tested in future research' – not only provides a window into understanding the possible nature and intensity of the contact situation and/or the degree of bilingualism of speakers involved, but also allows us to make predictions about the same. As Wichmann & Wohlgemuth point out, '… if a language already has a strategy and changes this or adds another one, then the new strategy's placement in the hierarchy relative to the earlier strategy would be determined by the relative degree of bilingualism in the source language or languages.' This is evident from the cases of English loan-verbs in Marathi and Hindi where in addition to the light-verb strategy – which has been the default – at least some speakers can and do allow direct integration of English verbs in these languages. This isn't surprising given that each generation of Indian language speakers has an increased exposure to English and could said to be more proficient in English than the previous one – including cases where younger generations from Indian metro cities are being brought up bi/multilingual with English being one of the languages.

Thus, characterizing the different strategies of loan-verb integration as correlates of their degree of bilingualism provides an explanation for the presence of one or more strategies of loan-verb integration as well for some strategies of loan-verb integration being entirely optional.

3.2 Strategies of loan-verb integration and Indian languages

In this subsection, I will briefly review each of the loan-verb integration strategies, with a focus on their attestation in Indian languages.

In the 'light verb strategy', the loan-verb is accommodated using a light verb which

carries the necessary verbal morphology. This is perhaps the default strategy for Indian languages when adapting English loan-verbs, as seen in Kannada (2), Bangla (3), Marathi (7-8), as well as Manange (20a) spoken in neighbouring Nepal.

In 'indirect insertion strategy', the loan-verb is accommodated using an affix. Apart from Nepali loan-verbs in Manange (seen in 20b), certain cases of Hindi verbs appearing in what Bhatia (1989) calls 'Filmi English' seem to be cases like indirect insertion, where the addition of morpheme '*-o-*' to a Hindi verb *g^{h}erna* 'encircle' allows the resulting stem to appear with English past tense morphology. Bhatia thus calls '*-o-*' stem forming vowel.[15]

(22) He was ***g^{h}era***-o-ed by more girls than he could handle.
encircle-o-PST
'He was encircled by more girls than he could handle.' (Bhatia 1989: 269)

The indirect insertion strategy is similar to the light verb strategy in the sense that in both the strategies, a borrowed verb needs to undergo licensing but differs in terms of the licensing element involved. In the light verb strategy, it is the light verb which acts as a licensing verbalizer; whereas the loan-verb is verbalized by an affix in the indirect insertion strategy. Since a loan-verb still needs to be licensed in the indirect insertion strategy, it should not appear without the verbalizing affix. This prediction is well borne out as Bhatia points out that the borrowed Hindi verbs cannot appear in English without the '*-o-*' morpheme which acts as a verbalizer (23).

(23) *He was ***g^{h}era***-ed by more girls than he could handle.
encircle-PST
(Intended) 'He was encircled by more girls than he could handle.' (Bhatia 1989: 269)

The 'direct insertion' strategy (seen in 4, 6, 9-19) refers to the cases where a loan-verb can take the host language's inflectional morphology without having to undergo any overt morphosyntactic changes.

When '… the loan verb is not adapted to the recipient language's morphology at all but is borrowed along with significant parts of the donor language's verbal morphology which maintains its function' (Wichmann & Wohlgemuth 2008: 102), it constitutes a case of 'paradigm transfer'. Such scenarios are much rarer than other strategies of loan-verb integration and are seen only in intensive contact situations. In the Indian context, certain Kannada borrowings in Konkani pointed out by Miranda (1977: 263) constitute cases of paradigm transfer. He points out that Kannada employes suffix '*-isu*' to 'foreign' nouns to derive denominal verbs. Many of such denominal verbs have been borrowed into Konkani along with the '*-isu*' suffix which maintains its denominative status in Konkani.

(24) Paradigm transfer in Kannada-Konkani borrowings
a) Sanskrit *ad^{h}ar* (assistance) → Kannada *ad^{h}arisu* → Konkani *ədərʃi* (help)
b) Sanskrit *upəyogə* (utilization) → Kannada *upəyogisu* → Konkani *upyogʃi* (use)

Thus, it is evident that all four strategies of loan-verb integration are attested in Indian

[15] For arguments against '*-o-*' being introduced for purely phonological reasons, see Bhatia (1989: 271)

languages, albeit to differing extents. The light verb strategy – associated with least proficiency of speakers – has been the 'default' one when it comes to borrowing English verbs in Indian languages. There is also comparatively recent tendency, however, of English loan-verbs getting directly integrated in some Indian languages, albeit in limited domains which likely reflects the usage of highly proficient bilinguals (as discussed briefly in §3.1). Cases of indirect insertion are attested in Nepali loan-verbs in Manange – where the two languages have been in long-term contact and many Manange speakers have been educated with Nepali being the medium of instructions[16]; as well as Hindi verbs appearing in Filmi English – a variety spoken by highly proficient Hindi-English bilinguals. Lastly, paradigm transfer is observed in Kannada loan-verbs in Konkani, where the two languages have been in an intense contact situation for centuries.

Each of these instances thus seem consistent with Wichmann & Wohlgemuth's proposal that the usage of strategies roughly correlate with the degree of bilingualism of speakers, and in turn, the intensity of language contact; and thus, provide evidence in support of the loan-verb integration hierarchy itself.

3.3 On grammatical category of borrowed verbs

The next question I deal with is of the syntactic category of borrowed verbs. More precisely, I look at whether the borrowed verbs retain their syntactic category once they have been borrowed. The literature on loan-verb typology is not in agreement about whether borrowed verbs retain their 'verb hood' in the host language. Moravcsik (1975) claims that verbs are borrowed as nouns[17], whereas Wichmann and Wohlgemuth (2008) opine that there is no clear evidence of verbs being borrowed as nouns but think that they are borrowed as non-verbs. On the other hand, usage of the term 'bilingual compound verb' in the literature suggests presence of two verbs, each from a different language. Considering the patterns of borrowing and loan-verb integration attested in the previous literature and the novel data presented in this paper, as well as given the role degree of bilingualism of speakers plays in optionality of direct integration of loan-verbs, I would like to propose the following:

> During the process of borrowing, speakers with a high degree of bilingualism retain the information about the syntactic identity of the borrowed item, irrespective of whether its syntactic category is preserved in the recipient language. That is, a verb, when borrowed, is borrowed with an understanding that it is a verb, even though it may not function like one in the recipient language immediately upon entering its system.

The evidence in support of this comes from various counts:

Annamalai (1989: 50-51) points out that balanced and imbalanced Tamil-English bilinguals differ in terms of how they borrow English verbs. In Tamil, verbs can be formed from nouns by the addition of verb *pəɳɳu* 'do', as in *kəlyaɳəm pəɳɳu* 'marry' (lit. 'marriage

[16] For more on Nepali-Manange contact, see Hildebrandt 2009

[17] '… the class of borrowed constituents in a language does not include lexically homolingual constituents that are verbs in both languages' (Moravcsik 1975: 4)

do'). In case of English borrowings in Tamil, imbalanced bilinguals follow the Tamil pattern and use English nominals with *pəɳɳu* to form a verb (25a), whereas balanced bilinguals use the English verbal forms to do the same (25b).

(25) English borrowings in Tamil

a) *əvən* *enne* ***confusion***-*pəɳɳiʈʈan*
he me confusion-did

b) *əvən* *enne* ***confuse***-*pəɳɳiʈʈan*
he me confuse-did

'He confused me.'

A case parallel to the one above is of how verbal and nominal borrowings behave once borrowed into the host language. If we assume that verbs are borrowed as nouns, then they should be treated on par with the nominal borrowings by the speakers. That is, once borrowed into the host language, borrowed nouns as well as verbs should behave the same way, and should be subjected to similar morphosyntactic processes. This prediction is not borne out, for direct integration of a borrowed English verb is possible in Marathi (26b), but an English noun cannot take the host language's verb morphology (26a). Similarly, borrowed English verbs cannot take host language's nominal morphology, only borrowed nouns can.

(26) Direct integration borrowed English verbs and nouns in Marathi

a) *maʤʰə* ***confusion*** *ʤʰa-lə / ***confusion**-lə*
I.GEN confusion be-PRF / confusion-PRF

b) *mi* ***confuse*** *ʤʰa-lo / **confuse**-lo*
I confuse be-PRF / confuse-PRF

'I got confused.'

Another argument in support of this comes from instances of bilingual children's accidental or inadvertent language usage. Bilingual children, including cases where they have been exposed to more than one language but aren't necessarily equally proficient in both, often accidentally infuse verbs from one language in the sentence from another along with the host language's morphology (27a-c).

(27) Inadvertent language mixing by children

a) *to* *mə-la* ***gʰur***-*toy* (Hindi verb *gʰurna* 'stare' in Marathi)
he 1-DAT stare-PROG
'He is staring at me.' (Chinmay Dharurkar, p.c.)

b) *mi* *ata* ***bʰag***-*te* (Hindi verb *bʰagna* 'run' in Marathi)
I now run-IMPRF
'I will run now.'

c) I am ***zop***-ing (Marathi verb *zopɳe* 'sleep' in English)
sleep
'I am sleeping.'

Though such cases where bilinguals accidentally 'misuse' words from one language by

using them in another aren't exactly surprising[18], what is remarkable here is the children's ability to insert the verb stem from one language into the frame of another while the inserted verb takes on the verbal inflections of the host language – something that only highly proficient bilinguals can do. This perhaps reflects children's implicit knowledge of syntactic categories of lexical items in both languages, and hints at such knowledge being acquired rather early on. Milton & Donzelli (2013: 443), for example, mention that certain theories of (early) second language acquisition make a distinction between forms – which contain information about morphophonogical forms of words, and lemmas – which contain information about the meaning and syntactic categories of words, and that the L2 learners grasp the idea of lemma very early on in their learning process. Haznedar & Garuseva (2013: 346) also mention 'Lexical Learning Hypothesis' which predicts the knowledge of lexical categories to be acquired early on[19].

This provides evidence in support of the idea that bilinguals with a higher degree of proficiency in both languages retain the information about the syntactic category of the lexical item they are borrowing.

4 Summary

In this paper, I have a provided an overview of loan-verb integration strategies attested in Indian languages, with a focus on English verbal borrowings in Hindi and Marathi. Such borrowings can either be accommodated using a supporting light verb or be directly integrated to appear with the host language's morphology. This optionality of loan-verb integration cannot be adequately explained by syntactic analyses of verbal borrowings such as the one proposed by Amritavalli (2017) which posits common verbal functional structure to be a prerequisite to the direct integration of verbs. Following Matras (2007, 2011) and Wichmann et. al. (2008), I argue that it is the degree of bilingualism of speakers which makes direct integration of loan verbs into the target language possible. Further, I propose that proficient bilinguals retain the information about the syntactic identity of the borrowed elements which allows for their direct integration in the host language. In doing so, I provide evidence from verbal borrowings in Indian languages in support of Wichmann & Wohlgemuth's (2008) loan-verb integration hierarchy which posits a causal relationship between the degree of bilingualism of speakers and morphosyntactic complexity of loan-verb integration strategy used.

Acknowledgements

I am thankful to Chinmay Dharurkar and Ruta Paradkar for discussions on this project, and to Sander Nederveen for his detailed feedback on the paper. I am also thankful to reviewers

[18] See for example, Bialystok (2003: 108-109) on accidental/inadvertent 'mixing' by bilingual children: '… children's linguistic representations are organized according to the two languages, but terms and structures from the other language are selected if they are required to fill a gap in the language the child is attempting to use. Just as young monolinguals will use a close but incorrect label for an object they want to talk about, so too will bilingual children use an item from their other language when it is necessary to express their current meaning.'

[19] As opposed to functional categories which are acquired later sequentially.

and attendees of (f)ASAL-14 at Stony Brook, as well as reviewers of SLE-2020 and GlOW-43 for their comments and suggestions. All remaining errors are my own.

Contact: aaditya.kulkarni.94@gmail.com

References

Amritavalli, R. 2017. Application *diya*, apply *kiya*: borrowings and the light verb. *Indian Linguistics,* 78:1-2, pp. 9-23.

Annamalai, E. 1989. The language factor in code mixing. *International Journal of the Sociology of Language*, 75, pp. 47 – 54. https://doi.org/10.1515/ijsl.1989.75.47

Bhatia, T. 1989. Bilinguals' creativity and syntactic theory: evidence for emerging grammar. *World Englishes,* 8:3, pp. 265-276. https://doi.org/10.1111/j.1467-971X.1989.tb00668.x

Bhatia, T. & W. Ritchie. 2016. Multilingual Language Mixing and Creativity. *Languages, 1(1),* 6; https://doi.org/10.3390/languages1010006

Bialystok, E. 2003. *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press. https://doi.org/10.1017/CBO9780511605963

Dharurkar, C. 2019. Borrowed words make it *informal*! The pragmatics and aesthetics of code-mixing of English in Indian languages. In A. Rath, C. Chatterjee, & S. Ganapathy (Eds.), *Critical Essays in Literature, Language, & Aesthetics: A Volume in Honour of Milind Malshe*. Cambridge Scholars Publishing, pp. 249-261.

Haznedar, B. & Gavruseva, E. (2013). Childhood second language acquisition. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge Handbook of Second Language Acquisition*. Cambridge: Cambridge University Press, pp. 338-352. https://doi.org/10.1017/CBO9781139051729.021

Hildebrandt, K. A. 2009. Loanwords in Manange, a Tibeto-Burman language of Nepal. In M. Haspelmath & U. Tadmoor (Eds.), *Loanwords in the World's Languages*. Berlin: Mouton de Gruyter, pp. 447-470. https://doi.org/10.1515/9783110218442.447

Hook, P. E. & Pardeshi, P. 2009. The semantic evolution of Eat-expressions. In J. Newman (Ed.), *Linguistics of Eating and Drinking*. Amsterdam: John Benjamins, pp. 153-172. https://doi.org/10.1075/tsl.84.08hoo

Kulkarni, A. 2017. Syntax and semantics of HIT-Expressions in Marathi. Unpublished master's dissertation. Deccan College Post-Graduate & Research Institute.

Madrikar, K. S. 1954. A linguistics study of Marathi language in the 17th century with special reference to Dasbodh. Unpublished doctoral dissertation, Savitribai Phule Pune University.

Matras, Y. 2007. Borrowability of structural categories. In: Y. Matras & J. Sakel (Eds.), *Grammatical Borrowing in Cross-Linguistic Perspective*. Berlin, New York: De Gruyter Mouton, pp. 31-74. https://doi.org/10.1515/9783110199192.31

Matras, Y. 2011. Universals of structural borrowing. In: P. Siemund (Ed.), *Linguistic Universals and Language Variation*. Berlin, New York: De Gruyter Mouton, pp. 200-229. https://doi.org/10.1515/9783110238068.204

Milton, J. & Donzelli, G. 2013. The lexicon. Herschensohn & M. Young-Scholten

(Haz.), *The Cambridge Handbook of Second Language Acquisition*. Cambridge: Cambridge University Press, pp. 441-460. https://doi.org/10.1017/CBO9781139051729.027
Miranda, R. 1977. The assimilation of Dravidian loans to Konkani phonological and morphological patterns. Indo-Iranian Journal 19, pp. 247-265. https://doi.org/10.1007/BF00183519
Moravcsik, E. 1975. Verb borrowing. *Wiener Linguistische Gazette*, 8, pp. 3-30.
Moravcsik, E. 1978. Language contact. In: Greenberg, J. (Ed.), C. A. Ferguson, & E. A. Moravcsik (Ass. Eds.), *Universals of Human Language 1*. Stanford: Stanford University Press, pp. 93-122.
Muysken, P. 2000. *Bilingual speech: a typology of code-mixing*. Cambridge University Press.
Muysken, P. 2016. From Colombo to Athens: Areal and Universalist Perspectives on Bilingual Compound Verbs. *Languages, 1(1), 2*. https://doi.org/10.3390/languages1010002
Poonam. 2020. A case study of Hindi-English hybridized verbs on Twitter. Unpublished doctoral dissertation, University of Punjab.
Romaine, S. 1995. *Bilingualism*. Blackwell: Oxford, UK.
Wichmann, S. and Wohlgemuth, J. 2008. Loan verbs in a typological perspective. In: Stolz T., R. S. Palomo & D. Bakker (eds.): *Aspects of Language Contact. New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Processes*. Berlin, New York: Mouton de Gruyter, pp. 89-121. https://doi.org/10.1515/9783110206043.89

On the interaction of aspect and ability in two Hindi/Urdu constructions

PRERNA NADATHUR, *The Ohio State University / New York University*

ABSTRACT

Complex predicates with the Hindi/Urdu light verb *le* ('take') show an unexpected pattern of interpretation in composition with grammatical aspect. Perfective *le* has a completive meaning (Singh 1990), but a *dispositional* (modal) interpretation arises in the imperfective (Butt 1997). This paper pursues a unified analysis of *le*: I compare *le* predicates to uses of the English implicative *manage*, and its aspectual alternation to the *actuality entailments* of the Hindi/Urdu ability modal *sak* (Bhatt 1999). The account builds on prior work (Nadathur 2023a,b) to argue that all three predicates share reference to a complex causal structure, predicting the observed patterns of interpretation in combination with the contrastive semantics of (im)perfective aspects.

1 Introduction

Butt (1997) describes an unexpected **dispositional** reading for certain complex Hindi/Urdu predicates in the imperfective aspect. While the 'simple' imperfective in (1a) indicates that Ila drives habitually, the complex alternative in (2a)—where the main verb is modified by the 'light' auxiliary *le* ('take')—indicates that Ila can (is able to) drive and regularly chooses to exercise this skill. Comparing (2a) to (1a) thus suggests that *le* introduces modal semantics, but this is at odds with its apparent role in (2b): perfectively-marked *le* predicates are typically associated with **completion** (or *culmination*; see, e.g., Singh 1998).

(1) a. Ila gaaṛii calaa-tii (hai).
Ila car drive-IMPF.F (be.PRS)
'Ila (regularly) drives a car.'

b. Ila-ne gaaṛii calaa-yii.
Ila-ERG car drive-PFV.F
'Ila drove a car.'

(2) a. Ila gaaṛii calaa le-tii (hai).
Ila car drive take-IMPF.F (be.PRS).
'Ila (can and) does drive a car.'

b. Ila-ne gaaṛii calaa l-ii.
Ila-ERG car drive take-PFV.F
'Ila drove a car.'

The contrast in (2) is reminiscent of another, more famous interaction between aspect and modality. First described by Bhatt (1999), **actuality entailments** arise when ability modals compose with overt perfectivity: as shown in (3b) for Hindi/Urdu *sak* ('can', 'be able to'), perfective ability entails its prejacent. (3b) contrasts with its imperfective alternative in (3a), which remains compatible with the prejacent's non-realization.

(3) a. Ila gaaṛii calaa sak-tii thii (lekin us-ne gaaṛii nahĩĩ calaa-yii).
Ila car drive can-IMPF.F be.PST.F (but 3SG-ERG car NEG drive-PFV.F)

'Ila could drive a car (but she did not drive a car).' (Ila had the ability to drive)

b. Ila gaaṛii calaa sak-ii (#lekin us-ne gaaṛii nahĩĩ calaa-yii).
Ila car drive can-PFV.F (#but 3SG-ERG car NEG drive-PFV.F)
'Ila managed (was able) to drive a car (#but she did not drive a car).

Although *le* and *sak* constructions differ in their relationship to an embedded predicate—*le* constructions uniformly realize this predicate, while *sak* constructions do so only in the perfective—(2)-(3) show an intriguing parallelism. In both cases, a modal meaning which can be detected under imperfective marking seems to be counteracted by the perfective aspect. This similarity argues against explaining either contrast as the result of lexical ambiguity; such an account is particularly unlikely for *sak* in view of the crosslinguistic prevalence of actuality entailments (see, e.g., Hacquard 2020).

This paper pursues a unified explanation of the effects in (2)-(3). I compare *le* and *sak* to the English **implicative** verb *manage* (Karttunen 1971), drawing on data from Butt (1997) and Bhatt (1999). I argue that the causal semantics proposed for *manage* in Nadathur (2023b) offers a path towards unifying the dispositional and completive uses of *le*, as well as an account of the ability-actuality alternation in (3) (Nadathur 2023a). I propose that *le, sak*, and *manage* share a presuppositional reference to a background structure in which the subject of the complex predication must take some action to *bring about* (causally precipitate) an event in the denotation of the embedded predicate. *Manage* and *le* assert that the causing action was realized, thus licensing inferences to the embedded predication regardless of the aspectual marking. Abilitative *sak*, on the other hand, establishes only its subject's (stative) capacity for realizing the causing action: this produces a 'pure ability' reading in the imperfective, but is systematically reinterpreted in composition with an eventive-selecting perfective operator, leading to the actualized interpretation in (3b).

The paper is organized as follows. §2 provides background on complex *le* predicates, and sketches an account of the dispositional reading. §3 examines the connection between ability and implicativity, arguing that the semantic structure of implicative verbs can also explain the behavior of *sak*. §4 spells out the result of composing implicatively-structured *le* with (im)perfective aspects and discusses some challenges for the proposal. §5 concludes.

2 The dispositional complex predicate

Hindi/Urdu has a rich system of complex predicates which combine a bare main verb with an inflected 'light' auxiliary (Hook 1974). **Light verbs** (LVs) come from a restricted set of lexical verbs (Table 1), but their semantic contribution to the complex predicate is bleached by comparison to 'heavy' uses. LVs affect the interpretation of a complex predicate in a variety of ways: some add (in)volitionality entailments (see 4), while others perform operations like passivization or permissive causativization (Butt 1993). LVs can also introduce *aspectual* content: the *paṛ* ('fall') predicate in (4a) focuses on the inception of a singing

event, but the *le* predicate in (4b) emphasizes completion (Singh 1990, 1998).[1]

Based on (di)transitives	Based on intransitives
le ('take')	*aa* ('come')
de ('give')	*jaa* ('go')
ḍaal ('put')	*paṛ* ('fall')

Table 1: A non-exhaustive list of Hindi/Urdu light verbs (Butt 1993)

(4) a. Ila gaanaa gaa paṛ-ii
Ila song sing fall-PFV.F
'Ila burst out in song.'
(spontaneously, involuntarily)

b. Ila-ne gaanaa gaa li-yaa
Ila-ERG song sing take-PFV.M
'Ila sang a song.'
(fully, deliberately)

The link between *le* and culmination is particularly clear in composition with **telic** predicates: the complex perfective in (5a) licenses a *culmination entailment* which contrasts with the weaker reading of the 'simple' perfective in (5b) (Arunachalam & Kothari 2011).

(5) a. Maayaa-ne biskaṭ khaa li-yaa #lekin use puuraa nahĩĩ khaa-yaa.
Maya-ERG cookie eat take-PFV.M but it.ACC whole NEG eat-PFV.M
'Maya ate the cookie, #but did not finish it.'

b. Maayaa-ne biskaṭ khaa-yaa lekin use puuraa nahĩĩ khaa-yaa.
Maya-ERG cookie eat-PFV.M but it.ACC whole NEG eat-PFV.M
'Maya ate the cookie but did not finish it.'

The dispositional reading in (2a) challenges existing analyses of *le* as an essentially aspectual LV (Singh 1998; Butt 1993, a.o.). The core aspectual contrast in Hindi/Urdu is between a habitual imperfective and an episodic (terminating but non-culminating) perfective (see 1). Thus, if *le* adds the semantics of culmination, as (5a) suggests, we predict some combination of habituality and culmination from imperfective *le* claims, but there is no obvious source for the modal component of the observed dispositional reading.

Butt (1997) likens the dispositional complex predicate (DCP) to *existentially-interpreted* English generics (Lawler 1973): like (6), the DCP indicates that its subject has some property which enables realization of the embedded predicate, and moreover chooses to exercise this ability on a regular basis. Butt emphasizes that *regularity* should be understood in a conditional sense: the ability is exercised under some contextually-relevant set of circumstances (i.e., when necessary, but not necessarily at all conceivable opportunities; see 6).

[1] Intransitive-derived LVs often indicate spontaneity, while (di)transitive-derived alternatives require ergative case—here a marker of volition (Mohanan 1990, a.o.)—and indicate the subject's *conscious choice*.

(6) My pet toad will eat flies. (Lawler 1973)
∼ My pet toad can and does eat flies under the right circumstances (but not necessarily in all eating situations, and not necessarily to the exclusion of other foods).

The dispositional meaning is crucially accompanied by a non-modal entailment: (2a) cannot be coherently followed by the claim that Ila does not drive at all. This differentiates the DCP from imperfective uses of the ability modal *sak* (e.g., 3a) and renders the former especially suitable as a counter to negative expectations. Example (7), for instance, is well-suited to a context in which Ila's ability to drive is in question (perhaps because the addressee has never seen her drive; R. Bhatt, p.c.); this example also highlights the qualified (conditional) nature of the regularity associated with the DCP.

(7) climate change-kii vajah-se vo aaj-kal gaaṛii nahĩĩ calaa
climate change-GEN reason-INST 3SG.NOM today-tomorrow car NEG drive
rahii hai, lekin bilkul vo gaaṛii calaa le-tii hai.
PROG.F be.PRS, but certainly 3SG.NOM car drive take-IMPF.F be.PRS.
'Due to climate change, she's not in the habit of driving these days, but she certainly (can and) does drive.'

Finally, Butt notes that the actualization contrast between the DCP and imperfective *sak* (see 3a) affects their relative appropriateness in conditional constructions. (8a) describes what Ila *will* do if she encounters a good road, while the oddness of (8b) is due to the suggestion that it is only in these circumstances that she will develop the ability to cycle.[2]

(8) a. agar raastaa pakkaa ho, Ila saikal calaa le-gii.
if road correct be.INF, Ila cycle drive take-FUT.F
'If the road is good, Ila will (choose to) ride a bicycle.'

b. ??agar raastaa pakkaa ho, Ila saikal calaa sak-egi.
if road correct be.INF, Ila cycle drive can-FUT.F
'If the road is good, Ila will be able to ride a bicycle.'

Butt concludes that dispositional *le* warrants a modal semantic treatment. In pursuit of a satisfactory analysis, she suggests a connection to (certain uses) of the Sinhala **involitive**, which is analyzed by Inman (1993) as introducing a 'happenstantial' modality.

2.1 Happenstance: Insights from Sinhala

Sinhala verbs alternate between a default *volitive* form and a morphologically marked **involitive** form. The volitive in (9a) is typically used to describe intentional acts, and the involitive (9b) is associated with accidentality, but Inman (1993, pp.102–104) argues that the contrast between the forms cannot be about (lexically-specified) volition, since involitive claims asymmetrically entail their volitive counterparts, as shown in (9c).

[2]The reading in (8a) shows that the dispositional interpretation of *le* is not restricted to imperfective contexts, further motivating a univocal treatment of its LV uses.

(9) a. laməya kooppe binda
child.NOM cup break.PST
'The child broke the cup.'

b. laməya atiŋ kooppe (?hitəla) biňduna.
child ERG cup (?intend.PTCPL) break.INV.PST
'The child (?intentionally) broke the cup.'

c. (9b) → (9a), (9a) ↛ (9b)

In addition to accidentality, the involitive stem has a *dispositional* use, which is exemplified by (10): like the DCP, this interpretation is well-suited to counter-to-expectation contexts. Example (10) is neutral with respect to Mahatun's volitionality, but conveys the speaker's surprise that Mahatun can and does realize the embedded predication: (10) is thus well-paraphrased by the DCP in (11).

(10) Mahatuŋ atiŋ mee kæmə hoňdəṭə hædenəwa
Mahatun ERG this food well make.INV.PRS
'Mahatun makes this food well (unexpectedly).' (Sinhala; Inman 1993, p.100)

(11) Mahatuŋ ye khaanaa acchaa banaa le-taa hai.
Mahatun this food well make take-IMPF.M be.PRS
'Mahatun (can and) does make this food well.' (Hindi/Urdu)

Inman proposes to unify the accidental and dispositional uses of INV in terms of 'happenstantial' modality, which he associates with the semantics of the English implicative *happen (to)*. As (12) shows, the inferential profile of *happen (to)* parallels that of the involitive marker: *x happens to P* entails that *x* does *P*, and can be paraphrased with "can and does", taken together with some indication of countered expectation.

(12) Mahatun happens to make this dish well (#but he does not make it well).
∼ *As it turns out, Mahatun can and does make this dish well.*

Happenstantial modality is formalized as *non-necessity* in (13). Inman treats INV as a propositional operator, requiring that its argument ϕ holds in the evaluation world but not across the entire relevant modal domain. To capture the contrast between accidentality and unexpectedness (in the dispositional reading), Inman suggests that INV alternates between *teleological* modality, where $\mathsf{opt}_{f,g}(w)$ comprises circumstantially-accessible worlds (cr) which are optimal with respect to the subject's intentions (tel), and an *epistemic* flavour, where the optimal worlds are maximally stereotypical (nm) with respect to the speaker's beliefs (ep). The resulting interpretations for (9b) and (10) are paraphrased in (14).

(13) $[\![\text{INV}]\!]^{w,f,g} := \lambda\phi_{st}.\phi(w) \,\&\, \neg\forall w' \in \mathsf{opt}_{f,g}(w)[\phi(w')]$

(14) a. Accidental: $\mathsf{opt}_{f,g}(w) = \mathsf{opt}_{\mathsf{cr},\mathsf{tel}}(w)$
(9b) ∼ *The child broke the cup and there is some world compatible with her intentions and circumstances in which she did not do so.*

b. Dispositional: $\mathsf{opt}_{f,g}(w) = \mathsf{opt}_{\mathsf{ep},\mathsf{nm}}(w)$
(10) ∼ *Mahatun makes this dish well and there is some world compatible with the speaker's beliefs and expectations in which he does not do so.*

2.2 Happenstance and the DCP

Inman's analysis of the dispositional involitive offers a promising first pass at the semantics of the DCP. As spelled out in (15), this proposal captures the entailment from the DCP to its simple alternative, while appropriateness in counter-to-expectation contexts follows (as in 14b) from the second entailment to non-necessity.

(15) $[\![\text{le}]\!]^w := \lambda\phi.\phi(w) \,\&\, \neg\forall w' \in \mathsf{opt}_{\mathsf{ep},\mathsf{nm}}(w)[\phi(w')]$

Nevertheless, (15) falls short on Butt's (1997) desiderata. For one, if *le* is analyzed as a propositional operator, it will not have access to the sentential subject and thus cannot impose any volitionality constraints, meaning that (15) does not capture the sense of *conscious choice* that invariably attaches to *le* predicates—for the DCP, this amounts to the inference that $\phi(w)$ results from the subject's deliberate decision to exercise their ability. (15) also fails to capture the *conditional* nature of the DCP: Butt's own suggestion is that *le* should be analyzed as conditional necessity, with a modal domain containing "the speaker's expectations and the conditions under which the subject [...] will perform the given action" (p.10), but it is not immediately clear how to implement this idea, nor how it should be integrated with the non-necessity that derives counter-to-expectation effects for Inman.

I propose that the happenstantial semantics in (15) can be reconciled with both volitionality and conditionality requirements of the DCP by making a few key modifications. First, a satisfactory account should distinguish the embedded proposition (the dispositional *target*) from a second proposition, corresponding to the subject's *choice* to exercise their ability. Second, I propose to condition the dispositional target on the relevant choice. By treating this choice as contextually *determinative* (necessary and sufficient) for the dispositional target, we can accommodate both Inman's non-necessity and Butt's conditional necessity: the embedded proposition will go unrealized in any accessible world in which the subject does not choose to exercise ability, but is guaranteed whenever a positive choice is made. As long as the modal domain for a complex *le* claim includes worlds in which the subject chooses positively as well as worlds in which the choice is negative, we capture the inference that realizing the dispositional target is a matter of volition.

Finally, motivated by Inman's comparison of INV to *happen (to)*, I suggest that the modal component of the DCP should be treated as projective (*not at-issue*) content. The natural interpretation of a negated *happen (to)* claim is one on which negation targets the non-modal entailment: (16) conveys that the cup was not broken and preserves the intuition that both breaking and non-breaking were possible in context. If *happen (to)* in fact entails non-necessity, we should also expect (16) to have a reading on which the child broke the cup and this outcome was (teleologically or epistemically) necessary. In the absence of clear prosodic focus on *happen*—a device which is independently known to introduce metalinguistic effects (e.g., Beaver & Clark 2008)—this reading does not seem to be available.

(16) The child did not happen to break the cup. → *The child did not break the cup.*

Inman does not discuss the interpretation of negated involitives, so I cannot compare INV with *happen (to)* in this regard. The DCP itself cannot be negated (see §4.2). The

hypothesis that its modality is presupposed thus remains provisional; taking a broader view, however, this move would bring its at-issue contribution closer to that of completive *le* perfectives (such as 2b, 4b, 5a), thereby holding out hope for a univocal analysis.

Example (17) sketches a revised treatment of the DCP. Unlike (15), (17) does not take LV *le* to directly assert the realization of an embedded proposition; this entailment follows instead from the joint effect of a modal presupposition (17a) and the at-issue resolution of the determinative choice in (17b).

(17) Given a one-place predicate P and an agent x, $le(P)(x)$:
   a. *presupposes* that a **(prior) choice** $A(x)$ is **necessary** and **sufficient** for $P(x)$
   b. *asserts* the truth of $A(x)$ (that x realized A)

As we will see, (17) is structurally similar to Nadathur's (2023b) analysis of English *manage*, which—like *happen (to)*—is semantically **implicative** (Karttunen 1971). The similarity is particularly notable given the parallels between LV *le* and abilitative *sak* (see §1), since Bhatt (1999) independently likens the actualized interpretation of *sak* to that of past-tense *manage*. The emerging picture, then, is suggestive of an underlying uniformity in the semantics of implicativity, ability, and the DCP. The next section explores the ability-implicativity link in more detail.

3 From implicativity to ability

Explaining the behaviour of *sak* in (3) requires an account of the crosslinguistic phenomenon of *actuality entailments* (AEs), exemplified in (3b).

(3) a. Ila gaaṛii calaa sak-tii thii (lekin us-ne gaaṛii nahĩĩ calaa-yii).
   Ila car drive can-IMPF.F be.PST.F (but 3SG-ERG car NEG drive-PFV.F)
   'Ila could drive a car (but she did not drive a car).' (Ila had the ability to drive)
   b. Ila gaaṛii calaa sak-ii (#lekin us-ne gaaṛii nahĩĩ calaa-yii).
   Ila car drive can-PFV.F (#but 3SG-ERG car NEG drive-PFV.F)
   'Ila managed (was able) to drive a car (#but she did not drive a car).

AEs resist explanation on standard treatments of aspect and modality. Grammatical aspects are usually treated as providing a particular perspective on a situation by constraining its temporal relationship to a reference time supplied by tense (Kratzer 1998): on this approach, episodic perfectives include the runtime of the target situation in the reference time, as in (18). Within the linguistic literature, ability modals are most frequently analyzed as *circumstantial* possibilities (but see §3.2): in composition with (18), (19) predicts an 'opportunity' reading for (3b), on which the possibility of driving is bounded by the reference time t^*. Nothing requires that Ila acted on her opportunity, so (20) falls short of an AE.

(18) $[\![\text{PFV}]\!] := \lambda w \lambda t \lambda P_{vt}.\exists e[\tau(e) \subseteq t \,\&\, P(e)(w)]$

(19) $[\![\text{can}_{\text{ability}}]\!] := \lambda w \lambda P_{vt} \lambda e.\exists w' \in \mathsf{opt}_{\mathsf{cr},\emptyset}(w)[P(e)(w')]$ (cf. Hacquard 2009)

(20) $[\![\text{Ila gaaṛii calaa sakii}]\!]^{w^*,t^*} = [\![\text{PST}(\text{PFV}(\text{can}_{\text{ability}}(\text{Ila drive a car})))]\!]^{w^*,t^*} = 1$
iff $\exists e[\tau(e) \subseteq t\{\prec t^*\} \,\&\, \exists w \in \text{opt}_{\text{cr},\emptyset}(w^*)[\text{drive-car}(\text{I})(e)(w)]]$

Bhatt (1999) offers an important insight into the nature of the actualizing effect. While much of the subsequent literature treats AEs as cases of modal erasure (with the perfective undoing the modal's contribution), Bhatt points out that actualized ability is better paraphrased by English *manage* than by a simple (non-modal) assertion of the modal's prejacent. Alongside complement entailments (shown for *manage* in 21), actualized ability and *manage* also share a projective inference: (22a)-(22b) both suggest that riding a bicycle was difficult or somehow non-trivial for Ila, regardless of matrix polarity.

(21) Ila managed to drive a car (#but she did not drive a car). (compare to 3b)

(22) a. Ila managed / did not manage to ride a bicycle.
b. Ila saikal (nahĩĩ) calaa sak-ii
Ila cycle (NEG) drive can-PFV.F
'Ila was (not) able to ride a bike.'
(22a), (22b) → *Cycling was effortful/difficult/non-trivial for Ila.*

On the strength of this comparison, Bhatt argues that ability modals share the lexical semantics of *manage*. Unfortunately, however, this approach cannot explain why *sak* and *manage* diverge in non-episodic contexts, with *manage* lacking the non-entailing reading of imperfective *sak* (compare 23 to 3a). This difference cannot simply be attributed to the lack of overt grammatical aspect in English: as (24) shows, the complement entailments of French implicative *réussir* ('manage', 'succeed') are likewise insensitive to aspect.[3]

(23) Ila manages to drive, #but she does not drive.

(24) Ila { réussissait / a réussi } à conduire, #mais elle n'a pas conduit.
Ila { managed-IMPF / -PFV } to drive, #but she NEG-has NEG drive.PP
'Ila managed to drive, #but she did not drive.'

The above data argue against the lexical equivalence of *sak* and *manage*, but do not undermine a weaker version of Bhatt's hypothesis, on which the equivalence between actualized ability and *manage* is analytically derived. Pursuing this idea requires identifying the key semantic components of lexical implicativity, and then determining how these elements may be (re)assembled in the composition of perfective aspect and abilitative modality.

3.1 Causal semantics for *manage*

Implicative verbs are characterized by complement entailments which reverse with matrix negation (so that negating 21 entails that Ila did not drive), paired with a projective inference that blocks full equivalence between the implicative and its complement (Karttunen

[3]Bhatt explains the 'pure ability' reading in (3a) by assuming that IMPF optionally introduces a generic operator which shifts the lexically-specified prejacent entailment to a set of "ideal" worlds (see also Hacquard 2009). Under lexical equivalence, the same derivation should be available for *manage* and *réussir*.

1971). Despite agreement on these points, it has proven difficult to pin down what *manage* projects: *difficulty* or *unexpectedness* are plausible in many cases, but the felicity of naturally-occurring data like (25a)-(25b) shows that these inferences cannot be lexically encoded (Coleman 1975; Baglini & Francez 2016; Nadathur 2023b).

(25) a. By 1998, [...] gun manufacturers had easily managed to bypass the laws [...]
b. The Sozialdemokratiet managed to strengthen their position [...] as expected.

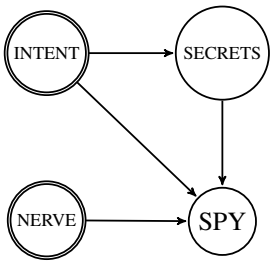
The picture becomes clearer when we consider an implicative like *dare*, which is more specific than *manage* in its projective content. As (26) shows, both positive and negative *dare* indicate that acting bravely (*being daring*) is required in order for the sentential subject to realize the complement; however, whether or not this prerequisite is realized depends on matrix polarity. The polarity of prerequisite inferences therefore aligns with that of *dare*'s complement inferences, with the result that (26a)-(26b) are well-paraphrased by claims which establish a causal relationship between Ria's bravery and Ria's opening of the door.

(26) a. Ria dared to open the door. → *Ria acted bravely*
∼ *Ria opened the door because she acted bravely.*
b. Ria did not dare to open the door. → *Ria did not act bravely*
∼ *Ria did not open the door because she did not act bravely.*
(26a), (26b) → *Opening the door required Ria to act bravely.*

The pattern in (26) is shared by a wide range of implicatives, motivating a templatic account on which these verbs presuppose that some prerequisite action is both **causally necessary** and **sufficient** for their complements (Nadathur 2023b). The prerequisite's realization is settled as at-issue content, deriving the desired pattern of complement entailment when presupposition and assertion are combined. Implicatives differ in what and how much they specify about the causal prerequisite: *manage*, like *dare*, establishes the existence of this prerequisite, but leaves its nature underspecified. As a result, *manage* complements are understood to be non-trivial because they cannot be realized without first satisfying some prerequisite, but whether this results in difficulty, unlikeliness, or something else will depend on what is known about salient causal relationships in the utterance context.

I formalize the causal components of the proposal using **structural equation models** (Pearl SEMs; 2000), treated here as discourse parameters which track contextually-relevant causal information. An SEM corresponds to a *directed acyclic graph* (Figure 1a) whose nodes are (unvalued) propositional variables and whose edges represent an atomic notion of *causal relevance*; the graph is paired with a set of *structural equations* (Figure 1b) that specify how the truth values of 'downstream' (dependent) variables are determined by the values of their immediate causal ancestors. Figure 1 models the toy context in (27).

(27) *Context:* In the infamous Dreyfus affair (1894–1906), Captain Dreyfus was wrongfully accused of sharing French military secrets with the Germans. Assume that:
a. SPYing requires three things: (a) harboring treasonous INTENT, (b) collecting military SECRETS, and (c) taking risks to transmit these secrets (NERVE)



(a)

SPY := INTENT & SECRETS & NERVE

SECRETS := INTENT

(b)

Figure 1: SEM for the Dreyfus context in (27)

b. SECRET collection depends entirely on the presence of treasonous INTENT

Given a situation s (a partial assignment of truth values to variables), we can use an SEM to work out a set of causal consequences (cf. Schulz 2011). For instance, in a hypothetical situation which establishes that Dreyfus has treasonous intent and acts daringly, we infer that he will collect secrets and ultimately spy: Figure 2 illustrates the stepwise causal reasoning which takes us from starting situation 2a to its *maximal causal development* 2c.

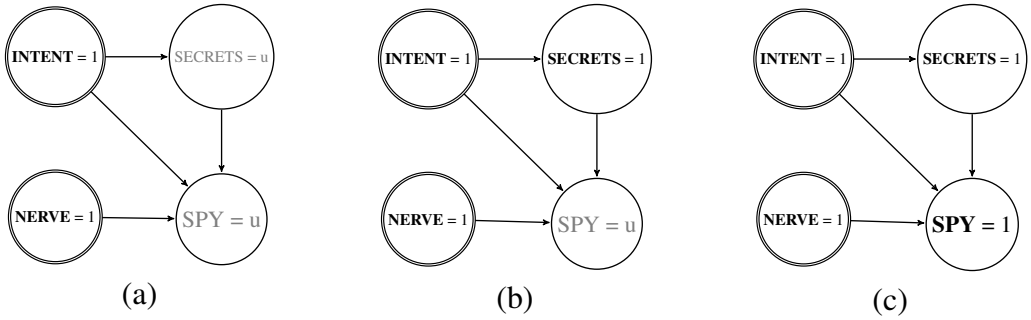


Figure 2: Reasoning with causal models

In this framework, *causal necessity* and *causal sufficiency* are structural relationships that may obtain between a valued variable and a valuation for one of its descendants, relative to a specific situation. The variable-value pair $\langle C, c\rangle$ ($c \in \{0,1\}$) is *causally necessary* for $\langle E, e\rangle$ (where $e \in \{0,1\}$ and E is downstream of C) in any situation s such that all causally-consistent extensions of s which assign value e to E also assign value c to C. $\langle C, c\rangle$ is *causally sufficient* for $\langle E, e\rangle$ in s just in case the situation $s[C \mapsto c]$ (which is identical to s except perhaps at C) assigns value e to E in its maximal causal development.

To illustrate, Figure 3a depicts a situation in which being daring is both causally necessary and sufficient for Dreyfus to spy: the only consistent way to expand this situation into one which makes SPY true requires verifying NERVE, and (since INTENT $= 1$ guarantees that SECRETS $= 1$), adding NERVE $= 1$ causally ensures that SPY $= 1$. This is exactly the right sort of context for implicative *dare*. Which of (28a)-(28b) is accurate depends on how

Dreyfus actually behaves: (28a) asserts that he acted with daring, causally entailing that he spied, while (28b) derives his failure to spy from an asserted lack of daring.

(a) (b)

Figure 3: Two contexts for the Dreyfus scenario

(28) a. Dreyfus dared to spy for the Germans.
b. Dreyfus did not dare to spy for the Germans.

(29) a. Dreyfus managed to spy for the Germans.
b. Dreyfus did not manage to spy for the Germans.

The reality of the affair is better represented by Figure 3b, which establishes Dreyfus's lack of treasonous intent. In this scenario, the causal semantics rightly predict that neither (28a) nor (28b) will be appropriate: *dare* presupposes that acting daringly is causally determinative for spying, but there is no consistent way to extend 3b to a situation which makes SPY true, regardless of the value of NERVE. The lack of a consistent causal pathway from 3b to $\text{SPY} = 1$ also rules out (29a)-(29b): while *manage* does not require the necessity/sufficiency of $\text{NERVE} = 1$ in particular, it does require the existence of some causally determinative condition for $\text{SPY} = 1$, and no such condition exists. Crucially, (29b) is infelicitous despite the contextual truth of its complement entailment—that Dreyfus did not spy. This provides clear support for the proposed causal background: use of an implicative does not simply inform the listener about complement truth, but requires a context in which this complement is both non-trivial and causally realizable, under conditions which may (or may not) be descriptively constrained by the matrix verb.

As spelled out below, the implicative profile of *manage* relies on two things: the presupposition of a causal prerequisite for the complement and an assertion which settles whether or not this prerequisite occurred. Following Kaufmann (2013) (see also Nadathur 2023a, Ch.5), the causal laws encoded in an SEM can be mapped to a causal ordering source (cs), which (when paired with a circumstantial modal base) allows (30) to be expressed in the more standard terms in (31) (where $\text{in}(t,w,\beta_{vt}) \equiv \exists e.\tau(e) \subseteq t \,\&\, \beta(e)(w)$; Nadathur 2023c).

(30) Given a one-place predicate P and an agent x, *manage*$(P)(x)$:
a. *presupposes* that some action $A(x)$ is *causally necessary and sufficient* for $P(x)$
b. *asserts* $A(x)$

(31) $[\![\text{manage}(P)(x)]\!] :=$
$\lambda w \lambda t \lambda e.(\iota A_{evt}.\forall w' \in \mathsf{opt}_{\mathsf{cr},\ \mathsf{cs}}(w,t)[\mathsf{in}(t,w',A(x)) \leftrightarrow \mathsf{in}(t,w',P(x))])(e)(w)$

Modulo the use of causal modality, (30) parallels the modified happenstantial semantics for *le* in (17). I revisit this similarity in §4, after discussing how implicative structure is involved in the interpretation of abilitative *sak*.

3.2 Ability and actuality entailments

Our working hypothesis is that AEs are instances of implicativity, derived via the compositional (re)assembly of the semantic components in (30). If this is correct, then *sak* must be given an analysis which produces the structure in (30)/(31) in combination with the perfective—but not the imperfective—aspect. Such an analysis is given below.

(32) Given a one-place predicate P and an agent x, *sak*$(P)(x)$:
   a. *presupposes* that some action $A(x)$ is causally necessary and sufficient for $P(x)$
   b. *asserts* that $A(x)$ is in x's *choice set* (x can do A)

(33) $[\![\text{sak}(P)(x)]\!] :=$
$\lambda w \lambda t.(\iota A.\forall w' \in \mathsf{opt}_{\mathsf{cr},\mathsf{cs}}[\mathsf{in}(t,w',A(x)) \leftrightarrow \mathsf{in}(t,w',P(x))])(x) \in \mathsf{CH}(x,w,t)$

Like *manage*, *sak* requires a context where some prior action is causally determinative for the embedded proposition. The predicates diverge in assertion: *manage* realizes the complement-causing action, but *sak* establishes only that this action is *available* to x. I capture availability by using the notion of a *choice set* ($\mathsf{CH}(x,w,t)$) comprising possible actions for agent x at world w and time t: including $Q(x)$ in $\mathsf{CH}(x,w,t)$ expresses the possibility that x *chooses* the modal alternative which verifies $Q(x)$ at t ($\forall w,t,x,Q_{evt}[Q(x) \in \mathsf{CH}(x,w,t) \rightarrow \exists w' \in \mathsf{hist}(w)[\mathsf{in}(t,w',Q(x))]]$; (Belnap & Perloff 1988, Nadathur 2023a).

A complex structure for ability can be motivated by comparing the conditions under which abilities, as opposed to circumstantial possibilities, may be attested. For instance, a single (potentially fluky) witness for a proposition $P(x)$ entails the corresponding circumstantial claim, but is not enough to justify ability: the latter seems to require additional evidence that the performance of $P(x)$ can be reliably repeated (Kenny 1976, a.o.).

(34) *Context:* Rookie golfer Tara makes a hole in one on her first game (Maier 2018)
   a. It is possible for Tara to make a hole in one.
   b. ??Taaraa hole-in-one kar sak-tii hai.
      ??Tara hole-in-one do can-IMPF.F be.PRS
      '??Tara has the ability to make a hole in one.' (Hindi/Urdu)

Proposal (32)/(33) explains the effect in (34). Ability modals are structured here as doubly modal, expressing *hypothetical guarantees* (Mandelkern et al. 2017) in which a potential action $A(x)$ acts as a prejacent-ensuring strategy (i.e., a means by which $P(x)$ can reliably be realized). While this basic structure is shared by several existing analyses of ability (including Mandelkern et al.; see also Brown 1988; Louie 2015; Maier 2018, a.o.), (32)/(33)

adds two novel components, enforcing a *causal* link between $A(x)$ and $P(x)$ and strengthening this relationship to one of necessity as well as sufficiency. These modifications capture the sense of non-triviality which typically attaches to claims of ability and bring ability modals into alignment with implicative *manage*, as anticipated by Bhatt (1999).

With (32)/(33) in hand, *manage* and *sak* differ only in their treatment of the causing action $A(x)$: if we are to derive AEs as implicative entailments, perfective marking must convert the assertion in (32b) into the one in (30b), forcing the subject of an ability claim to act on the prejacent-causing choice. As it turns out, there is a good deal of evidence to suggest that this is precisely what the addition of an episodic perfective does.

Manage and *sak* differ in aspectual class: prerequisite-realizing *manage* claims are eventive, but *sak* is at-base stative, assigning a static property to its subject.[4] Ability modals belong, moreover, to a special class of *dynamic capacity* statives, describing properties that hold of individuals in virtue of their propensity for certain kinds of action: the class includes behavioral predicates such as *be fast* and *be loud*. As the French data in (35) show, dynamic capacity statives have a distinctive pattern of aspectual interpretation. Imperfective (35a) describes the potential for speed-characterized action, but perfective (35b) is understood as a claim about action, describing an event in which Juno actually manifested her speed.

(35) a. Juno était rapide.
Juno was-IMPF fast
'Juno was capable of speed.'

b. Juno a été rapide.
Juno was-PFV fast
'Juno did something quickly.'

The pattern in (35) extends to a set of *enough* predicates which bridge the gap between lexical implicativity and ability: (36) attributes a dynamic capacity, and can be paraphrased in abilitative terms, as causally conditioning Juno's ability to win the race on her propensity for speed (Nadathur 2023a,c).[5] In aspect-marking languages like French, these constructions license complement entailments in the pattern of ability modals (Hacquard 2005). Taking the effect in (35b) into account, (37b) appears essentially implicative: where (37a) establishes Juno's capacity for the race-winning speed, perfective in (37b) triggers a performance reading, asserting that Juno ran at the required speed and thereby licensing the observed entailment.

(36) Juno is fast enough to win the race.
$\sim$ *Juno is able to win the race, in view of her (capacity for) speed.*

(37) a. Juno était assez rapide pour gagner la course.
Juno was-IMPF enough fast for win the race.
'Juno was fast enough to win the race.' (acceptable if she did not win)

b. Juno a été assez rapide pour gagner la course.
Juno was-PFV enough fast for win the race.

[4]Homer (2021) provides a number of good empirical arguments for the stativity of ability modals.

[5]Causal *enough* constructions thus bear the same relationship to (standard) ability as prerequisite-specifying implicatives like *dare* bear to (underspecified) *manage*.

'Juno was fast enough to win the race.' (contradictory/false if she did not win)
$\sim$ *Juno ran at the race-winning speed and consequently won the race*

The effects in (35b) and (37b) are instances of a more general pattern of eventivizing **aspectual coercion**, a much-observed effect in which stative predicates are reinterpreted when they occur in episodic (event-selecting) contexts, such as the scope of a perfective operator (de Swart 1998; Bary 2009; Homer 2021; Nadathur 2023a, a.o.). A particularly well-known instance of coercion involves the use of knowledge predicates (e.g., French *savoir*, *connaître*) to describe 'coming to know' (*learning, meeting*) events in the perfective: the effect can be formally derived by inserting an *inchoative* coercion operator (mapping statives to predicates of state-initiating events) between the underlying predicate and the perfective operator. The 'performance' effect in (35b) involves the application of a different form of coercion, variously termed *dynamic* (de Swart 1998), *evidential* (Fernald 1999; Nadathur 2023c), *actualistic* (Homer 2021), or *instantiative* (Nadathur 2023a): whatever it is called, this operation ultimately replaces a dynamic capacity stative with a predicates of actions that manifest (provide evidence for) the underlying capacity.

On the causal analysis in (32)/(33), *sak* attributes a dynamic capacity to its subject: specifically, the capacity for action of a type which will bring about the embedded proposition. In the scope of an episodic perfective, then, *sak* is a candidate for the same performance-inducing form of aspectual coercion which applies in (35b) and (37b). The result, as sketched in (38), is that a claim like (3b) makes the same assertion as the corresponding *manage* claim, indicating here that Ila performed the proximate cause of the ability complement, with the causal consequence in (38d): i.e., the desired actuality entailment.

(38) $[\![\text{Ila gaaṛii calaa sakii}]\!]^{w^*,t^*} = [\![\text{PST}(\text{PFV}(\text{sak}(\text{Ila drive a car})))]\!]^{w^*,t^*}$

a. *Presupposition:* $\exists A : \forall w \in \mathsf{opt}_{\mathsf{cr,cs}}(w^*)[\mathsf{in}(t^*, w, A(\mathsf{I}) \leftrightarrow \mathsf{in}(t^*, w, \mathsf{drive\text{-}car}(\mathsf{I}))]$
Some action for Ila is the determinative (proximate) cause of driving

b. *Base assertion (stative):* $A(\mathsf{I}) \in \mathsf{CH}(\mathsf{I}, w^*, t^*)$
The proximate cause of driving is an immediate option for Ila

c. *After coercion (eventive):* $\mathsf{in}(t^*, w^*, A(\mathsf{I}))$
Ila acted on her capacity for the proximate cause of driving

d. *Causal consequence:* $\mathsf{in}(t^*, w^*, \mathsf{drive\text{-}car}(\mathsf{I}))$
Ila drove a car

4 Dispositions revisited

Proposal (17) for (dispositional) *le* is nearly identical to the *manage* semantics in (30), and the gap can be further narrowed if we take the modality of (17) to be causal in flavour. To the extent that Inman's (1993) 'happenstantial' modality draws on a stereotypical ordering source, this is a natural move: intuitions about what is normal in any situation are plausibly structured by knowledge about the causal relationships between salient events.

Assigning *le* the full implicative semantics would result in the lexical entry in (39), expressing that the subject of the complex construction takes some action which is contextually causally determinative for the embedded predication:

(39) $[\![\mathsf{le}(P)(x)]\!] := \lambda w \lambda t \lambda e.(\iota A_{evt}.\forall w' \in \mathsf{opt}_{\mathsf{cr},\ \mathsf{cs}}[\mathsf{in}(t,w',A(x)) \leftrightarrow \mathsf{in}(t,w',P(x))])(w)(e)$

This cannot be quite right, since it obscures an important difference between *manage* and *le*: namely, that *manage* complements can be unintended, while *le* requires the embedded predication to be deliberately realized.[6] The initial characterization of $A(x)$ as a target-directed *prior choice* was intended to capture the latter restriction, but this conceptualization is lost in (39), which allows $A(x)$ to be any action with the right relationship to $P(x)$.

Constraining $A(x)$ to be in x's choice set at the relevant world-time indices—even assuming that $\mathsf{CH}(x,w,t)$ contains only options of which x is aware—is still not quite enough: the subject of a *le* predication must choose to realize the embedded predicate itself. One solution might be to treat $A(x)$ as a choice in a very literal sense—i.e., as the act of choosing (from some set of alternative paths) the unique course of action that leads to $P(x)$. I leave the appropriate formalization of this restriction as a topic for future investigation.

These limitations notwithstanding, Proposal (39) represents important progress towards a unified analysis of LV *le*. As I argue below, the proposed implicative semantic structure turns out to be compatible with both the dispositional and 'aspectual' uses of *le*, once the contrastive semantics of Hindi/Urdu (im)perfectives are taken into account.

4.1 Habitual and episodic readings for implicative *le*

Eventive predicates (whether telic or atelic) receive habitual readings in the Hindi/Urdu imperfective (see 1a). We can capture this effect by assuming that IMPF selects for statives (treated here as predicates of times (Nadathur 2023c)), triggering insertion of a stativizing coercion operator when it composes with eventives. (40) offers a preliminary proposal for Habitual coercion, taking Hab to map eventive predicates P to predicates of times during which P is instantiated at all intervals satisfying some characterization R of contextual relevance. Building on Schubert & Pelletier's (1989) analysis of the generic operator, I assume that R minimally picks up any presuppositions of the input predicate P. Using Hab, we derive the interpretation in (41) for imperfective *le* predicates (using the implicative structure in 39, and taking IMPF to contain the reference time within the target situation).

(40) $[\![\mathsf{Hab}]\!] := \lambda w \lambda t \lambda R_{it} \lambda P_{vt}.\forall t'[t' \subset t \,\&\, R(t')][\mathsf{in}(t',w,P)]$

[6] A situation in which Ila intentionally presses a button without being aware that it will open a door is perfectly well described by (1a) but cannot be described by (1b): the *le* construction requires Ila's intention to target the embedded predication.

(1) a. Ila managed to open a door.

b. Ila-ne darvaazaa khol li-yaa.
Ila-ERG door open take-PFV.M
'Ila chose to open the door.'

(41) $[\![\text{IMPF}(\mathsf{Hab}(\text{le}(P)(x)))]\!] = \lambda w \lambda t.\exists t'[t' \supset t \,\&\, \forall t''[t'' \subset t' \,\&\, \mathsf{rel}(t'') \,\&$
$\exists! A.\forall w' \in \mathsf{opt}_{\mathsf{cr},\ \mathsf{cs}}(w)[\mathsf{in}(t,w',A(x)) \leftrightarrow \mathsf{in}(t,w',P(x))]][\mathsf{in}(t'',w,A(x))]]$

The resulting truth conditions express that the reference interval is contained within a period during which all relevant situations where x has a causally determinative choice for $P(x)$ are situations in which x acts on this choice. *Modulo* the question of how choice should best be represented, this seems to capture the right interpretation for the DCP, with the contextual-relevance restriction building in the desired notion of conditionality (see §2).

Since *le*, like *manage*, is eventive, it can compose directly with the episodic Hindi/Urdu perfective. This produces the interpretation in (42):

(42) $[\![\text{PFV}(\text{le}(P)(x))]\!] =$
$\lambda w \lambda t.\exists e[\tau(e) \subseteq t \,\&\, (\iota A.\forall w' \in \mathsf{opt}_{\mathsf{cr},\ \mathsf{cs}}[\mathsf{in}(t,w',A(x)) \leftrightarrow \mathsf{in}(t,w',P(x))])(w)(e)$

(42) requires a context in which x has a causally determinative choice for outcome $P(x)$, and establishes that the agent acts on this choice, thus capturing both the actualization and volitionality requirements of the complex *le* perfective (again, *modulo* a suitable characterization for the relationship between $A(x)$ and the embedded predicate). Coupled with the interpretation in (41), this result provides strong evidence that an implicative-structured semantics is on the right track towards a univocal account of LV *le*.

4.2 Further complications

Even setting aside the question of choice, several challenges remain for the implicative approach to *le*. In the remainder of this section, I briefly discuss the two problems which seem to me to be the most immediate, and conclude by sketching a potential way forward.

Negation. As noted in §2.2, complex *le* predicates, like other aspectual complex predicates, are known to be incompatible with negation (Bhatia 1973; Hook 1974, a.o.):

(43) a. *us-ne gaanaa nahĩĩ gaa li-yaa
*3SG-ERG song NEG sing take-PFV.M
Intended: 'He didn't (choose to) sing a song.'

b. *vo gaanaa nahĩĩ gaa le-taa
*3SG song NEG sing take-IMPF.M
Intended: 'He doesn't/won't (choose to) sing songs.'

These data are not readily explained on the implicative approach. Lexical implicatives are compatible with negation, licensing entailments to the non-realization of their complements (as a consequence of the subject's failure to act on a necessary prerequisite; see 26b).[7] By the same token, assigning a ("choosy") implicative semantics to *le* should result in the intended interpretations in (43), but these sentences are uniformly rejected.

[7] Negating *sak* is also perfectly acceptable, and—under coercion-triggering perfective marking—gives rise to an interpretation more or less parallel to that of negated *manage*.

The only available explanation of the facts in (43) is due to Singh (1990). The account relies on a characterization of *le* as aspectual in nature: Singh proposes that it *le* emphasizes or focuses on the natural completion point of some eventuality in the denotation of the modified predicate.[8] The idea, roughly speaking, is that this effect becomes incoherent in the presence of negation, since it would require emphasizing the culmination of an event which must be either nonexistent or at best incomplete. The explanation seems reasonable enough (if tricky to formalize), but it is not obvious how it may be integrated with an implicative account of *le*, which does not make clear reference to culmination.

Culmination. Recall from §2 that a link between *le* and culmination can be motivated by its effect on telic predicates. As shown in (5), modifying a telic predicate with perfective *le* licenses a culmination entailment which is not present in the simple perfective alternative:

(5) a. Maayaa-ne biskaṭ khaa li-yaa #lekin use puuraa nahĩĩ khaa-yaa.
Maya-ERG cookie eat take-PFV.M but it.ACC whole NEG eat-PFV.M
'Maya ate the cookie, #but did not finish it.'

b. Maayaa-ne biskaṭ khaa-yaa lekin use puuraa nahĩĩ khaa-yaa.
Maya-ERG cookie eat-PFV.M but it.ACC whole NEG eat-PFV.M
'Maya ate the cookie but did not finish it.'

However the terminating but crucially non-culminating interpretation in (5b) is explained, this contrast poses a challenge for the implicative approach to *le*. On the current proposal, *le* adds information about the conditions under which the embedded predicate *P* is undertaken. This should not alter the structure of *P* itself, nor the semantics of the perfective with which the complex predicate combines. Thus, while the *le* perfective will ensure that $P(x)$ was intentionally initiated (i.e., *chosen*), it should remain compatible with a non-culminated instantiation of $P(x)$. (5a) shows that this prediction is not upheld.

Towards a resolution. I have so far ignored a very important point: unlike *manage*, *le* is not a clause-embedding predicate. Instead—like other aspectual LVs—it combines with a lexical verb to form a single predicate structure which behaves syntactically like a simple verb (Butt 1993). This motivates a parallel semantic analysis on which the meaning of a complex *le* predicate involves a kind of lexical merger: the LV contributes aspectual structure and (in)volitionality entailments to the complex predicate, which otherwise inherits its content from the main verb (Butt et al. 1990; Butt 1993; Butt & Ramchand 2005, a.o.). Putting this idea together with recent work from Nadathur & Filip (2021) offers a path towards reconciling the culmination facts in (5) (and potentially also the negation facts in 43, if Singh (1990) is on the right track) with the implicative approach to *le*.

On standard theories of aspectual class, telic predicates are taken to denote exclusively culminated eventualities; (selectively) intensionalized aspectual operators must then be made responsible for instances of telic non-culmination. Nadathur & Filip suggest an alternative approach, proposing that telic event types correspond to *causal models* in which

[8] Other aspectual LVs, such as *paṛ* (see 4a) are instead proposed to focus points of event inception.

the lexically-specified culmination condition of the underlying predicate *P* occurs as a dependent variable: this induces a rich mereological structure which crucially permits both culminated and non-culminated events to validly instantiate *P*. The analysis permits a straightforward, uniformly *partitive* account of grammatical aspects (see also Altshuler (2014)) on which the difference between terminating and culminating perfectives lies in whether the events they instantiate are required to be 'locally' maximal (i.e., corresponding to the maximal instantiation of *P* within the reference time) or maximal with respect to the denotation of *P* itself. On this view, the availability of readings like (5b) is entirely dependent on the inclusion of non-culminated events in the uninflected predicate: if such candidates are excluded, even a 'weak' (terminating) perfective will license culmination entailments.

The hypothesis I wish to entertain is this: if merged with the lexical representation of a telic predicate *P* (structured *à la* Nadathur & Filip), an implicative semantics for *le* will have the effect of 'pruning' the denotation of the input predicate of any non-maximal eventualities. Implicative *le* provides a causal template on which the volitional initiation of a particular type of event is fully determinative of its complete realization in the reference context. Thus, when merged with a telic predicate *P*, the result should be to ensure that volitional initiation of a *P*-event is both causally necessary and—crucially—causally sufficient for the realization of *P*'s lexically specified culmination condition. If this suggestion can be suitably formalized, the resulting denotation for the complex telic predicate is one on which initiation uniformly guarantees culmination. This should produce the interpretation in (5a) even in combination with a weak (non-culminating) Hindi/Urdu perfective: the only eventualities available for instantiation by PFV are, by construction, culminated.

5 Conclusion

I began by pointing out a parallel in the aspectually-driven interpretation of two Hindi/Urdu constructions. For both complex *le* predicates and abilitative *sak* claims, perfective marking appears to eliminate a modal meaning which is detectable in other contexts. Building on Butt (1997) and Bhatt (1999) as well as on prior work on implicative semantic structure in the lexical representation of ability (Nadathur 2023a,b), I proposed a unified account of both phenomena: specifically, that *le* and *sak* share reference to a causal background in which some choice or action for *x* is causally determinative for the realization of the embedded predicate. Dispositional *le* and abilitative *sak* differ in what they establish about the presupposed causing condition: this difference corresponds to a contrast in aspectual class, with consequences for the predicates' respective (re)interpretations under grammatical aspects. If this analysis is on the right track, the perfective's 'de-modalizing' effect turns out to be largely illusory, and the interpretations of both *le* and *sak* claims are compositionally predicted, once the selectional restrictions of (im)perfective aspects are taken into account.

While an implicative semantic structure promises to unify the dispositional and completive uses of LV *le*, a number of analytical problems remain. If the suggestions at the end of §4 should prove fruitful—as I hope they will—the behavior of *le* supports a richer

semantic view of aspectual LVs than anticipated in previous work, and may ultimately lend support to an emerging view of even lexically simple eventuality predicates as invoking *causal models*—richly structured representations of causal information (cf. Nadathur & Filip 2021)—thereby paving the way for a new aspect calculus and a new way of accounting for the distinguishing features of distinct aspectual classes.

Acknowledgements

My thanks are due to Miriam Butt for sharing her original data and insights about the DCP, and to Rajesh Bhatt for additional discussion. Any mistakes and oversights are mine.

References

Altshuler, Daniel. 2014. A typology of partitive aspectual operators. *Natural Language and Linguistic Theory* 32. 735–75.

Arunachalam, Sudha & Anubha Kothari. 2011. An experimental study of Hindi and English perfective interpretation. *Journal of South Asian Linguistics* 4. 463–508.

Baglini, Rebekah & Itamar Francez. 2016. The implications of managing. *Journal of Semantics* 33. 541–60.

Bary, Corien. 2009. *Aspect in Ancient Greek: A semantic analysis of the aorist and imperfective*: Radboud University dissertation.

Beaver, David & Brady Clark. 2008. *Sense and sensitivity: How focus determines meaning* Explorations in Semantics. Oxford: Wiley-Blackwell.

Belnap, Nuel & Michael Perloff. 1988. Seeing to it that: a canonical form for agentives. *Theoria* 54. 175–99.

Bhatia, Tej K. 1973. On the scope of negation in Hindi. *Studies in the Linguistic Sciences* 3. 1–27.

Bhatt, Rajesh. 1999. *Covert modality in non-finite contexts*: University of Pennsylvania dissertation.

Brown, Mark. 1988. On the logic of ability. *Journal of Philosophical Logic* 17. 1–26.

Butt, Miriam. 1993. *The structure of complex predicates in Urdu*: Stanford University dissertation.

Butt, Miriam. 1997. Dispositional complex predicates in Urdu. Talk, *Sinn und Bedeutung*.

Butt, Miriam, Michio Isoda & Peter Sells. 1990. Complex predicates in LFG. Ms, Stanford University.

Butt, Miriam & Gillian Ramchand. 2005. Complex aspectual structure in Hindi/Urdu. In N. Erteschik-Shir & T. Rapaport (eds.), *The syntax of aspect: deriving thematic and aspectual interpretation*, 117–53. Oxford: Oxford University Press.

Coleman, Linda. 1975. The case of the vanishing presupposition. In *Proceedings of the Berkeley Linguistics Society*, vol. 1, 78–89.

Fernald, Theodore. 1999. Evidential coercion: using individual-level predicates in stage-level environments. *Studies in the Linguistic Sciences* 29. 43–63.

Hacquard, Valentine. 2005. Aspects of *too* and *enough* constructions. In *Proceedings of Semantics and Linguistic Theory*, vol. 15, 80–97.
Hacquard, Valentine. 2009. On the interaction of aspect and modal auxiliaries. *Linguistics and Philosophy* 32. 279–312.
Hacquard, Valentine. 2020. Actuality entailments. In D. Gutzmann, L. Matthewson, C. Meier, H. Rullman & T.E. Zimmerman (eds.), *The Wiley Blackwell Companion to Semantics*, Oxford: Wiley.
Homer, Vincent. 2021. Actualistic interpretations in French. *Semantics and Pragmatics* 14. 12.1–56.
Hook, P.E. 1974. *The compound verb in Hindi*. The University of Michigan: Center for South and Southeast Asian Studies.
Inman, Michael. 1993. *The semantics and pragmatics of colloquial Sinhala involitive verbs*: Stanford University dissertation.
Karttunen, Lauri. 1971. Implicative verbs. *Language* 47. 340–58.
Kaufmann, Stefan. 2013. Causal premise semantics. *Cognitive Science* 37. 1136–70.
Kenny, Anthony. 1976. Human ability and dynamic modalities. In J. Manninen & R. Tuomela (eds.), *Essays of explanation & understanding*, 209–32. Dordrecht: Reidel.
Kratzer, Angelika. 1998. More structural analogies between pronouns and tense. In *Proceedings of Semantics and Linguistic Theory*, vol. 8, 92–110.
Lawler, John. 1973. Tracking the generic toad. In *Proceedings of the Chicago Linguistic Society*, vol. 9, 320–31.
Louie, Meagan. 2015. *The temporal semantics of aspect and circumstance in Blackfoot*: University of British Columbia dissertation.
Maier, John. 2018. Ability, modality, and genericity. *Philosophical Studies* 175. 411–28.
Mandelkern, Matthew, Ginger Schultheis & David Boylan. 2017. Agentive modals. *The Philosophical Review* 126. 301–43.
Mohanan, Tara. 1990. *Arguments in Hindi*: Stanford University dissertation.
Nadathur, Prerna. 2023a. *Actuality inferences: causality, aspect, and modality*. Oxford: Oxford University Press.
Nadathur, Prerna. 2023b. Causal semantics for implicative verbs. *Journal of Semantics* 40. 311–58.
Nadathur, Prerna. 2023c. Variable implicativity in *enough* constructions: causation, coercion, and composition. Ms, The Ohio State University.
Nadathur, Prerna & Hana Filip. 2021. Telicity, teleological modality and (non)culmination. In *Proceedings of the West Coast Conference on Formal Linguistics*, vol. 39, 1–8.
Pearl, Judea. 2000. *Causality*. Cambridge: Cambridge University Press.
Schubert, Lenhart & Francis Pelletier. 1989. Generically speaking, or, using discourse representation theory to interpret generics. In G. Chierchia, B. Partee & R. Turner (eds.), *Properties, types and meaning*, vol. II, 193–268. Dordrecht: Kluwer/Reidel.
Schulz, Katrin. 2011. If you'd wiggled a, then b would've changed. *Synthese* 179. 239–51.
Singh, Mona. 1990. The aspectual content of compound verbs. In Y. No & M. Libucha (eds.), *Proceedings of the eastern states conference on linguistics*, vol. 7, 260–71.

Singh, Mona. 1998. On the semantics of the perfective aspect. *Natural Language Semantics* 6. 171–99.
de Swart, Henriëtte. 1998. Aspect shift and coercion. *Natural Language and Linguistic Theory* 16. 347–85.

Verb Root Allomorphy in Indo-Aryan Languages

PAROMA SANYAL, *Indian Institute of Technology Delhi*

ABSTRACT

This paper aims to theoretically describe the phonological basis for the systematic verb root allomorphy seen in some of the Indo-Aryan languages. To highlight the similarities and differences in allomorphy patterns, three languages, Bangla, Hindi and Odia have been compared. The paper proposes that these languages typologically choose to achieve *Paradigm Uniform*ity (Steriade 2000) within the verbal paradigm by introducing additional phonological processes, rather than by blocking the phonological processes active in the rest of the language.

1 Introduction

Natural languages tolerate allomorphy where a single lexical or syntactic item corresponds to more than one surface phonological representation. Nevertheless, within languages, such one to many mappings are restricted to small lexical sets within specific morpho-syntactic paradigms. In fact, morphological well-formedness concepts like *Paradigm Uniformity* (Kiparsky 1982, Kenstowicz 1998, Steriade 2000) have been proposed to account for the markedness of the phenomena. This has been defined in (1).

(1) Paradigm Uniformity
All output forms of α, where α is a morpheme shared by all members of a morphological paradigm P, must be identical.

In this paper, I analyze the phonological phenomenon of vowel allomorphy in the verb roots of three modern Indo-European languages, Hindi, Odia and Bangla. These languages have been selected for comparison because they share a substantial portion of the lexicon but have very distinct phonological characteristics with respect to vowels.

Hindi shares the property of vowel length distinction in non-mid vowels with other western Indo-Aryan languages. While this distinction is phonemic in non-verbs (2a), it is used to distinguish morphologically related forms in verbs (2b).

(2) Long and short vowel distinction in Hindi

a) Non-verbs

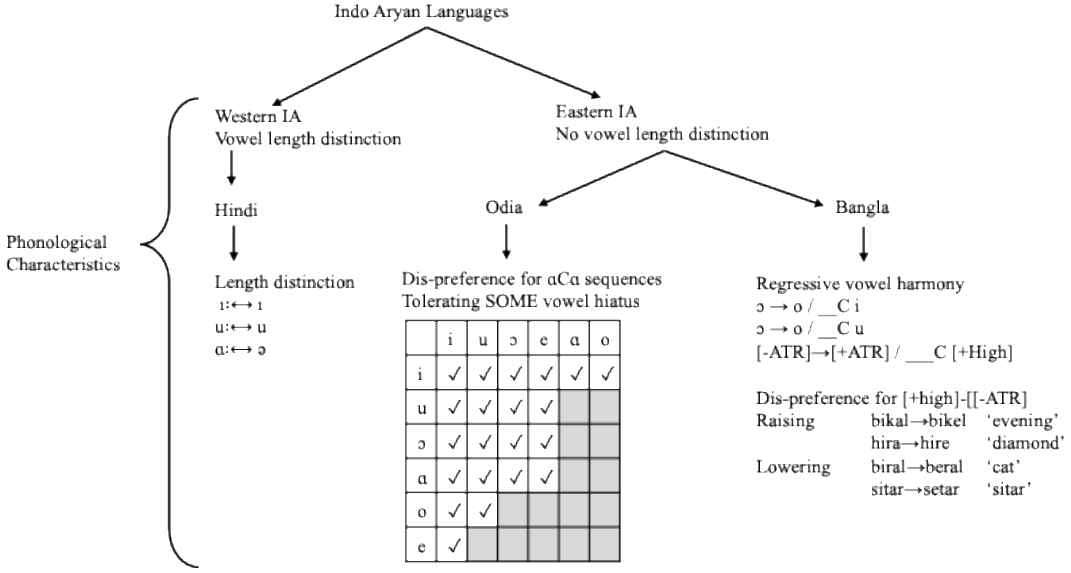
d̪i:n	'poor'	d̪in	'day'
d̪u:r	'distance'	d̪ur	'bad'
ʈa:n	'string of musical notes'	ʈən	'body'

b) Non-verbs

Transitive.perf		Intransitive.perf	
pi:t-a:	'beat up'	pit-a:	'got beaten up'
lu:t-a:	'robbed'	lut-a:	'got robbed'
ka:t-a :	'cut'	kət-a:	'got cut'

Unlike Western Indo-Aryan languages, the Eastern Indo-Aryan languages have no phonemic contrast in vowel length. However, these languages exhibit distributional restrictions on the cooccurrence of vowels within the phonology of the language. In Odia, which tolerates the marked phonological context of vowel hiatus, the distributional restrictions appear in this context. In contrast, in Bangla such distributional restrictions are spread throughout the language and have been analysed as different types of vowel harmonies in preceding literature (Chatterjee 1926/1978, Dasgupta 1982, Dan 1998, Sanyal 2011). The contrast in the phonological characteristics of the vowel systems of these three languages have been schematically represented in (3).

(3) Schematic representation of Phonological characteristics



	i	u	ɔ	e	ɑ	o
i	✓	✓	✓	✓	✓	✓
u	✓	✓	✓	✓		
ɔ	✓	✓	✓	✓		
ɑ	✓	✓	✓	✓		
o	✓	✓				
e	✓					

Verb root allomorphy in Hindi, Odia and Bangla involve change in the root vowel. This has been shown in (4) with the example of the verb 'sleep' which surfaces as [so] and [su].

(4) Change in root vowels in Verbs

a) Hindi

(i) s**o:** rəha: hũ:
sleep prog. Be.1p
I am sleeping.

(ii) s**u**l-a: rəha: hũ:
sleep-cause prog. Be.1p
I am putting someone to sleep.

b) Odia

(i) s**o**-uch-i
sleep-prog.1p
I am sleeping.

(ii) s**u**-o-uch-i
sleep-cause-prog.1p
I am putting someone to sleep.

c) Bangla

(i) s**u**-cch-i
sleep-prog.1p
I am sleeping.

(ii) s**o**-a-cch-i
sleep-cause-prog.1p
I am putting someone to sleep.

Since, vowel phonology is central to the analysis of this phenomenon, we will begin the discussion with a theoretical discussion on vowel systems in sub-section 1.1 followed by a short description of the vowel systems of the three languages in sub-section 1.2.

1.1 Vowel Primes

Following the theory of Government Phonology (Kaye et al. 1985, 1990, 1992, 1995 and subsequent work) we assume three Primes [I, U and A] that are the base formatives of the vocalic system each having a marked ***Hot*** feature. In addition, there is a cold vowel *v* that lacks any hot feature.

(5) Primes

a) $I=\begin{bmatrix} \mathbf{-Back} \\ +High \\ -Round \\ -ATR \\ -Low \end{bmatrix}$ b) $A=\begin{bmatrix} +Back \\ \mathbf{-High} \\ -Round \\ -ATR \\ +Low \end{bmatrix}$ c) $U=\begin{bmatrix} +Back \\ +High \\ \mathbf{+Round} \\ -ATR \\ -Low \end{bmatrix}$ d) $\text{Ɨ}=\begin{bmatrix} +Back \\ +High \\ -Round \\ +ATR \\ -Low \end{bmatrix}$

The Primes can combine with one another to form complex primes through the operation of combining a Head and an Operator. The resultant complex copies in the Hot feature of the Head and the remaining feature values from the Operator.

(6) Prime correlates to vowels

Head Operator Head Operator V

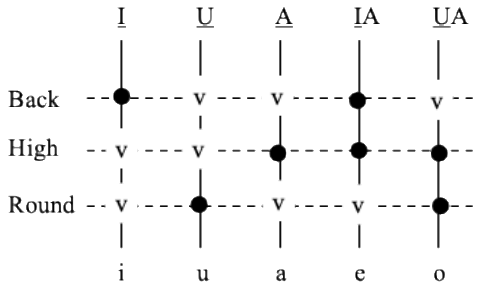
$I=\begin{bmatrix} \mathbf{-Back} \\ +High \\ -Round \\ -ATR \\ -Low \end{bmatrix} . \quad \text{Ɨ}=\begin{bmatrix} +Back \\ +High \\ -Round \\ +ATR \\ -Low \end{bmatrix} \longrightarrow \underline{I}=\begin{bmatrix} -Back \\ +High \\ -Round \\ +ATR \\ -Low \end{bmatrix}$ /i/

$I=\begin{bmatrix} \mathbf{-Back} \\ +High \\ -Round \\ -ATR \\ -Low \end{bmatrix} . \quad A=\begin{bmatrix} +Back \\ \mathbf{-High} \\ -Round \\ -ATR \\ +Low \end{bmatrix} \rightarrow |\underline{I}A|=\begin{bmatrix} \mathbf{-Back} \\ \mathbf{-High} \\ -Round \\ -ATR \\ +Low \end{bmatrix} . \quad \text{Ɨ}=\begin{bmatrix} +Back \\ +High \\ -Round \\ +ATR \\ -Low \end{bmatrix} \quad |\underline{I}A|=\begin{bmatrix} -Back \\ -High \\ -Round \\ +ATR \\ -Low \end{bmatrix}$ /e/

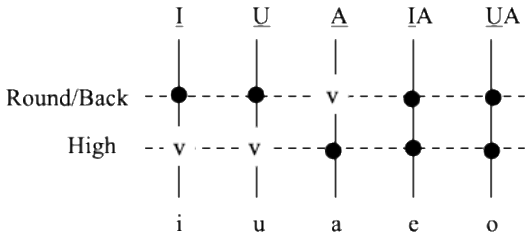
The Hot features also form autosegmental tiers on which the phonemes can be plotted. These tiers can have fused exponence in the vocalic systems of particular languages with the consequence that the primes associated with the fused tiers primes cannot combine to form a complex sound. This has been demonstrated in (7) with a respect to a toy grammar with a five-vowel vocalic system.

(7) Vocalic systems with Fused tiers

a) Distinct Tiers



b) Fused Tiers



In the following section, the vowel repertoire of the three Indo-Aryan languages, Hindi, Odia and Bangla have been characterized using this formulation.

1.2 Vowel Primes in Hindi, Odia and Bangla

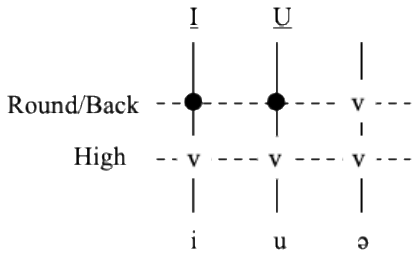
Hindi has a basic repertoire of three short and seven long vowels[1]. Among the long vowels there is optionality between the diphthongs /ɛe/ and /ɔo/, with the monophthongs /ɛː/ and /ɔː/ respectively. Leaving out these two long vowel phonemes, the rest of the long vowels have a distribution of primes identical to the one demonstrated in (7b). The Prime and (syllable) Rhyme based representation of the vocalic phonemes of Hindi are shown in (8).

(8) Vowel phonemes in Hindi

Representations→ Prime-based Rhyme-based

a) Short vowels

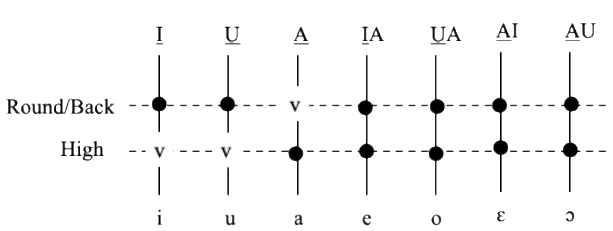
mil-na: ‘to find’
juɖ-na: ‘to join’
cəl-na: ‘to walk’



[1] The long vowels can also have oral and nasal counterparts. (Think about it)

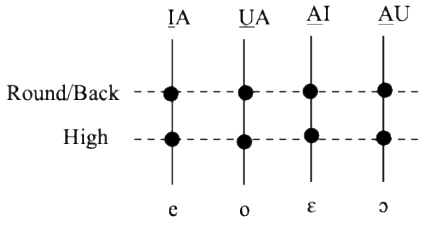
b) Long vowels

pi:ʈ-na: ‘to beat up’
lu:ʈ-na: ‘to loot’
c^ha:p-na: ‘to print’
ɖe:k^h-na: ‘to see’
k^ho:l-na: ‘to open’
bɛ:$ʈ^h$-na: ‘to sit’
ɖɔ:ɖ-na: ‘to run’



c) Diphthongs

bɛeʈ^h-na: ‘to sit’
ɖɔoɖ-na: ‘to run’



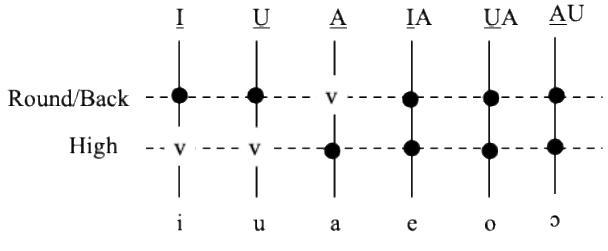
Hindi has word final phonetic lengthening, where the word final vowel is always a long vowel. Since the vowel schwa [ə] does not have a long counterpart, it fails to appear word finally. In section 2.1 we argue that this is the reason CV verb roots do not surface with the vowel schwa.

Unlike Hindi, where vowel length plays an important role in the phonological distribution, in the Eastern Indo-Aryan languages, all vowels are generally pronounced as phonetically long without any phonemic length distinction. The Prime and Rhyme based representation of the Odia vowels are shown in (9).

(9) Vowel phonemes in Odia

Representations→ Prime-based Rhyme-based

piʈ ‘beat up.imp’
luʈ ‘loot.imp’
ɖak ‘call.imp’
ɖek^h ‘see.imp’
k^hol ‘open,imp’
bɔs ‘sit.imp’
ɖɔ:ɖ ‘to run’

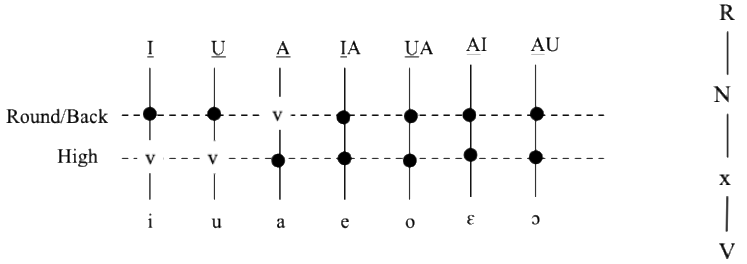


With respect to the vowel repertoire, Bangla differs minimally from Odia by having one phoneme more than the latter. However, the vowel allomorphy patterns in verb roots of these two sister languages differ from each other substantially as will be seen in section 2.2 and 2.3. In Bangla verb roots, the vowels [i, u] never appear in the same morphological paradigm with the vowels [ɛ, ɔ], though these form minimal pairs in the non-verbal domain. So, in (10) I am using non-verb examples to showcase the vowel repertoire of Bangla.

(10) Vowel phonemes in Bangla

Representations→ Prime-based Rhyme-based

d̪in ‘day’
d̪ul ‘earring
d̪an ‘given object’
ʈel ‘oil
kol ‘lap’
d̪^{h}ɛn ‘attention’
kɔl ‘tap’



In the following section I show how these vowels of Hindi, Odia and Bangla show systematic patterns of alternation in different morpho-syntactic contexts.

2 Verb Root Allomorphy

Unlike languages like English that form lexical causatives (11a), many of the Indo-Aryan languages form systematic morphological causatives (11b-d).

(11) Lexical vs Morphological causatives

Lexical Causative	a) English	[eat]↔[feed]	[see]↔[show]
Morphological Causative	b) Hindi	[k^{h}a:-na:]↔[k^{h}il-a:-na:]	[d̪ek^{h}-na:]↔[d̪ik^{h}-a:-na:]
	c) Odia	[k^{h}a-i-ba]↔[k^{h}u-o-i-ba]	[d̪ek^{h}-i-ba]↔[d̪ek^{h}-o-i-ba]
	d) Bangla	[k^{h}a-wa]↔[k^{h}a-wa-no]	[d̪ɛk^{h}-a]↔[d̪ɛk^{h}-a-no]

Despite the overt morphological marking of causatives, the verb roots of causatives in all three languages form a sub-paradigm within which the vowel of verb roots gets modified. In Odia, this is the only environment for systematic allomorphy in CV verb roots. Hindi shows identical patterns of verb root allomorphy in causative and intransitive contexts. Thus the unaccusative and causative forms of transitive verb roots in Hindi have identical allomorphic forms. The same phonological pattern is seen in the causative forms of unergative and ditransitive verbs in Hindi as well. Unlike these two, the Bangla allmomorphy pattern is slightly more complex as it can be divided into three phonological sub-paradigms that correspond to five distinct morphological environments. The following sections 2.1, 2.2 and 2.3 discuss these correspondences between the phonological and morphological contexts for allomorphy in Hindi, Odia and Bangla respectively.

2.2 Patterns of verb root allomorphy in Hindi

In Hindi transitive verbs can be morphologically derived to form the corresponding unaccusative verb by systematically modifying the vowel in verb roots. This has been shown in (12).

(12) Transitive-Unaccusative alternation in Hindi

Transitive	Unaccusative	Causative	
pi:ʈ-na:	piʈ-na:	piʈ-wa:-na:	‘to beat up’
d̂e:k^{h}-na:	d̂ik^{h}-na:	d̂ik^{h}-a:-na:	‘to see’
ĝ^{h}is-na:	ĝ^{h}is-na:	ĝ^{h}is-wa:-na:	‘to rub’
lu:ʈ-na:	luʈ-na:	luʈ-wa:-na:	‘to loot’
k^{h}o:l-na:	k^{h}ul-na:	k^{h}ul-wa:-na:	‘to open’
bun-na:	bun-na:	bun-wa:-na:	‘to knit’
c^{h}a:p-na:	c^{h}əp-na:	c^{h}əp-wa:-na:	‘to print’
ţ̂əl-na:	ţ̂əl-na:	ţ̂əl-wa:-na:	‘to fry’

Each of the long vowel from the transitive verb root changes to a short vowel in the corresponding unaccusative and causative verb root. Similar changes take place in deriving the causative forms for unergative verbs as well, shown in (13).

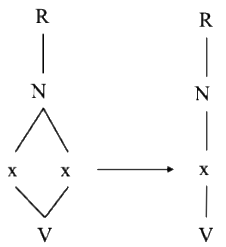
(13) Unergative-Causative alternation in Hindi

Unergative	Causative		
ci:k^{h}-na:	cik^{h}-wa:-na:	*cik^{h}-na:	‘to scream’
le:ʈ -na:	liʈ-a:-na:	*liʈ -na:	‘to lie down’
gir-na:	gir-a:-na:		‘to fall’
c^{h}u:ʈ-na:	c^{h}uʈ-wa:-na:	*c^{h}uʈ-na:	‘to escape’
bo:l-na:	bul-wa:-na:	*bul-na:	‘to say’
sun-na:	sun-a:-na:		‘to listen’
ma:ŋ-na:	məŋ--wa:-na:	*məŋ-na:	‘to print’
cəl-na:	cəl-a:-na:		‘to fry’

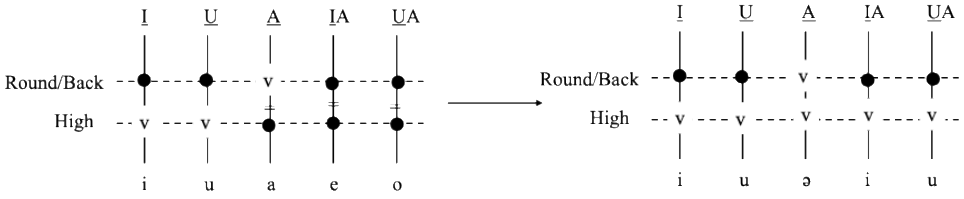
The phonological pattern of change in these modifications involves two changes. First, there is a reduction at the segmental tier where a binary branching nucleus changes to an unbranched one with just one segmental position, shown in (14a). Secondly, there is a loss in the prime |A| in the root vowels which contain the prime |A|. Thus, the two complex long vowels |I̲A| and |U̲A| reduce to short |I̲| and |U̲|, and the long vowel |A| with a single prime loses its hot feature to become the reduced vowel schwa that surfaces without any hot feature. This has been schematically represented in (14b) as the delinking of the root vowel segments from the [High] tier resulting in the loss of |A|.

(14) Phonological changes in Hindi Verb roots

(a) Change in skeletal tier



(b) Change in Primes



The complex vowels |AI| and |AU| that show optionality between dipththong and long monophthong surface realization also tend to show lexical variation with respect to this vowel reduction as seen in (15).

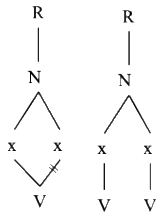
(15) Vowel reduction in Diphthongs

Unergative	Causative		
bɛeʈʰ-na: ~ bɛ:ʈʰ-na:	bɛeʈʰ-a:-na: ~ biʈʰ-a:-na:		'to sit'
t̪ɛer-na: ~ t̪ɛ:r-na:	t̪ɛer-a:-na:	*t̪ir-a:-na:	'to swim'
d̪ɔod̪-na: ~ d̪ɔ:d̪-na:	d̪ɔod̪-a:-na:	*d̪ud̪-a:-na:	'to 'run'
cɔ̃õk-na: ~ cɔ̃:k	cɔ̃õk-a:-na:	*cũk-a:-na:	'to shock'

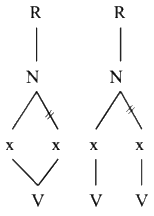
Whether a native speaker is able to reduce the root vowel in causatives depends on whether they interpret it as a monophthong or a diphthong. For those cases where the vowel is categorized as a diphthong by the individual speaker, there is no reduction. This confirms that the change in segmental tier is though the delinking of the doubly linked V (16a), rather than the change of the branching nucleus to a non-branching one. Were it the latter case (16b), all long vowels including diphthongs would have shown the same reduction process in causatives.

(16) Delinking within the skeletal tier

a) Only Monophthongs get reduced



b) All long vowels get reduced



In case of Hindi, there is no phonological motivation for this vowel reduction. Such vowel reduction in lexical vowels is not seen anywhere else in the language as well. This reduction, in fact results in accidental neutralization of the lexical distinction within the causative paradigm, as shown in (17).

(17) Neutralization of lexical distinction in causatives

Verb root	Causative verb root	
kʰa:-na:	kʰil-a:-na:	'to eat'

kʰel-na:	kʰil-a:-na:	‘to play’
kʰil-na:	kʰil-a:-na:	‘to bloom’

Hindi tolerates such lexical neutralization in verb roots in order to create a distinct paradigm of morphologically derived unaccusatives that differ from their transitive counterparts without additional suffixation. There is a semantic commonality between the transitive-causative and transitive-unaccusative derivation, because both cases involve a change in the theta role of the transitive verb .

2.3 Patterns of verb root allomorphy in Odia

In Odia verb root allomorphy is restricted to the CV roots. Further, among the CV roots *Paradigm uniformity* is restricted to the Causative paradigm. This means that in the non-causative morphological contexts the CV verb roots could surface with more than one allomorph depending on the phonological context and the resulting non-uniformity within the paradigm is tolerated. Unlike Hindi, the allomorphy pattern in Odia is partly phonologically motivated as it avoid illicit sequences of vowel hiatus. This has been shown in (18).

(18) Neutralization of lexical distinction in causatives

a) Non-causatives

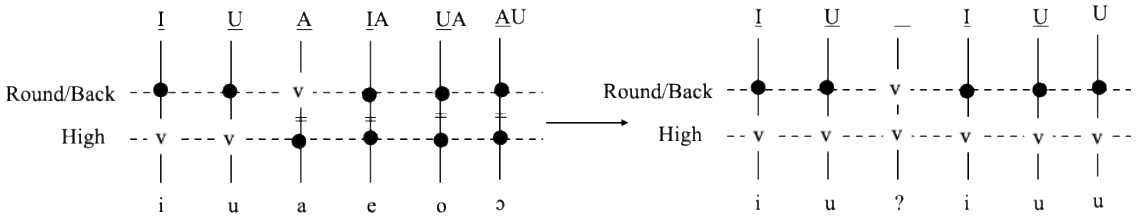
	hab.1p	hab.2p	prog.1p	perf.1p	
	-e	*-ɔ*	*-u-cʰi*	*-i-cʰi*	
pi	pi-e	pi-ɔ	pi-u-cʰi	pi-i-cʰi	‘to drink’
cʰũ	cʰũ-e	cʰũ-ɔ	cʰũ-u-cʰi	cʰũ-i-cʰi	‘to touch’
d̰e	**d̰i-e**	**d̰i-ɔ**	**d̰o-u-cʰi**	d̰e-i-cʰi	‘to give’
so	**su-e**	**su-ɔ**	so-u-cʰi	so-i-cʰi	‘to sleep’
kʰa	kʰa-e	kʰa-ɔ	kʰa-u-cʰi	kʰa-i-cʰi	‘to eat’
nɔ	nɔ-e	nɔ-ɔ	nɔ-u-cʰi	nɔ-i-cʰi	‘to bend’

b) Causatives

	hab.1p	prog.1p	perf.1p	
	-a-e	*-o-u-cʰi*	*-e-i-cʰi*	
pi	pi-a-e	pi-o-u-cʰi	pi-e-i-cʰi	‘to drink’
cʰũ	cʰũ-a-e	cʰũ-o-u-cʰi	cʰũ-e-i-cʰi	‘to touch’
d̰e	**d̰i-a-e**	**d̰i-o-u-cʰi**	**d̰i-e-i-cʰi**	‘to give’
so	**su-a-e**	**su-o-u-cʰi**	**su-e-i-cʰi**	‘to sleep’
kʰa	**kʰu-a-e**	**kʰu-o-u-cʰi**	**kʰu-e-i-cʰi**	‘to eat’
nɔ	**nu-a-e**	**nu-o-u-cʰi**	**nu-e-i-cʰi**	‘to bend’

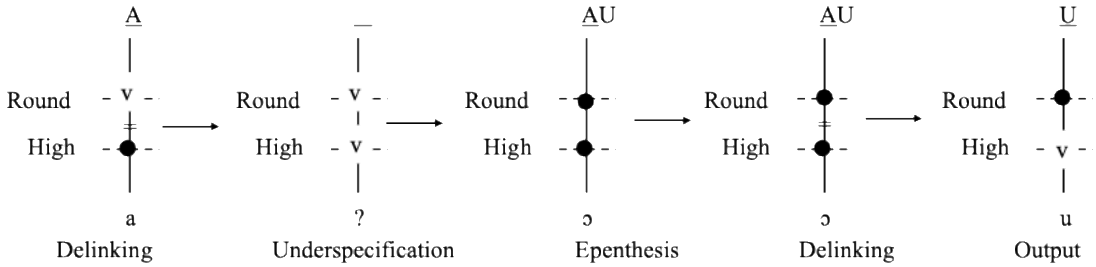
The causative paradigm of Odia is phonologically more complex than Hindi, as there are distinct vowels that appear as the causative morpheme in the habitual, progressive and perfect contexts which are also vowel-initial suffixes. Phonologically, all the scenarios of verb root allomorphy, save one [d̰o-u-chi] involve delinking of the [High] feature and the consequent erasure of the prime |A| from the CV roots. This is schematically shown in (19).

(19) Phonological change in Odia Causative verb roots



While Odia is similar to Hindi in delinking the associations with the [High] tier as part of the Causative morphology, it differs from the latter in the realization of the vowel without any hot feature. Unlike Hindi which has a phoneme schwa to corresponds with the cold feature set, Odia does not have any output representation that matches this description. To surface in a structure preserving manner (Kiparsky 1985) this underspecified form must now acquire the characteristics of the default epenthetic vowel for the language. In Odia, this vowel is [ɔ] with the primes |A̲I|. Once characterized as [ɔ] it follows the same delinking and reduction process that takes place for underlying cases of |A̲I|. This derivational route has been shown in (20).

(20) Process of reduction of |A̲| to |U̲| in Odia Causatives



Independent phonological evidence supporting the analysis in (20) can be found at two different places within the Odia lexicon. The first is the case of delinking an underlying [a] from the [High] tier when it is followed by another [a] in the consecutive syllable. This is found in both verbs as well as non-verbs, but in non-verbs it becomes apparent only when the forms are compared to cognates in other Indo-Aryan languages. This has been shown in (21).

(21) Dis-preference for consecutive [a] vowels in Odia

a) Verbs

	hab.1p	Caus.hab.1p	Caus.prog.1p	Caus.perf.1p	
ɖ̲ak	ɖ̲ak-e	**ɖ̲ɔk**-a-e	**ɖ̲ɔk**-o-u-c^{h}i	**ɖ̲ɔk**-e-i-c^{h}i	‘to call’

b) Non-Verbs

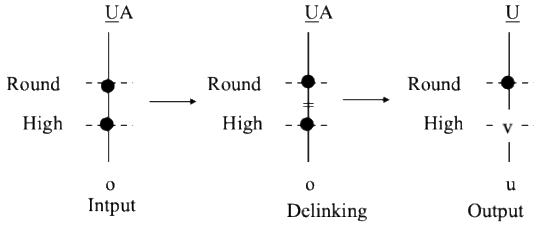
Odia	Hindi	Bangla	
c^{h}ɔ̲ta	c^{h}a:ʈ̲a:	c^{h}a̲ta	‘umbrella’
ɔʈa	a:ʈa:	aʈa	‘flour’

In the case of (21a) and (21b), the second sequence of delinking from (20) does not take place since the output forms with [ɔ] followed by [a] with an intervening consonant is not a marked sequence that requires repair. In case of the causative form of verbs, the phonological change is overapplied across all the causative forms of the lexical item to maintain *Paradigm Uniformity*.

The second evidence comes from the distributional restriction on the vowel [o] in different dialects of Odia. While Katki Odia, spoken around the east central region of the state restricts [o] to initial syllables, Sambalpuri Odia, spoken along the western flank prefers to restrict it to the final syllable (Guru and Nayak, 2024). A comparison of cognates between the two dialects reveals that the underlying [o] surfaces as [u] in prosodic positions where the former is not licensed. This has been shown in (22).

(22) Reduction of [o] to [u] in Odia dialects

UR	seo ‘apple’	gora ‘fair’
Delinking		
SR		
Sambalpuri	seo	**gura**
Katki	**seu**	gora



2.4 Patterns of verb root allomorphy in Bangla

Bangla has two distinct processes of vowel harmony that could have potentially made the verbal paradigm non-uniform. These two processes are listed in (23).

(23) Phonological processes in Bangla

a) Regressive harmony

ɟɔj	‘victory’	ɟoj-i	‘victor’
pɔt̪^{h}	‘path’	pot̪^{h}-ik	‘traveller’

b) Prosodic misalignment adjustment

bikal∼bikel	‘evening’	biral∼beral	‘cat’	biʃal	‘huge’
ɟuṭa∼ɟuṭo	‘shoe’	ḍukan∼ḍokan	‘shop’	ʃikar	‘hunt’
piṭ̪ol∼peṭ̪ol	‘brass’	ʃikɔr∼ʃekor	‘root’	iʃʃɔr	‘god’

In (23a) when the mid lax vowel [ɔ] is followed by the vowels [i] or [u] in the following syllable, the former raises to [o]. Except for a handful of prefixes, this process applies throughout the lexicon. The second one (23b), is analyzed as a case of prosodic misalignment by Sanyal (2011). I observe that there is a distinct dis-preference for disyllabic lexical strings where a [+high] vowel is followed by a [-ATR] one, and often

results in alternative pronunciations that address this context by either lowering the first vowel or raising the second one. In Sanyal (2011) I argue that this dis-preference is caused by prosodic mis-alignment where the stresses initial syllable contains vowels with lowest sonority and its corresponding unstressed second syllable contains the maximally sonorous vowels.

In the lexical domain of verb roots, both these phonological processes could have potentially induced phonologically conditioned allomorphy. In (24) I demonstrate this point using the hypothetical verb root template [kvl], where [v] is replaced with the seven vowels or the repertoire, with actual inflectional markers from Bangla.

(24) Hypothetical case of phonological processes applying in the verbal domain

Root vowel	*Verb	Habitual.1p	Gerund
i	kil	kil-i→kili	kil-a→kela
e	kel	kel-i→keli	kel-a→kela
ɛ	kɛl	kɛl-i→keli	kɛl-a→kɛla
u	kul	kul-i→kuli	kul-a→kola
o	kol	kol-i→koli	kol-a→kola
ɔ	kɔl	kɔl-i→koli	kɔl-a→kɔla
a	kal	kal-i→kali	kal-a→kala

In (24) the cases of potential homophony are highlighted. Not only would these phonological processes make the verb root paradigm phonologically non-uniform, but it would produce homophonous outputs within the paradigm. To avoid this scenario, the verb roots in Bangla are divided into three phonological sub-paradigms that correspond to different syntactic contexts. These three sub-paradigms are shown in (25).

(25) Phonological sub-paradigms in Bangla verb roots

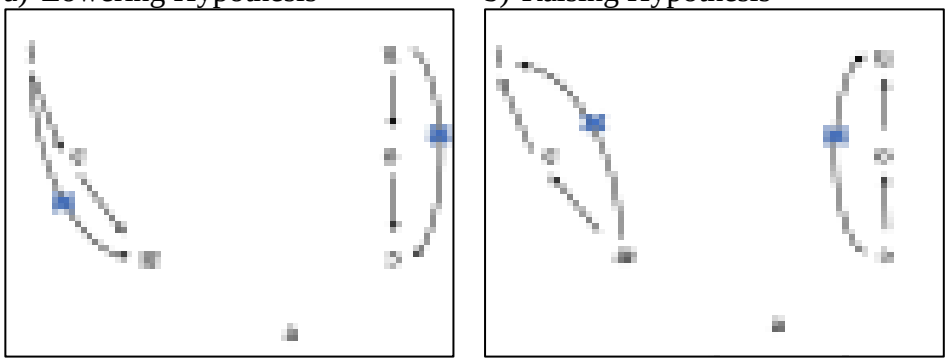
Sub-Paradigm	A	B	C	
Root Vowels	[e, o, a, ɔ, ɛ]	[i, u, a, o, e]	[i, u, o, e]	
a) CVC roots	Imperative.2p	Permissive/ (Prohibitive)	Perf.present.1p	
	lek^{h}-o	lik^{h}-o (na)	lik^{h}-e-c^{h}-i	‘write’
	ʃon-o	ʃun-o (na)	ʃun-e-c^{h}-i	‘listen’
	rak^{h}-o	rak^{h}-o (na)	rek^{h}-e-c^{h}-i	‘keep’
	kɔr-o	kor-o (na)	kor-e-c^{h}-i	‘do’
	d̪̂ɛk^{h}-o	d̪̂ek^{h}-o (na)	d̪̂ek^{h}-e-c^{h}-i	‘see’

Sub-Paradigm	A	C		
Root Vowels	[o, a, e]	[i, u, e]	[i, u, e]	
b) CV roots	Imperative.1p	Permissive/ (Prohibitive)	Perf.present.1p	
	ca	ce-o (na)	ce-e-c^{h}-i	‘ask for’
	ʃo	ʃu-o (na)	ʃu-e-c^{h}-i	‘sleep’
	ne	ni-o (na)	ni-e-c^{h}-i	‘keep’

In (25a) CVC roots form three sub-paradigms in such a way that each verb root has two allomorphs that can be derived from the other form by chain shift lowering or raising.

(26) Schematic representation of chain shift

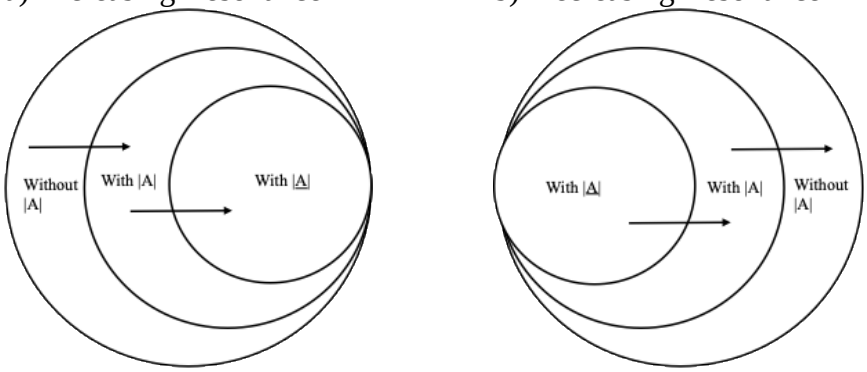
a) Lowering Hypothesis b) Raising Hypothesis



The schematics in (26a) and (26b) both depict the patterns root vowel change between sub-paradigms A and B in (25a). Here, the vowels that undergo change can be divided into three sets: Those without |A|, those with |A| and those with |A| as head. Since the change is symmetric, we could either consider A or B as the base paradigm and derive the other. This has been shown using primes in (27).

(27) Prime-based representation change in resonance for Root vowel

a) Increasing Resonance b) Decreasing Resonance



The Low vowel [a] does not participate in allomorphy between sub-paradigms A and B. However, it participates in C by failing to surface and showing a decrease in resonance, similar to B. Based on this pattern of change between A to C, I concur with preceding literature that the process in Bangla CVC verb roots could be characterised as vowel raising or decrease in sonority. To prevent homophony between items in the within the paradigm, the Sub-paradigm A, henceforth Root 1, does not contain [+high] vowels. Root 2 is derived by making a gradual change with respect to the realization of the prime |A|. In (25b) we find that with CV roots, the sub-paradigm B is missing altogether and the change is from pattern A in Root 1 to pattern C in Root 2. Further, there are only two levels of change in

the resonance of CV roots, instead of the three levels of change in the phonological system of CVC roots[2].

There are clear morpho-syntactic paradigms associated with Sub-paradigms B and C in Bangla where the verb roots have uniform exponence. These have been listed in (28a) and (28b) below.

(28) Morpho-syntactic paradigms

a) Sub-Paradigm B

Past.1p	Past.hab.1p	Past.prog.1p	Present.prog.1p	Fut.1p	
lik^{h}-l-am	lik^{h}-t-am	lik^{h}-c^{h}i-l-am	lik^{h}-c^{h}-i	lik^{h}-b-o	'write'
ʃun-l-am	ʃun-t-am	ʃun-c^{h}i-l-am	ʃun-c^{h}-i	ʃun-b-o	'listen'
rak^{h}-l-am	rak^{h}-t-am	rak^{h}-c^{h}i-l-am	rak^{h}-c^{h}-i	rak^{h}-b-o	'keep'
kor-l-am	kor-t-am	kor-c^{h}i-l-am	kor-c^{h}-i	kor-b-o	'do'
ḍ̂ek^{h}-l-am	ḍ̂ek^{h}-t-am	ḍ̂ek^{h}-c^{h}i-l-am	ḍ̂ek^{h}-c^{h}-i	ḍ̂ek^{h}-b-o	'see'

Past.2p	Past.hab.2p	Past.prog.2p	Present.prog.2p	Fut.2p	Subjunctive.2p
lik^{h}-l-e	lik^{h}-t-e	lik^{h}-c^{h}i-l-e	lik^{h}-c^{h}-o	lik^{h}-b-e	lik^{h}-o
ʃun-l-e	ʃun-t-e	ʃun-c^{h}i-l-e	ʃun-c^{h}-o	ʃun-b-e	ʃun-o
rak^{h}-l-e	rak^{h}-t-e	rak^{h}-c^{h}i-l-e	rak^{h}-c^{h}-o	rak^{h}-b-e	rak^{h}-o
kor-l-e	kor-t-e	kor-c^{h}i-l-e	kor-c^{h}-o	kor-b-e	kor-o
ḍ̂ek^{h}-l-e	ḍ̂ek^{h}-t-e	ḍ̂ek^{h}-c^{h}i-l-e	ḍ̂ek^{h}-c^{h}-o	ḍ̂ek^{h}-b-e	ḍ̂ek^{h}-o

Past.3p	Past.hab.3p	Past.prog.3p	Present.prog.3p	Fut.3p	Subjunctive.3p
lik^{h}-l-o	lik^{h}-t-o	lik^{h}-c^{h}i-l-o	lik^{h}-c^{h}-e	lik^{h}-b-e	lik^{h}-uk
ʃun-l-o	ʃun-t-o	ʃun-c^{h}i-l-o	ʃun-c^{h}-e	ʃun-b-e	ʃun-uk
rak^{h}-l-o	rak^{h}-t-o	rak^{h}-c^{h}i-l-o	rak^{h}-c^{h}-e	rak^{h}-b-e	rak^{h}-uk
kor-l-o	kor-t-o	kor-c^{h}i-l-o	kor-c^{h}-e	kor-b-e	kor-uk
ḍ̂ek^{h}-l-o	ḍ̂ek^{h}-t-o	ḍ̂ek^{h}-c^{h}i-l-o	ḍ̂ek^{h}-c^{h}-e	ḍ̂ek^{h}-b-e	ḍ̂ek^{h}-uk

All non-present forms along as well as those marked with progressive aspect belong to the Sub-Paradigm B. All affixal forms that attach to the verb root are consonant-initial, with the notable example of the 2p subjunctive. The vowel of the affix has no effect on the root morpheme. The paradigm B emerges in certain CV roots.

Past.2p	Past.hab.2p	Past.prog.2p	Present.prog.2p	Fut.2p	Subjunctive.2p
ca-i-l-e	ca-i-t-e	ca-i-c^{h}i-l-e	ca-i-c^{h}-o	ca-i-b-e	ca-o *ca-i-o
k^{h}e-l-e	k^{h}e-t-e	k^{h}a-c-c^{h}i-l-e	k^{h}a-c-c^{h}-o	k^{h}a-b-e	k^{h}a-o

With respect to CV roots, the differences between Sub-paradigms B and C are not as clearly defined as it is in case of CVC roots, and I will not comment of this matter any further in this paper.

[2] The vowel [ɔ] and [ɛ] do not appear word-finally in lexical words in Bangla, and also fail to occur root finally in CV verb roots.

b) Sub-Paradigm C

Perf.1p	Perf.2p	Perf.3p	
lik^{h}-e-c^{h}-i	lik^{h}-e-c^{h}-o	lik^{h}-e-c^{h}-e	'write'
ʃun-e-c^{h}-i	ʃun-e-c^{h}-o	ʃun-e-c^{h}-e	'listen'
rek^{h}-e-c^{h}-i	rek^{h}-e-c^{h}-o	rek^{h}-e-c^{h}-e	'keep'
kor-e-c^{h}-i	kor-e-c^{h}-o	kor-e-c^{h}-e	'do'
ɖ̂ek^{h}-e-c^{h}-i	ɖ̂ek^{h}-e-c^{h}-o	ɖ̂ek^{h}-e-c^{h}-e	'see'

Past.perf.1p	Past.perf.2p	Past.perf.3p	Conjunctive participle
lik^{h}-e-c^{h}i-l-am	lik^{h}-e-c^{h}i-l-o	lik^{h}-e-c^{h}i-l-e	lik^{h}-e (eʃ-o)
ʃun-e-c^{h}i-l-am	ʃun-e-c^{h}i-l-o	ʃun-e-c^{h}i-l-e	ʃun-e (eʃ-o)
rek^{h}-e-c^{h}i-l-am	rek^{h}-e-c^{h}i-l-o	rek^{h}-e-c^{h}i-l-e	rek^{h}-e (eʃ-o)
kor-e-c^{h}i-l-am	kor-e-c^{h}i-l-o	kor-e-c^{h}i-l-e	kor-e (eʃ-o)
ɖ̂ek^{h}-e-c^{h}i-l-am	ɖ̂ek^{h}-e-c^{h}i-l-o	ɖ̂ek^{h}-e-c^{h}i-l-e	ɖ̂ek^{h}-e (eʃ-o)

Phonologically, Sub-Paradigm C is a specialized case of Sub-paradigm B where the root vowel [a], which is phonologically inert in B, also undergoes the same phonological transformation as other vowels with |A| head. Morpho-syntactically too, the Sub-paradigm B is the larger set containing all non-present tense forms as well as aspect markers. Similar to the phonology, the Perfect aspect, can also be considered as a specialized morpho-syntactic characterization which stands out from the general set B by forming a specialized paradigm C.

Elsewhere in the language as well, there is independent evidence that the Perfect is morpho-phonologically distinct from the regular Tense-Aspect morphology of Bangla. For example, it has a fused exponence with negation as seen in (29).

(29) Negation in Bangla

a) Present/Habitual

tumi	o-ke	bol-o		tumi	o-ke	bol-o	na
2p	3p-acc	tell.subj.2p		2p	3p-acc	tell.subj.2p	Neg
You tell him.				You don't tell him.			

b) Perfect

tumi	o-ke	bol-e-c^{h}o		tumi	o-ke	bɔl-o	ni
2p	3p-acc	tell.perf.2p		2p	3p-acc	tell.2p	Neg.perf
You told him.				You didn't tell him.			

Finally, the remaining or elsewhere morpho-syntactic paradigm shows non-uniform phonological exponence. When the CVC verb root is immediately followed by a inflection that begins with a [+high] vowel, the environment for regressive vowel harmony, the Root 2 is surfaces in sub-paradigm B. Everywhere else, including causatives, gerund and imperative the Root appears in Sub-paradigm A. This has been shown in (30).

(30) Elsewhere

a) Sub-paradigm B: With [High] vowels

Habitual.1p	Habitual.1p.fam	Caus.Perf	Imp.Hon	
lik^{h}-i	lik^{h}-iʃ	lik^{h}-i-e-c^{h}i-l-o	lik^{h}-un	‘write’
ʃun-i	ʃun-iʃ	ʃun-i-e-c^{h}i-l-o	ʃun-un	‘listen’
rak^{h}-i	rak^{h}-iʃ	rak^{h}-i-e-c^{h}i-l-o	rek^{h}-un	‘keep’
kor-i	kor-iʃ	kor-i-e-c^{h}i-l-o	kor-un	‘do’
d̪ek^{h}-i	d̪ek^{h}-iʃ	d̪ek^{h}-i-e-c^{h}i-l-o	d̪ek^{h}-un	‘see’

b) Sub-paradigm A

Imp.2p.fam	Imp.2p	Hab.Hon	Hab.3p	Gerund
lek^{h}	lek^{h}-o	lek^{h}-en	lek^{h}-e	lek^{h}-a
ʃon	ʃon-o	ʃon-en	ʃon-e	ʃon-a
rak^{h}	rak^{h}-o	rak^{h}-en	rak^{h}-e	rak^{h}-a
kɔr	kɔr-o	kɔr-en	kɔr-e	kɔr-a
d̪ɛk^{h}	d̪ɛk^{h}-o	d̪ɛk^{h}-en	d̪ɛk^{h}-e	d̪ɛk^{h}-a
Caus.1p	Caus.prog.1p	Caus.past.1p	Caus.prog.past.1p	
lek^{h}-a-i	lek^{h}-a-c-c^{h}-i	lek^{h}-a-l-am	lek^{h}-a-c-c^{h}-i-l-am	‘write’
ʃon-a-i	ʃon-a-c-c^{h}-i	ʃon-a-l-am	ʃon-a-c-c^{h}-i-l-am	‘listen’
rak^{h}-a-i	rak^{h}-a-c-c^{h}-i	rak^{h}-a-l-am	rak^{h}-a-c-c^{h}-i-l-am	‘keep’
kɔr-a-i	kɔr-a-c-c^{h}-i	kɔr-a-l-am	kɔr-a-c-c^{h}-i-l-am	‘do’
d̪ɛk^{h}-a-i	d̪ɛk^{h}-a-c-c^{h}-i	d̪ɛk^{h}-a-l-am	d̪ɛk^{h}-a-c-c^{h}-i-l-am	‘see’

Summarizing this sub-section on Bangla verb root allomorphy, we see that the pattern for vowel alternation cannot be simply stated with respect to either a single phonological or morphological context. The morphology of the language, cognizant of the phonological well-formedness requirements of Bangla, conspires to create phonological sub-paradigms A and B that bleed away the context for phonological rule application.

3 Discussion

In this paper we have seen that all the three Indo-Aryan languages Hindi, Odia and Bangla show vowel umlaut in verb roots. In each one of the contexts, the prime |A| plays a central role in the phonological description of the allomorphic pattern. Further, in each of the patterns the vowel repertoire gets restricted in the marked morphological environments. The “marked” morphological context, literally works to set the paradigm distinctively apart from the rest of the inflectional morphology in the verbal paradigm.

This “marking out” is witnessed clearly in the case of Hindi and Odia causatives where the umlaut simply doubles up as an additional morphological factor in a context that is already marked by a causative inflection.

(31) Double Morphological marking

	sleep.Non-fin	sleep.causative. Non-fin
Hindi	so:-na:	sul-a:-na:
Odia	so-i-ba	su-e-i-ba

These two languages also differ in the extent of such marking out within the verbal lexicon. Hindi, which has no other systematic vowel cooccurrence restrictions, applies this across the board to the morphological causatives of all verbs, while Odia which has phonological restrictions on vowel hiatus contexts, applies the umlaut to only those verbs roots CV and CaC ones, that would have undergone phonological modification anyway. In that way, Odia and Bangla are similar as the latter also uses umlaut as a means to avoid irregularity within "marked out" paradigms. Also, similar to Bangla, Hindi could have easily avoided surface homophony in the "marked" morphological paradigms by avoiding the short vowels in the non-causative and non-accusative verb roots. This would have optimized the morphological "marking out" of these paradigms even better as the transitive and unaccusative forms of all verbs would be distinct. At present the ones with the short vowels become homophonous on the surface.

(32) Hindi transitive-unaccusative homophony

Transitive				Unaccusative			
rəvi	səmoːse:	t̪ʰəl rəha:	tʰa:	səmose	t̪ʰəl	rəhẽ:	tʰẽ:
ravi	samosa.pl.	fry prog.M	be.past.M	samosa.pl	fry	prog.pl	be.past.pl
Ravi was frying the samosas.				The samosas were being fried.			

Unlike Hindi and Odia, Bangla phonology places more restrictions on vowel cooccurrence, so there is a greater need to optimize the verbal lexicon. So, the language has removed vowels without |A|, [i, u], from the phonological sub-paradigm A altogether. Most of the verbal morphology falls into sub-paradigm B where the complex vowels with |A| head fail to surface in verb roots. This group includes all morphological paradigms where the post-verbal inflection begins with a [High] vowel, as well as all the consonant-initial inflections. It also includes subjunctives though they are vowel initial and do not begin with [High] vowel. In Sanyal (2017), I argue that the subjunctives are included in the allomorphic paradigm to avoid surface homophony with the imperatives. So, we see that Bangla avoids homophony not just between lexemes within a paradigm, but also between the outputs of distinct paradigms.

(33) Homophony avoidance in Bangla

do.imperative (Sub-paradigm A)	do.subjunctive (Sub-paradigm B)
kɔr-o	kor-o

Finally, as part of the concluding remarks a word about the distinct patterns between CVC and CV verb roots is warranted. Most of the verb roots in all Indo-Aryan languages are CVC monosyllables with a handful of CV and disyllabic roots. Ideally, CVC roots should not have much phonological issue with morphological suffixation as the final C will get syllabified tauto-syllabically and hetero-syllabically with C-initial and V-initial suffixes. In case of Hindi, this is not a major concern as most inflections are independent clitics and not suffixes. In Bangla the verbal inflections are all suffixes and could be either C-initial

or V-initial. All the C-initial suffixes fall into the same morphological sub-paradigm, but the V-initial ones show divergence.
CV verb roots in all three languages have non-uniform morphological paradigms. In Odia, this non-uniformity in verb root exponence is restricted to non-causatives, but in Bangla and Hindi all paradigms of CV are somewhat irregular.

(34) Irregularities in CV paradigms

a)	Hindi	kʰa:-na: 'to eat'	kʰil-a:-na:	'to feed'	*kʰəl-a:-na:
		ga:-na: 'to sing'	gə-wa:-na:	'to make sing'	*gil-a:-na:
b)	Bangla	kʰa-wa 'to eat'	kʰe-ṭ-am	'eat.past.hab'	*kʰa-t-am
		ga-wa 'to sing'	gai-ṭ-am	'sing.past.hab'	*ge-ṭ-am

Disyllabic verb roots are those with a morphological causative built into it. For example, the Hindi verb root [bula:-na:] 'to call', has not non-causative root [bul-na:]. In Bangla, such forms usually surface with two possible pronunciations and have been discussed extensively in Dasgupta (2018).

(35) Alternative pronunciations for denominal verbs in Bangla
kor-a-no ~ kur-o-no 'to lift from the ground'
ceb-a-no ~ cib-o-no 'to chew'

In conclusion, the key task is to imagine a theory of morpho-phonological interface where morphological well-formedness factors such as Paradigm Uniformity and Anti-Homophony are being balanced with core phonological distributional restrictions to arrive at Optimal Paradigms. This paper is just a small step in that direction.

Acknowledgements

I am grateful to Anasuya Nayak for consultation on the verbal paradigm of Odia.

References

Althaus, Nadja, Sandra Kotzor, Swetlana Schuster, Aditi Lahiri. 2022. Distinct orthography boosts morphophonological discrimination: Vowel raising in Bengali verb inflections, *Cognition*, Volume 222.
Bhatt, Rajesh, and David Embick. 2017. Causative Derivations in Hindi-Urdu. *Indian Linguistics* 78. 93-151.
Calcutta, New Delhi: Rupa & Co. Original published 1926, Calcutta University.
Charette, Monik. 1991. *Conditions on Phonological Government*. CUP, Cambridge.
Chatterji, Suniti Kumar. 1926/1978. *The origin and development of the Bengali language*
Crosswhite, Catherine. 2001. *Vowel reduction in Optimality Theory*. New York and London: Routledge.

Dan, Mina. 1998. Bangla verb morphology: The actual derivation. *Indian Linguistics* 59 43-79.
Dasgupta, Probal. 1982. Phonology and the Bangla Verb. *Indian Linguistics* 43.17-28.
Dasgupta, Probal. 2000. The verb stem formative /a/ in Bangla. In R. E. Asher and R Harris, eds., *Linguisticolliterary: A Festschrift for Professor D.S. Dwivedi*, 115-125. Delhi: Pilgrim.
Dasgupta, Probal. 2018. On Certain Vowels in the Substantivist Morphology of Bangla. *Indian Linguistics* 79. 23-43.
Guru, Pratikshya, Anusuya Nayak. 2024. Microvariation in the distribution of /o/. Paper presented at the Himalayan Languages Symposium (HLS-27), June 12-14, IIT Guwahati, India.
Harris, John. 1990. Segmental Complexity and phonological government. *Phonology* 7, 255-300
Harry, Otelemate, and Larry M. Hyman. 2014. Phrasal Construction Tonology: The case of Kalabari. *Studies in Language* 38 649-689.
Ichimura, Larry K. 2006. *Antihomophony blocking and its productivity in transparadigmatic relations*. Doctoral Dissertation: Boston University.
Kaye, Jonathan, Jean Lowenstamm, Jean-Roger Vergnaud. 1985. The internal structure of phonological representations: a theory of charm and government. *Phonology Yearbook* 2, 305-328.
Kaye, Jonathan, Jean Lowenstamm, Jean-Roger Vergnaud. 1990. Constituent structure and government in phonology. Phonology 7, 193-231.
Kenstowicz, Michael. 1996. Base Identity and Uniform Exponence: Alternative to Cyclicity. In Jacques Durand and Bernard Laks eds., *Current Trends in Phonology: Models and Methods*, Volume 1, by, 363-393. University of Salford.
Kenstowicz, Michael. 1998.Uniform Exponence: Exemplification and Extension. Rutgers Optimality Archive. https://rucore.libraries.rutgers.edu/rutgers-lib/39727/PDF/1/play/
Kiparsky, Paul. 1982. Analogical change as a Problem for Linguistic Theory. In *Explanation in Phonology*, De Gruyter, 217-236.
Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook*, volume 2, 82-136.
McCarthy, John J. 2005. Optimal Paradigms. In Laura Downing, Tracy Hall and Renate Raffelsiefen, eds., *Paradigms in Phonological Theory*, 295-371. Oxford: OUP.
McCarthy, John J., and Alan Prince. 1994. The Emergence of the Unmarked. NELS 34. Amherst: University of Massachusetts. 333-379.
Paul, Jayashree. 1986. Bangla verb morphology: a reconstruction. *Indian Linguistics* 47 73-79.
Sanyal, Paroma, Vyom Sharma, and Ankita Prasad. 2021. Hindi root allomorphy: Insights from phonological and morpho-syntactic theory. In Ghanshyam Sharma and John J. Lowe, eds., *Trends in South Asian Linguistics*, 381-414. De Gruyter Mouton.
Sanyal, Paroma. 2011. *Prosodic Well-formedness and vowel licensing in Bangla*. Doctoral dissertation. EFL University. Hyderabad.

Sanyal, Paroma. 2017. Anti-Identity constraints in morphological paradigms. Paper presented at the Formal Approaches to South Asian Languages (FASAL) 7, Massachusetts Institute of Technology, Cambridge USA, March 2017.
Smith, Jennifer. 2011. Category Specific Effects. In Marc van Oostendorp, Colin Ewen, Beth Hume and Keren Rice, eds., *The Blackwell companion to phonology*, by, 2439-2463. Malden, Massachusetts: Wiley-Blackwell.
Steriade, Donca. 2016. "The morphome vs Similarity-based syncretism." In Ana Luis and Ricardo Bermudez-Otero, eds., *The Morphome Debate*, 112-172. Oxford: OUP.
Steriade, Donca. 2023. Paradigm Structure in Sanskrit Reduplicants. In Darya Kavitskaya and Alan C. L. Yu, eds., *The Life Cycle of Language: Past, Present and Future*, OUP.

Malayalam-Kannada Code-mixing

PAROMA SANYAL, *Indian Institute of Technology Delhi*
THERESE LIAM TOM, *Indian Institute of Technology Delhi*

ABSTRACT

This paper presents the results of a pilot study undertaken to see the possible limits of code-mixing among Malayalam-Kannada bilinguals in a syntactic context of featural mismatch. The results of the study reiterate that the fact that despite being influenced by psycholinguistic factors like being primed with the language of the task paragraph, simultaneous (balanced) bilinguals can make the choice of strategy based on structural factors, but sequential bilinguals do not access syntactic structure while making these decisions. The latter's choice of strategy is based either on the acquisitional factor of MT/non-MT or on psycholinguistic factors from the task design.

1 Introduction

A bilingual person can mix two languages within a single syntactic construction. This shows that the syntactic machinery offers the possibility of combining the syntactic frames from different languages. (1) below shows such a possibility where the English and Malayalam frames have been mixed.

(22) [I told John [that *avande co:r njaan kazhichu*]
I tell.PST John COMPL he-GEN rice I eat.PAST
'I told John that I ate his rice'

Here the matrix clause is in English, and the embedded proposition (*His rice, I ate*) is in Malayalam. Note that the Complementizer embedding the Malayalam clause, is in English, the language of the matrix sentence.

In addition to clausal level of mixing, bilinguals also insert lexical items (Muysken 2000) from one language into the syntactic frame of the other language as seen in (2)

(23) I eat *co:r* in the morning.

Here, the Malayalam lexical item, *co:r*, 'rice' replaces the English word, *rice* in an English frame sentence. Henceforth, in this paper we refer to the process in (1) as code-switching, and the one in (2) as code-mixing[1].

Neither (1) nor (2) require interactions among the functional features of the two languages since there are no overt functional features from English, the frame language, that are syntactically dependent on the Malayalam part. However, if the code-mixed lexical

[1] The terms code-mixing and code-switching have been used in literature (MacSwann 2014, Muyksen 2000, Grimstad et al. 2018, Myers-Scotton 1993) to mean a number of different kinds of mixing patterns between languages. Sometimes they include loan-words and sometimes they do not. For clarity and consistency, we are making our use of the terms transparent.

item needs to syntactically interact with a functional head from the other language there is a possibility for feature mismatch between them. This potential mismatch between the functional structures of two different languages is the scope of enquiry for this paper.

In naturalized environments, balanced bilinguals, who are proficient in both languages, are likely to avoid code-mixing in circumstances where there is a potential mismatch in syntactic characteristics between the two languages. So instead of looking for well-formed cases of naturalized code-mixing within a bilingual corpus, we are eliciting code-mixed data from bilingual participants in our study. This is a pilot study designed to understand what people would intuitively do if they had to code-mix in a certain way and had no time restriction to produce their response.

1.1 The participants in the code-mixing study

For this study on code-mixing, we have selected two Dravidian languages, Malayalam and Kannada, because they are lexically and syntactically similar, and have some structural distinctions that could produce mismatch of functional features in code-mixing contexts. Since the two languages are spoken in neighboring geographical regions, one would have expected balanced bilinguals among both Malayalam and Kannada speakers. However, that is not the case since the border region between the states of Kerala (where Malayalam is the official language) and Karnataka (where Kannada is the official language) has several other Dravidian languages such as Tulu, Kodagu etc. Residents of this region speak both Kannada and Malayalam but neither of them natively. Consequently, we have not tried the elicitation with people from this region.

All the participants in our study are Malayalam speakers who are long-time residents of Karnataka, Bangalore city to be specific, and use Kannada as a lingua franca. Ideally, we should also have had another set of participants who are Kannada speakers living in Kerala for a long time. However, due to the directionality of population migration, such a demographic is not readily available. Consequently, we have created two subsets, shown in (3), within the Malayalam-Kannada bilingual population available to us.

(3) Sequential and Simultaneous Bilinguals

Sequential Bilinguals	Participants who acquired L1 in childhood, and later acquired L2 as an adult.
Simultaneous Bilinguals	Participants who acquired both L1 and L2 in childhood

Both the groups of Malayalam-Kannada bilingual participants were presented with the same elicitation task. In the section 4 of the paper, where we analyze the sentences produced by them, we will see if the two groups are using divergent strategies to handle feature mismatch.

1.2 The mismatch contexts in the code-mixing study

Ordinarily, inflectional morphemes do not attach to borrowed words directly. Instead, the borrowed word is embedded within another lexical item from the frame language. This is observed in (4) where the English verb form 'buy' combines with the Malayalam tensed "do" verb *cejjudu* to fit into the syntactic frame of the Malayalam tensed clause.

(4) ɲa:n dosha **bɑɪ** cejjudu
I dosa buy do.PST
'I bought dosa'

Here, the English verb form 'buy' cannot attach directly with the tense morphology of Malayalam, and thus requires 'do support'. Will the same kind of restriction hold for code mixing between Malayalam and Kannada which share lexical and syntactic similarities?

Theoretically, one prominent distinction between the verbal inflections of Kannada and Malayalam is that the former has subject agreement for person and number while the latter has no agreement marking. Examples (5) and (6) demonstrate this.

(5) Kannada

a. naanu mane-ge hogid-**e**
1Sg home-DAT go.PST.1Sg
'I went home.'

b. avalu mane-ge hogid-**alu**
3SgF home-DAT go.PST.3Sg.F
'She went home.'

(6) Malayalam

a. njaan viitil-ottu pooji
1Sg home-DAT go.PST
'I went home.'

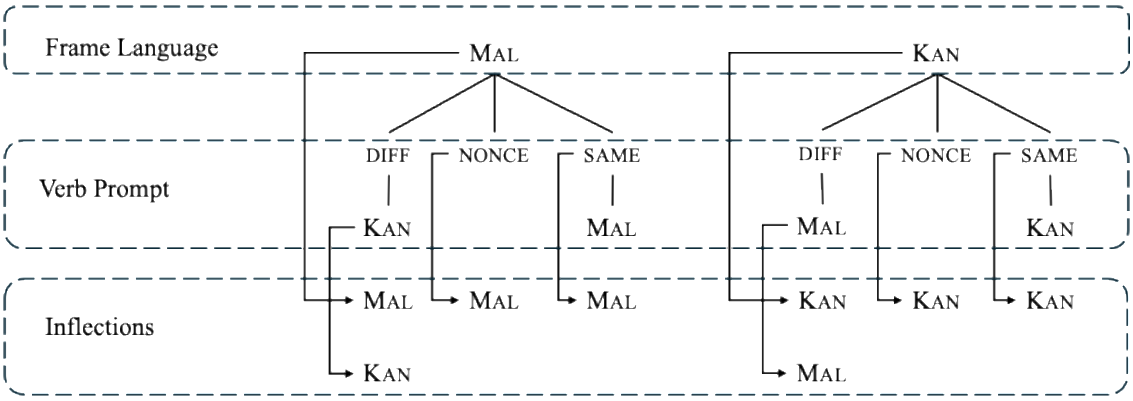
b. aval viitil-ottu pooji
3SgF home-DAT go.PST
'She went home.'

We designed our elicitation paradigm using this distinction between the two languages by inserting a blank in the place of specific verbs which the participant had to inflect and use. The language of the frame sentence was distinct from the language of the lexical verb prompts. For example, a specific verb is replaced with blanks within a Kannada text paragraph and the participant is prompted with the Malayalam lexical verb root rather than the Kannada one. While reading the text, the participant has a free choice to use 'do support' from Kannada, or inflect the Malayalam verb with either Kannada or Malayalam inflections. The same process is done with Kannada verb prompts and Malayalam Text.

Further, to determine whether the strategy used results from a preference for a particular option or a dispreference for the other, we have used a control set of nonce

prompts for both Kannada and Malayalam text contexts. If there was no interference from the inflectional paradigm of the language of the verb prompt, there should be no distinction between the choice of strategy between the nonce and verb prompt contexts. This entire paradigm along with the potential outcomes has been schematically presented in (7).

(7) Schematic presentation of the code-mixing contexts



The use of Kannada inflection in the DIFF verb prompt context has the further matter of whether the Kannada agreement markers have been used along with TAM inflections or if they have been left out, similar to Malayalam.

In the following section, we elaborate on theoretical reasons to distinguish between TAM inflections and Agreement markers. Following that is a section elaborating on the specifics of the elicitation task design and the results obtained.

2 The syntax of verb inflections

The theory of generative syntax (Chomsky 1970, 1981, 1995) has always noted the distinction between lexical and functional items in the vocabulary. While lexical items are independent meaningful words with corresponding phonological strings that are stored in the lexicon, functional items may or may not have phonological strings associated with them and have "meaning" only in the context of the syntactic arrangement of which they are a part. Within generative syntax, there are two broad ways in which these distinctions are analyzed. The two approaches are commonly referred to using the terms Lexicalist (Chomsky & Lasnik 1993) and the Non-lexicalist (Halle & Marantz 1993, Borer 2003, 2015, Embick & Noyer 2007) and hereafter we will use the same.

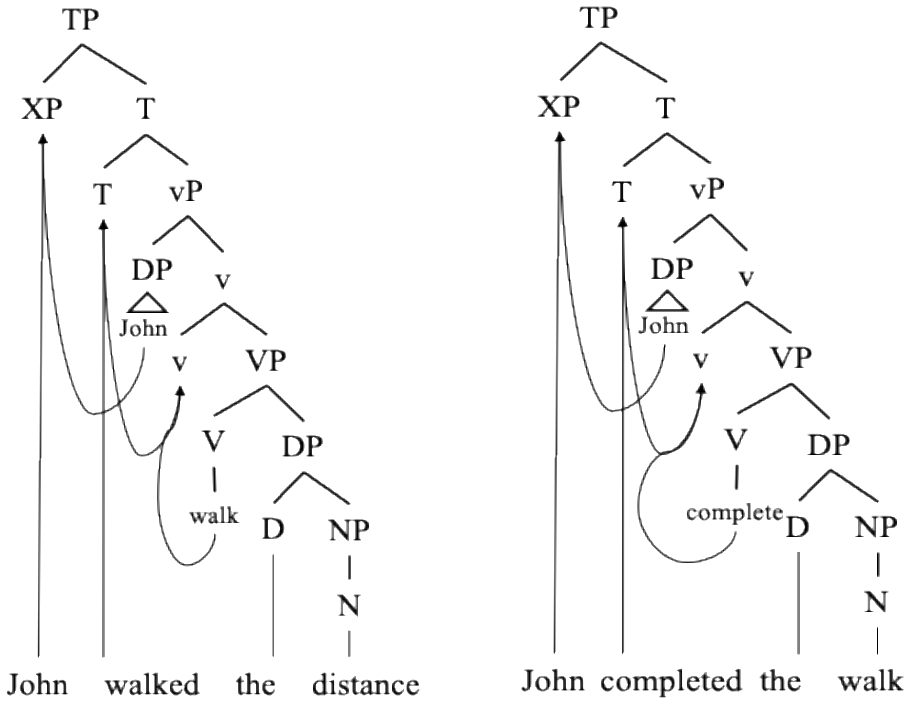
In the Lexicalist approach syntactically categorized lexical items are stored in the lexicon and enter the syntactic module by projecting a syntactic structure headed by the lexical item. Each such structure then MERGEs with further functional heads that host the syntactic and semantic features associated with that lexical head. For example, a Noun would further project functional heads corresponding to Number, Gender, Definiteness etc and a Verb would project functional heads corresponding first to the event structure and

then the tense, aspect, mood (TAM). This latter T(ense) head is also argued to be the functional head that contains uninterpretable ϕ features corresponding to person, number and gender. These are cancelled by matching them with the corresponding interpretable feature values present within a nominal projection via the operation AGREE. So, the difference between a language with overt ϕ agreement, like Kannada, and a language without ϕ agreement, like Malayalam, rests in the nature of the functional head T that is projected from the lexical verb of the language. A Kannada T has uninterpretable ϕ features that need agreement, while a Malayalam T does not have uninterpretable ϕ features and therefore does not need agreement.

In the non-lexicalist approach the general structure within the syntax and the mechanisms of structure building through MERGE and ϕ agreement via AGREE are the same as in lexicalist approach. The difference lies in the way the syntactic module interacts with the lexicon. First, the lexical items are syntactically distinct from the computational structure generated in narrow syntax. The entire syntactic hierarchy is generated on the basis of the syntactic features present on functional heads within the syntactic module. The lexicon contains category-less √Roots that gain their syntactic category as well as specific contextual meaning with respect to the syntactic structure they MERGE under. This fundamental difference between the two theoretical approaches is shown in (8) and (9).

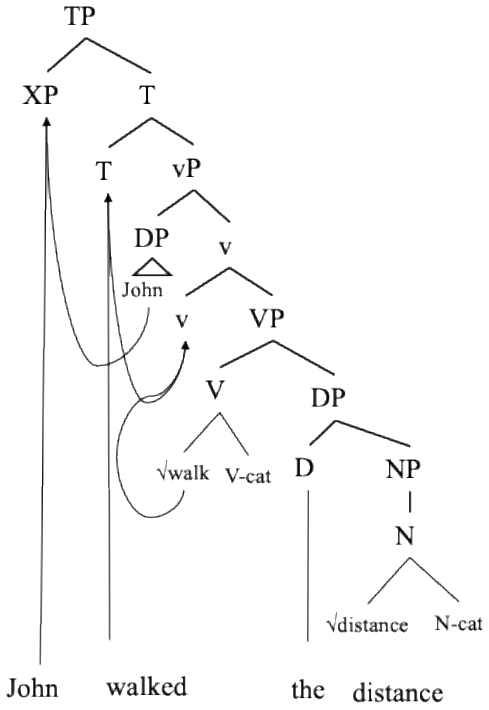
(8) Schematic presentation syntactic structure in the lexicalist approach

a. $[walk]_{verb}$ b. $[walk]_{noun}$

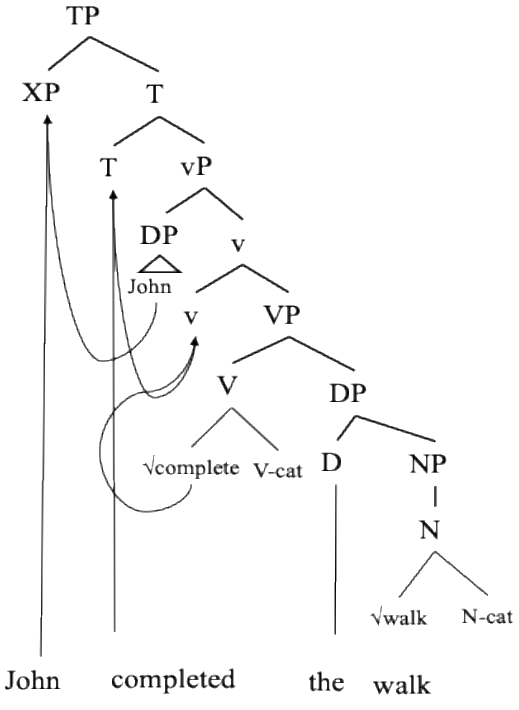


(9) Schematic presentation syntactic structure in the non-lexicalist approach

a. [√walk, v-cat]



b. [√walk, n-cat]



In (8) and (9) we see the first major distinction between these two theoretical approaches. In the lexicalist framework there are two different lexical entries corresponding to the phonological string [walk], one corresponding to verbal meaning and the other to the nominal one. In contrast, the lexicon contains just one single root √walk that corresponds to the both the nominal and the verbal outputs in the non-lexicalist framework. When the root merges with a n-categorizer under a NP-DP frame it gets a nominal meaning, and when it merges with a v-categorizer under a VP-TP frame it gets a verbal meaning.

These two theoretical approaches would predict different outcomes in the code-mixing context. If the syntactic structure is projected based on the features of the lexical item, then the Kannada verb will project a T with uninterpretable ϕ features while the Malayalam verb will project a T without it. The probe from this Kannada T will look for ϕ agreement with functional features within the Malayalam DP. The Malayalam DP will not have those functional heads since there is no ϕ probe in Malayalam. The structure will crash. On the other hand, if the syntactic structure is independent of the lexical choice the Kannada lexical item can nest under a Malayalam frame with Malayalam inflections and vice versa.

A second distinction between these two theoretical approaches also has crucial implication for our study. In the lexicalist approach all morphological derivations happen before the syntax and these lexical items already contain the phonological material

associated with inflections. In the non-lexicalist approach insertion of phonological material correlating to functional heads (non-roots) happens post-syntactically through an operation called Vocabulary Insertion (Embick, 2015). Apart from the un-categorized roots the lexicon also contains another set of phonological strings called vocabulary items (VI). Each VI corresponds to a syntactic substring containing a set of syntactic features. The same feature could be part of the defining correspondence of more than one VI. When that happens the Vocabulary Insertion progresses by disjunctive rule ordering, i.e. the more specific rule gets ordered above the less specific one. This has been demonstrated in (10).

(10) Past tense allomorphy in English

Suppletive past	Regular past
[sat]↔[past]/√sit [went]↔[past]/√go (…)	[-ed]]↔[past]

Each of the suppletive [past] VI insertion rule is specified for a particular set of roots and is therefore more specific than the regular past tense VI insertion rule. Like English, Malayalam also has multiple allomorphs corresponding to the syntactic feature [past]. There are two past tense markers [-i] and [-u] and both are specified for two distinct sets of roots.

There is a potential phonological pattern to the morphological contexts for marking the past tense with [-i] or [-u] in Malayalam. Malayalam has a phonemic contrast between voiceless geminate stops and voiced singleton stops in word medial contexts. There is a general tendency to geminate the stops before attaching the [-u] past marker and voice the stop consonants before attaching [-i]. However, this is not a clear case of complementary distribution since gemination can be seen in the context of past marking with [-i] as well. The corresponding data is shown in (11) (p.c and insights from Greeshma Joseph).

(11) Phonological context of past tense allomorphy in Malayalam

[past] marking with [-i]		[past] marking with [-u]	
paaḍ-i	sing	irunn-u	sit
caaḍ-i	jump	paranj-u	say
karakk-i	rotate	tot̪t̪-u	touch
uruṭ̪t̪-i	roll	koduṭ̪t̪-u	give

However, since voicing does not seem to accompany [-u] marking, we make the hypothesis that [-i] is the specific rule and [-u] the general one. This will be borne out in our elicitation task where the participants are theoretically predicted to use the general rule in case of nonce or novel conjugation. The disjunctive ordering for these is shown in (12).

(12) Past tense allomorphy in Malayalam

Suppletive past	Regular past
[-i]↔[past]/(√sing, √jump, √go, …)	[-u]]↔[past]

To summarize the discussion in this section so far, the lexicalist framework would predict that the TAM inflections would match the language of the verb prompt rather than the language of the remaining frame. In case of Kannada, the structure will crash since the uninterpretable ϕ features on the Kannada T would not be valued. But since the syntactic structure is literally generated out of the lexical verb given, the numeration of the participants will fail to generate an output with the Kannada verb prompt in the mismatch context, without "do support" from Malayalam. In contrast, the Malayalam verb prompt will not have this issue due to the absence of ϕ agreement. Participants will be able to produce the inflected Malayalam verbs in the Kannada frame.

The predictions from the non-lexicalist frameworks like the Exoskeletal approach (Grimstad et al., 2018) would differ. The functional projection of a Kannada V is compatible with a functional projection of D that has a Malayalam noun at its base. This is because the syntactic frame into which the Malayalam root is merged to form a noun is not generated out of the lexical item. It could contain the functional projections for person and number that could AGREE with the ϕ probe from the T. Similarly, the functional projection of a Kannada V need not have a Kannada lexical item at its base. So, a Malayalam verb prompt could end up with Kannada inflections including ϕ agreement in a Kannada frame paragraph.

Since the study design is open-ended and could result in varied outcomes, we also need an evaluative framework to analyze the responses. For this we are using a constraint-based evaluative framework. In the following section we elaborate of this evaluative framework.

3 The evaluative framework for analysis

Constraint based evaluative frameworks like Harmonic Grammar (Legendre et al., 2022) and Optimality Theory (Prince and Smolensky 1993) are used to analyze contexts where multiple factors are at play, to determine an outcome. These factors are termed as constraints. In Optimality Theory, the constraints have three core properties: violability, rankability, and universality. Of these we are going to use only the first two for our evaluative purpose. Since we are evaluating a performance output, all our constraints are not derived out of the formal aspects of the language module. In such a context, universality would be an inappropriate claim.

Violability refers to the fact that all constraints are in principle violable, and rankability means that all constraints are freely rankable with respect to all other constraints. To elaborate, suppose we are evaluating three candidate outputs A, B and C with respect to the constraints P, Q and R. Given that A violates P and Q, B violates Q and C violates P and R, the only possible outcomes are B and C and these are shown in (13).

(13) Evaluation Table for the possible outputs

a. Equal weights

	Con P	Con Q	Con R	Harmonic
	w=1	*w=1*	*w=1*	Value
Cand A	-1	-1		-2
☞Cand B		-1		-1
Cand C	-1		-1	-2

b. Q weighed up

	Con Q	Con P	Con R	Harmonic
	w=3	*w=1*	*w=1*	Value
Cand A	-1	-1		-4
Cand B	-1			-3
☞Cand C		-1	-1	-2

In (13) all the constraints are assumed to weights (*w=x*). Given the violations where each violation is denoted [-1], the harmonic value for each candidate, summation of violation and weight, is calculated for each candidate. The one with the highest value is the optimal output for that table. In (13a) where all the constraints are assumed to have equal weights, B surfaces since it has the highest harmonic value on account of minimum number of violations. Note that the violations of B being a proper subset of the violations of A, A is harmonically bound by B and will always be blocked by it from surfacing. On the other hand, C which is not harmonically bound, and also incurs two violations, can surface as the output if the weight of constraint Q is greater than the sum of the weights of P and R. This is seen in (13b).

The possible output candidates in our Kannada-Malayalam code-mixing set up are listed in (14).

(14) Candidate outputs

Denotation	DP Frame	Lexical Verb	TAM inflection	Agreement
a. MKM	M	K	M	
b. KMKK	K	M	K	K
c. MKKK	M	K	K	K
d. KMM	K	M	M	
e. MKK	M	K	K	
f. KMK	K	M	K	

(14a) is a context where a Kannada root is merged under a Malayalam VP frame and is comparable to (14b) where a Malayalam root is merged under a Kannada VP frame. Similarly, in (14c) and (14d) a Malayalam and Kannada root is merged under a Kannada and Malayalam DP frame. Technically, these four should have been the only contexts we should have expected. However, since our participants are native Malayalam speakers with

Kannada as a L2 there is a possibility that some of them would not have acquired the agreement paradigm of Kannada very well. Thus (14e) and (14f) are output candidates with a Kannada VP frame where the T does not have ϕ features.

Determining the nature of the potential constraints that influence the outcome is more challenging than determining the possible output candidates. We have divided these constraints into three sub-types based on our experimental paradigm. The first set of constraints are based on structure-based predictions from syntax and morphology, the second set from acquisition related factors, and the third considers psycholinguistics factors that might be induced by the elicitation task design. These have been described in (15).

(15) Potential evaluative constraints

a. Structural Constraints

Max F	*Project the Syntactic frame with maximal functional features* This constraint states that, given an option between two syntactic structures the one with the greater number of functional heads will be preferred than the one with the lesser one. This means both DP and VP frame choice will prefer Kannada over Malayalam.
Min F	*Project the Syntactic frame with minimal functional features* This constraint states that, given an option between two syntactic structures the one with the lesser number of functional heads will be preferred than the one with more features. This means both DP and VP frame choice will prefer Malayalam over Kannada.

b. Acquisition Factors

MT Faith	*Project the Syntactic frame of MT* This constraint states that, given an option between two syntactic structures where one of them is the mother tongue of the participant, there is a positive bias towards the MT frame. This means both DP and VP frame choice will prefer Malayalam over Kannada.
MT Anti-faith	*Do not project the Syntactic frame of MT with non-MT verb prompt* This constraint states that, given an option between two syntactic structures where one of them is the mother tongue of the participant, there is a negative bias towards the MT frame. This means both DP and VP frame choice will prefer non-Malayalam over Malayalam.

c. Psycholinguistic Factors

Match I-FT	*The language of the Inflection must match the language of the frame text*

This constraint states that given an option between two syntactic structures where one of them matches the context, there is a positive bias towards that frame. This means the VP frame choice will prefer the language of the context.

Match I-VP *The language of the Inflection must match that of the verb prompt*
This constraint states that given an option between two syntactic structures where one of them matches the root of the verb prompt, there is a positive bias towards that. This means the VP frame choice will prefer the language of the verb prompt.

While the constraints in (15b) and (15c) are self-explanatory, those proposed in (15a) warrant some additional discussion. The syntactic features of case and agreement have historically been linked to other syntactic features such as EPP. In the light of empirical data from languages with Multiple Agree, Addressee Agreement, and Split Agreement, agreement as a phenomenon is no longer linked with any particular syntactic position for the DP with the interpretable features. Further, Adger (2003) proposes a further distinction between these set of features. The semantically interpretable features such as number and definiteness do not get deleted after the match, but semantically vacuous purely syntactic features such as case and gender get deleted after the match. This [+/- interpretable] factor has been used by L2 acquisition researchers (Bel 2003, Díaz et al. 2008, Hulk & Müller 2000, Sorace 2003, Tsimpli 2001, Tsimpli & Dimitrakopoulou 2007, Tsimpli & Mastropavlou 2008) to observe that [-interpretable] features like gender and case tend to be problematic for L2 acquisition in a way that [+interpretable] features are not.

In the context of our study, since the agreement in Kannada involves only the [+interpretable] features that need to be acquired in L2, we are presuming that Max F will add to the depth of semantic information encoding, while Min F will prefer minimal semantic depth.

4 The elicitation task

In the code-mixing data elicitation task, the participant had to read out a paragraph of text from the screen. The sentences in the paragraph had blanks in place of the inflected verbs. The verb root was provided in brackets next to the blank (Figure 1) and participants were asked to inflect the verbs appropriately while reading out.

ನಮ್ಮ ಶರೀರವನ್ನು ಉತ್ತಮ ಅರೋಗ್ಯವಂತವಾಗಿ ಇಡಲು ನಾವು ವ್ಯಾಯಾಮ ಮಾಡಬೇಕು. ನಾನು ಪ್ರತಿದಿನ ಸಾಯಿಂಕಾಲ ನನ್ನ ಮನೆಯ ಪಕ್ಕ ಇರುವ ಪಾರ್ಕಿನಲ್ಲಿ ___ [nadi] ಹೋಗುತ್ತೀನಿ. ನನ್ನ ಗೆಳೆಯರು ನನ್ನ ಜೊತೆ ಪ್ರತೀ ದಿನ ಅಲ್ಲಿಯೇ ____[nadi]. ಹಿಂದಿನ ವಾರ ನನಗೆ ಜಾಸ್ತಿ ಕೆಲಸ ಇರಲಿಲ್ಲ ಅದಿಕೆ ಇನ್ನು ಸಮಯ ಸಿಕ್ತು. ಅದಿಕೆ ಆ ದಿನಗಳಲ್ಲಿ ನಾನು ದಿನ ಎರಡು ಸತಿ ____[nadi]

Figure 1. Reading Illustration.

Each participant read out six such paragraphs. While three of these paragraphs had Kannada as the context Language (Set A), three of them has Malayalam (Set B). For each set, the first paragraph has a verb prompt from the same language, the second from the other language and the third has a nonce prompt. Each paragraphs had three blanks that corresponded to an infinitive, present-habitual and a past tense marker. This task design is schematically shown in (16).

(16) Elicitation Task Design

Set	A			B		
Language of the Paragraph	K			M		
Language of Verb prompt	K	M	N	M	K	N
TAM context for Verb Prompt	NF	NF	NF	NF	NF	NF
	Hab	Hab	Hab	Hab	Hab	Hab
	Past	Past	Past	Past	Past	Past

The participant set had 10 Malayalam-Kannada bilinguals, 5 of whom were simultaneous bilinguals (SM), and the other 5 sequential bilinguals (SQ). The SM participants belonged to the age group 18-30, and the SQ participants belonged to the age group 25-50. The SM group includes some children of the migrant SQ population. Hence the age-groups intersect, but do not match.

The participants were first screened for fluency in both Malayalam and Kannada. Further, they were also asked about the age of acquisition of Kannada. For the reading elicitation task, they could choose to read the script in Kannada, Malayalam, or Roman. Accordingly, they were shown the task paragraphs in the script of their choice. The verb prompt was always in the roman script to maintain uniformity.

While most SQ bilinguals chose to read all the paragraphs in the Malayalam script, most SM bilinguals chose the Roman script for the Malayalam and the Kannada script for the Kannada paragraphs. This asymmetry shows that most participants chose a script that they learnt as part of their school education. Most SM bilinguals studied Kannada as part of their schooling in Bangalore. One SM bilingual chose to read both languages in Roman script. In the following section, as we present the results of the elicitation choice, we will see if the choice of script had some impact in priming for the syntactic frame of the language. Further, it might also have an impact on the priming for the vocabulary item irrespective of the frame.

5 Results and analysis

The entire set of 60 responses from the 10 Malayalam-Kannada participants in the code-mixing contexts KM_ and MK_ has been presented in (17). In this section two sets of participants SQ and SM are not differentiated. Once the overall analysis of the inflection

choice has been presented, we will discuss the differences in choice between the two populations in section 6.

(17) Results

Set	A			B		
Language of the Paragraph	K			M		
Language of Verb prompt	M *paDi* 'study'			K *kari* 'call'		
TAM context for Verb Prompt	NF	Hab	Past	NF	Hab	Past
Suffix K	*-yakke*	*-yutta:re*	*-ide*	*-yakke*	*-yutta:re*	*-daru*
Suffix M	*-kkya:n*	*-kkyum*	*-ccu*	*-kkya:n*	*-kkyum*	*-ccu*
SQ1	M	M	M	M	M	M
SM2	M	M	M	M	M	M
SM1	M	M	M	K	K	K
SQ4	M	M+K	~K	K	K+M	K+M
SM3	K	K	K	~M	~M	M
SM4	K	K	M	~M	~M	M
SQ5	K	T	~K	#M	#M	#M
SQ2	K	~K	K	M/K	M/K	M
SQ3	K	~K	~K	K	~M	K
SM5	K	K	~K	K	K	K

The results from the 10 participants in the study can be broadly categorized into four types. The diacritic (~) is used K to indicate absence of or incorrect agreement marking. With M the diacritic (~) shows that the affix initial [k] was not geminated. The diacritic (#) with M indicates that the participant reinterpreted the Kannada verb prompt as a similar sounding but contextually inappropriate verb from Malayalam and added Malayalam inflection to it.

The first group (SQ1 and SM2) consistently used the Malayalam inflectional form with both Kannada and Malayalam verbs ignoring the language of the paragraph altogether. We refer to them as the "pakka mallus". By ignoring the language of the paragraph, they bypassed the problem of ϕ agreement altogether. The evaluation table in (18) shows that either Min F or MT Faith weighed up by a factor *(+x)* would result in such an output. Weighing up Min F would indicate that the choice is syntactically driven while weighing up MT Faith would indicate that the choice is an acquisition related phenomenon. Since one of the participants is a sequential bilingual and the other a simultaneous bilingual, it is possible that both of them conflated at the same result due to different factors.

(18) Evaluation Table for uniform M inflection

KM_	Min F	MT Faith	Max F	MT anti-Faith	Match FT	Match VP	HV
	w=1(+x)	*w=1(+x)*	*w=1*	*w=1*	*w=1*	*w=1*	
☞KMM			-1	-1	-1		-3
KMK	-1	-1				-1	-3+*(-x)*

MK_	Min F	MT Faith	Max F	MT anti-Faith	Match FT	Match VP	HV
	w=1(+x)	*w=1(+x)*	*w=1*	*w=1*	*w=1*	*w=1*	
☞MKM			-1	-1		-1	-3
MKK	-1	-1			-1		-3+*(-x)*

The second group (SM1 and SQ4) show a slight inter-participant variation. While both prefer to inflect in the language of the verb prompt SQ4 shows a further distinction between non-finite and finite contexts. This distinction, indicates a clear impact of the syntactic frame since ϕ agreement needs to appear only in finite contexts. In such scenario, SQ4 shifts to a "do support" structure. The evaluation tables for these two participants are shown in (19) and (20) respectively.

(19) Evaluation Table for SM1

KM_	Match VP	Min F	Max F	MT Faith	MT anti-Faith	Match FT	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
☞KMM			-1		-1	-1	-3
KMK	-1	-1		-1			-3+*(-x)*
MK_	Match VP	Min F	Max F	MT Faith	MT anti-Faith	Match FT	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
MKM	-1		-1		-1		-3+*(-x)*
☞MKK		-1		-1		-1	-3

(20) Evaluation Table for SQ4

a. Non-finite context

KM_	Match VP	Min F	Max F	MT Faith	MT anti-Faith	Match FT	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
☞KMM					-1	-1	-2
KMK	-1			-1			-2+*(-x)*
MK_	Match VP	Min F	Max F	MT Faith	MT anti-Faith	Match FT	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
MKM	-1				-1		-2+*(-x)*
☞MKK				-1		-1	-2

b. Finite context

KM_	Match FT	Match VP	Min F	Max F	MT Faith	MT anti-Faith	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
KMM	-1			-1		-1	-3+*(-x)*
☞KMK		-1	-1		-1		-3

MK_	Match FT	Match VP	Min F	Max F	MT Faith	MT anti-Faith	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
☞MKM		-1		-1		-1	-3
MKK	-1		-1		-1		-3+*(-x)*

There is no distinction in functional features on the Malayalam and Kannada in the non-finite context, so Min F and Max F are not evaluated in (20a).

The third set (SM3, SM4, SQ5 and SQ2) includes participants who have preferred to inflect in the language of the text paragraph in both finite and non-finite contexts. Of these, the outputs of SM3 and SM4 are near identical, with just one variance. SM4 chose to uniformly use Malayalam past inflection, even when the text paragraph was Kannada. Given our earlier discussion of the Malayalam past suppletive morphology (see 11 and 12 for reference), we find that the general rule for Malayalam past vocabulary insertion is ordered over the regular past in Kannada. The Evaluation table for SM3 and SM4 is shown in (21).

(21) Evaluation Table for SM3 and SM4

KM_	Match FT	Match VP	Min F	Max F	MT Faith	MT anti-Faith	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
KMM	-1			-1		-1	-3+*(-x)*
☞KMK		-1	-1		-1		-3
MK_	Match FT	Match VP	Min F	Max F	MT Faith	MT anti-Faith	HV
	w=1(+x)	*w=1*	*w=1*	*w=1*	*w=1*	*w=1*	
☞MKM		-1		-1		-1	-3
MKK	-1		-1		-1		-3+*(-x)*

The evaluation table in (21) is also applicable to the SQ5 and SQ2. SQ5 in fact seems to distinguish Malayalam from Non-Malayalam, so the Non-Malayalam set includes both Kannada past marking without agreement as well as Tamil for the present-habitual. SQ2 differed from (21) in producing a second alternative output MKK for the MK_ context. By doing so, we see that the participant varys between weighing up Match FT and Match VP. The latter option is however superseded by the suppletive morphology of Malayalam in the Past context, just like SM4.

The fourth and final group (SQ3 and SM5), have chosen Kannada inflections in both KM_ and MK_ contexts. Once again, like the first group, the context has been ignored and the choice is either based on syntactic structure (Max F) or MT (MT anti-Faith). The evaluation table for this group is shown in (22).

(22) Evaluation Table for uniform K inflection

KM_	Max F	MT anti-Faith	Min F	MT Faith	Match FT	Match VP	HV
	w=1(+x)	*w=1(+x)*	*w=1*	*w=1*	*w=1*	*w=1*	

KMM	-1	-1			-1		-3+*(-x)*
☞KMK			-1	-1		-1	-3
MK_	Max F	MT anti-Faith	Min F	MT Faith	Match FT	Match VP	HV
	w=1(+x)	*w=1(+x)*	*w=1*	*w=1*	*w=1*	*w=1*	
MKM	-1	-1				-1	-3+*(-x)*
☞MKK			-1	-1	-1		-3

While consistently favoring MT anti-Faith, the sequential bilingual SQ3 went with the geminate-less Malayalam inflection in the habitual MK_ context instead of the agreement-less Kannada inflection that they used in the KM_ context. This small anomaly aside, our evaluation framework has successfully analyzed every piece of code-mixing data that was elicited in the task.

6 Discussion

In (23) we re-present the results table from (17) sorted according to SQ and SM.

(23) Results table sorted by SQ and SM

Set	A			B		
Language of the Paragraph	K			M		
Language of Verb prompt	M *paDi* ‘study’			K *kari* ‘call’		
TAM context for Verb Prompt	NF	Hab	Past	NF	Hab	Past
Suffix K	*-yakke*	*-yutta:re*	*-ide*	*-yakke*	*-yutta:re*	*-daru*
Suffix M	*-kkya:n*	*-kkyum*	*-ccu*	*-kkya:n*	*-kkyum*	*-ccu*
SM1	M	M	M	K	K	K
SM3	K	K	K	~M	~M	M
SM4	K	K	M	~M	~M	M
SM2	M	M	M	M	M	M
SM5	K	K	~K	K	K	K
SQ2	K	~K	K	M/K	M/K	M
SQ4	M	M+K	~K	K	K+M	K+M
SQ5	K	T	~K	#M	#M	#M
SQ1	M	M	M	M	M	M
SQ3	K	~K	~K	K	~M	K

The simultaneous bilinguals are expected to be more balanced in the relative proficiency between the two languages, than sequential bilinguals. The higher proficiency in Kannada, the L2, is clearly apparent in the fact that there is only one instance of (~K) among in the SM responses, whereas five out of seven cases of Kannada inflection usage in the SQ responses are instances of (~K). In contrast, the SM group has four instances of (~M) as opposed to a single instance in SQ. Considering that Malayalam is the L1 for all the

participants and nobody made any (~M) error in the control contexts with nonce, it is likely that all the (~M) errors are the effect of MT anti-Faith, where the participants produce an output that is morpho-syntactically M, but phonologically unlike M. If this is an innovation, then it is interesting to note that such innovation is more common among the SM group than the SQ.

The "do support" strategy is also conspicuous by its total absence from the SM group. Although the sample size of this pilot study is very small, it indicates that the SM group is comfortable to code-mix in the context of featural mismatch between the two languages. They have used three different strategies to code-mix. For the first and second strategy, they weighed up the Match VP and Match FT constraints respectively. In the third case, they went with Min F and Max F producing uniform M and K inflections. Thus, the MT-based constraints did not have any significant influence on the SM participants.

The SQ participants were a lot more varied in their outputs, and the cases of dual response, "do support", and complete lexical change in the verb (#M) responses indicate that this group is not so comfortable with code-mixing in the featural mismatch context. Their responses can be grouped into two sets based on whether they weigh up the
Match FT constraint, or one of the MT-based constraints. Evidence for the argument that uniform K inflection in SQ3 is a consequence of weighing up MT-anti faith rather than Max F, comes from the fact that the output lacks agreement marking.

References

Adger, D. (2003). Core syntax: A minimalist approach. Oxford, UK: Oxford University Press.

Alexiadou, A. & Anagnostopoulou, E. (1998). Parametrizing AGR: word order, V-movement and EPP-checking. Natural Language and Linguistic Theory, 16, 491-539.

Bel, A. (2003). The syntax of subjects in the acquisition of Spanish and Catalan.

Borer, H. (2003). Exo-skeletal vs. Endo-skeletal explanations: syntactic projections and the lexicon. In M. Polinsky and J. Moore (Eds.), The Nature of Explanation (pp.31-67). Chicago: Chicago University Press, (CSLI).

Borer, H. (2015). The category of roots. In A. Alexiadou, H. Borer and F. Schäfer (Eds.), The Roots of Syntax, the Syntax of Roots (pp.).Oxford: Oxford University Press.

Borer, H. (2015). Wherefore roots? Theoretical Linguistics 2014; 40(3/4), 343 – 359.

Chomsky, N. (1970). Remarks on nominalization. In R. Jacobs & P. Rosenbaum (Eds.), Readings in English transformational grammar (pp. 184–221). Waltham, MA: Ginn-Blaisdell.

Chomsky, N. (1981). Lectures on government and binding. Dordrecht, Netherlands: Foris.

Chomsky, N. (1995). The minimalist program. Cambridge, MA: MIT Press.

Chomsky, N., & Lasnik, H. (1993). The theory of principles and parameters. In A. Jacobs, W. von Stechow, & T. Vennemann (Eds.), Syntax. An international handbook of contemporary research (pp. 506–569) Berlin: de Gruyter.

Díaz, L., Bel, A., & Bekiou, K. (2017). Interpretable and uninterpretable features in the acquisition of Spanish past tenses. In *The role of formal features in second language acquisition* (pp. 484-512). Routledge.

Embick, D. (2015). The Morpheme: A Theoretical Introduction (Vol. 31). Walter de Gruyter GmbH & Co KG.

Embick, D. and Noyer, R. (2007). Distributed morphology and the syntax/morphology interface. In Oxford handbook of linguistic interfaces, chapter 9. Oxford : Oxford University Press

Grimstad, M. B., Riksem, B. R., Lohndal, T., & Åfarli, T. A. (2018). Lexicalist vs. exoskeletal approaches to language mixing. The Linguistic Review, 35(2), 187-218.

Halle, M. & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S.J. Keyser (eds.), The View from Building 20 (pp.111–176). Cambridge, Mass.: MIT Press.

Harley, H. (2014). On the identity of roots. Theoretical linguistics, 40(3-4), 225-276.

Hulk, A., & Müller, N. (2000). Cross-linguistic influence at the interface between syntax and pragmatics. Bilingualism: Language and Cognition, 3(III), 227–244

Legendre, G., Miyata, Y., & Smolensky, P. (2022). Harmonic Grammar–A formal multi-level connectionist theory of linguistic well-formedness: An application. In *12th Annual Conference. CSS Pod* (pp. 884-891). Psychology Press.

Liceras, J. M., Zobl, H. & Goodluck, H. (2008). Introduction. In J. M. Liceras, H. Zobl & H. Goodluck (Eds.), The role of formal features in second language acquisition (pp.1-19). New York: Lawrence Erlbaum Associates. Taylor & Francis Group.

MacSwan, J. (2014). A minimalist approach to intrasentential code switching. Routledge.

Mahootian, S. (1993). A null theory of codeswitching. Northwestern University.

Muysken, P. (2000). Bilingual speech: A typology of code-mixing. Cambridge University Press.

Myers-Scotton, C. M. (1993). Duelling languages: Grammatical structure in codeswitching. Oxford: Oxford University Press. Myers-Scotton, C. M. (2001). The matrix language frame model: Developments and responses. In R. Jacobson (ed.), Codeswitching worldwide II, pp. 23–58. Berlin: Mouton de Gruyte

Prince, A., & Smolensky, P. (2003). Optimality Theory in phonology. *International encyclopedia of linguistics*, *3*, 212-22.

Riksem, B. R. (2018). Language mixing in American Norwegian noun phrases. Journal of Language Contact, 11(3), 481-524.

Sorace, A. (2003). Near-nativeness. In M. Long & C. Doughty (Eds.), Handbook of second language acquisition (pp. 130–152). Oxford, UK: Blackwell.

Tsimpli, I.-M. (2001). LF-interpretability and language development: A study of verbal and nominal features in Greek normally developing and SLI children. Brain and Language, 77, 432–448.

Tsimpli, I.-M. & Dimitrakopoulou, M. (2007). The Interpretability Hypothesis: evidence from wh-interrogatives in second language acquisition. Second Language Research 23, 215-242.

Tsimpli, I-M. & Mastropavlou, M. (2008). Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. In J. M. Liceras, H. Zobl & H. Goodluck (Eds.), The role of formal features in second language acquisition (pp. 142-183). New York: Lawrence Erlbaum Associates. Taylor & Francis Group.

Familiar Definite Marking in Magahi[1]

AIDAN SHARMA, *Rutgers University*

ABSTRACT

This paper investigates the nominal suffix *-waa* in Magahi, an Eastern Indo-Aryan language. Existing accounts of *-waa* vary from analyzing it semantically in terms of familiarity and non-honorificity (Alok 2022), diminutivity (Atreya & Sinha 2020), or definiteness (Kumar 2020) and syntactically in terms of whether it projects a head in the nominal spine (Kumar 2020) or not (Alok 2012, 2022). I argue that *-waa* is a familiar definite marker, similar to the German strong article (Schwarz 2009) and Akan familiar article (Arkoh & Matthewson 2013), with additional presuppositions of non-uniqueness (Owusu 2022) and non-honorificity. Additionally, I argue that *-waa* can either be generated as the definite allomorph of the general classifier (Kumar 2020) and undergo CLF to D movement, or be be base generated in D (Simpson 2005).

1 Introduction

This paper proposes a semantic and syntactic analysis of the nominal suffix *-waa* in Magahi[2] (Eastern Indo-Aryan) as a familiar definite marker. Earlier studies on *-waa* differ in their approaches to both its semantics and syntax. Alok (2012, 2022) analyzes *-waa* as a nominal particle that encodes familiarity as a presupposition and non-honorificity as expressive content (Potts 2007). In particular, Alok maintains that the main contribution of *-waa* is not definiteness and *-waa* is not projected as a head. Atreya & Sinha (2020) treat *-waa* as a diminutive marker that can convey endearment, derogation, and specificity. However, Kumar (2020) focuses on Magahi's status as a numeral classifier language and argues that *-waa* is the definite allomorph of the classifier used in "bare classifier" phrases similar to the bare classifier constructions in Cantonese (Cheng & Sybesma 1999). Lahiri (2021) also glosses *-waa* as a classifier.

This paper proposes that *-waa* is a familiar definite marker similar to the German strong article (Schwarz 2009), but with the relevant notion of familiarity including both strong and weak familiarity as defined by Roberts (2003). Syntactically, my analysis is most similar to the one in Kumar (2020). However, I argue that definites with *-waa* are not bare classifier phrases, but full DPs. In particular, I argue that *-waa* can originate as the head of a classifier projection and undergo CLF to D head movement or that *-waa* can be base generated in D, presumably through reanalysis of frequent CLF to D movement (Simpson 2005).

[1]All uncited data in this paper comes from elicitations done by the author with five native speakers of Magahi, three from the Nawada district of Bihar and two from the Jehanabad district.

[2]Magahi is primarily spoken in Bihar and is considered one of the three main Bihari languages, along with Maithili and Bhojpuri. The number of speakers is estimated to be between 9 and 12 million (Verma 2003).

2 Familiarity and Uniqueness

Whether definite descriptions require uniqueness (Frege 1892; Russell 1905) or familiarity (Heim 1982) has been a long-standing debate in linguistics and philosophy. The following examples (Schwarz 2013: 535 based on Hawkins 1978) seem to vary in whether they require uniqueness or familiarity.

(1) a. **Anaphoric**
John bought *a book* and a magazine.
The book was expensive.

b. **Immediate Situation Uniqueness**
the desk (**Context:** uttered in a room with exactly one desk)

c. **Larger Situation Uniqueness**
the prime minister (**Context:** uttered in the UK)

d. **Bridging**[3]

i. **Producer-Product**
John bought *a book*.
The author is French.

ii. **Part-Whole**
John's hands were freezing as he was *driving* down the street.
The steering wheel was bitterly cold and he had forgotten his gloves.

English does not distinguish between these uses of definites and *the* is used for all of them. However, Schwarz (2009) showed that some languages do differentiate between these types of definites and have dedicated articles for the concepts of uniqueness and familiarity.

2.1 German Weak and Strong Articles

Schwarz (2009) shows that there are two types of definite articles in German, which are differentiated by their their phonological status following a preposition. The *weak article* in (2) contracts after a preposition, but the *strong article* in (3) does not.

(2) *Hans ging **zum** Haus*
Hans went **to_the**$_{weak}$ house
'Hans went to the house'.

(3) *Hans ging **zu dem** Haus*
Hans went **to the**$_{strong}$ house
'Hans went to the house'.

Schwarz shows that these two articles not only have different phonological behavior, but also different semantics. The weak article is used for unique definites, and the strong article is used for familiar definites. This can be seen with the different definites described by

[3]Hawkins (1978) uses the term associative anaphora. The term bridging is due to Clark (1975).

Hawkins (1978). While anaphoric definites require the strong article (4), both immediate (5) and larger situation (6) uniqueness definites require the weak article.

(4) **Anaphoric**

In der New York Bibliothek gibt es ***ein Buch*** *über Topinambur. Neulich*
in the New York library exists EXPL **a** **book** about topinambur recently
war ich dort und habe ***{#im / in dem}*** ***Buch*** *nach einer Antwort*
was I there and have **{#in_the$_{weak}$ / in the$_{strong}$}** **book** for an answer
auf die Frage gesucht, ob man Topinambur grillen kann.
to the question searched whether one topinambur grill can

'In the New York public library, there is **a book** about topinambur. Recently, I was there and searched **in the book** for an answer to the question of whether one can grill topinambur'. (Schwarz 2009: 30)

(5) **Immediate Situation Uniqueness**

Das Buch, das du suchst, steht ***{im / #in dem}*** *Glasschrank.*
the book that you look.for stands **{in_the$_{weak}$ / #in the$_{strong}$}** glass.cabinet

'The book that you are looking for is **in the glass cabinet**.' (Schwarz 2009: 39)

(6) **Larger Situation Uniqueness**

Armstrong flog als erster ***{zum / #zu dem}*** *Mond.*
Armstrong flew as first.one **{to_the$_{weak}$ / #to the$_{strong}$}** moon

'Armstrong was the first one to fly **to the moon**.' (Schwarz 2009: 40)

Schwarz claims the difference between the two is the presence of an anaphoric index, which blocks the contraction of the strong article. As for the semantics, the weak article takes a situation and a property as arguments. It presupposes there is a unique individual satisfying the property in the given situation and returns that individual. The strong article functions similarly to the weak article but also takes an index argument of type e, semantically equivalent to a pronoun. Again there is a uniqueness presupposition and the article returns a definite description but, for the strong article, the presupposition and definite description include a statement identifying the referent with the index argument. The structures and denotations for the two articles are given below (Schwarz 2019: 12).

(7) a. $[_{DP}[the_{weak}\ s]NP]$

b. $[\![the_{weak}]\!]^g = \lambda s_r \lambda P_{<e,st>} : \exists! x[P(x)(s_r)].\iota x[P(x)(s_r)]$

(8) a. $[_{DP}\ i[[the_{strong}\ s]NP]]$

b. $[\![the_{strong}]\!]^g = \lambda s_r \lambda P_{<e,st>} \lambda y : \exists!x[P(x)(s_r) \wedge x = y].\iota x[P(x)(s_r) \wedge x = y]$

Interestingly, the two types of bridging examined by Schwarz also make a distinction between the two articles.

(9) **Producer-Product Bridging**

Das	***Theaterstück***	*missfiel*	*dem*	*Kritiker*	*so*	*sehr,*	*dass*	*er*	*in*	*seiner*
the	**play**	displeased	the	critic	so	much	that	he	in	his
Besprechung	*kein*	*gutes*	*Haar*	***{#am / an dem}***				***Autor***	*ließ*	
review	no	good	hair	**{#on_the$_{weak}$ / on the$_{strong}$}**				**author**	left	

'**The play** displeased the critic so much that he tore **the author** to pieces in his review.' (Schwarz 2009: 53)

(10) **Part-Whole Bridging**

Der	***Kühlschrank***	*war*	*so*	*groß,*	*dass*	*der*	*Kürbis*	*problemlos*
the	**fridge**	was	so	big	that	the	pumpkin	without.a.problem
{im / #in dem}		***Gemüsefach***		*untergebracht*		*werden*	*konnte.*	
{in_the$_{weak}$ / #in the$_{strong}$}		**crisper**		stowed		be	could	

'**The fridge** was so big that the pumpkin could easily be stowed **in the crisper**.' (Schwarz 2009: 52)

In producer-product bridging, the product (the play) behaves as if it introduces an antecedent for the producer (the author), and thus the strong article is used with *Autor* 'author'. Meanwhile in part-whole bridging, the part (the crisper) behaves as if it is unique relative to the whole (the fridge) and the weak article is used with *Gemüsefach* 'crisper'.

2.2 Akan Familiar Article

Arkoh & Matthewson (2013) investigate the semantics of the definite article *nʊ́* in Akan (Kwa; Niger-Congo).[4] They argue it is a familiar definite article (glossed as FAM) and has the semantics of the German strong article.

(11) **Context**: beginning of conversation

mʊ̀-tɔ́-ɔ̀	*èkùtú*	*(*nʊ́).*	*èkùtú*	**(nʊ́)*	*yɛ̀*	*dɛ̀w*	*pápá.*
1SG.SUBJ-buy-PST	orange	FAM	orange	FAM	be	nice	good

'I bought an orange. The orange was really tasty.' (Arkoh & Matthewson 2013: 2)

[4] Arkoh & Matthewson also show that *nʊ́* can be used as a third person pronoun and a dependent clause marker, but I will only mention its use as a familiar article here.

Example (11) above shows that *nʊ* cannot be used with indefinites, even if specific, but is required for anaphoric definites. However, *nʊ* does not strictly require linguistic antecedence. It is compatible with nouns that are familiar to all discourse participants even if not introduced in the discourse.

(12) **Context:** a parent talking to his/her spouse about their children

mbofra nó wɔ́ dan nɔ́ mu
children FAM be room FAM in

'The children are in the room.' (Arkoh & Matthewson 2013: 7 from Saah 1994: 152)

While *nʊ* is compatible with these familiar definites, it is incompatible with larger situation uniqueness definites. These are expressed with bare nouns.

(13) *kwámì nyá-à kràtàá fí-ì èɡyá krɔ́nkrɔ́n póp hɔ́*
Kwame get-PST letter from-PST father holy pope there

'Kwame got a letter from the holy father Pope'. (Arkoh & Matthewson 2013: 11)

(14) *amstɔ́ŋ nyí nyímpá áà ɔ́-dzí-ì kán tú-ù kɔ́-ɔ̀ ɔ̀sìrán dʊ̀*
Armstrong is person REL 3SG.SUBJ-eat-PST first fly-PST go-PST moon top

'Armstrong was the first person to fly to the moon'. (Arkoh & Matthewson 2013: 11)

Because *nʊ* seems to require familiarity rather than uniqueness, Arkoh & Matthewson (2013) propose that it is the Akan equivalent of the German strong article. However, Owusu (2022) notes that adopting the semantics associated with the German strong article for *nʊ* does not rule out its use with larger situation uniqueness definites. While DPs such as *the Pope* or *the first man to fly to the moon* are necessarily unique, they can also be familiar. For example, in a conversation where the Pope has already been mentioned, *nʊ* would seem to be licensed with *èɡyá krɔ́nkrɔ́n póp* 'holy father Pope' based on the other Akan data. Additionally, Akan has no weak article that might rule out the strong article on the basis of competition. Thus, Owusu proposes that *nʊ* requires an additional presupposition of non-uniqueness, usually associated with demonstratives (Dayal & Jiang 2023).

(15) **Non-uniqueness**
$\exists s' s \leq s' \wedge |\{x \mid P(x)(s')\}| > 1$

Recall that the strong article takes a situation argument s_r and a property argument $P_{<e,st>}$. Then, (15) says that there must be a larger situation containing the situation argument of the determiner, in which there is more than one individual that satisfies P. In other words,

while the familiar determiner picks out an individual that is unique/maximal relative to *P* in the given situation, there do exist other individuals that satisfy *P*; they are just not in the given situation. Larger situation uniqueness definites cannot satisfy this presupposition and are therefore correctly predicted to be incompatible with *nʊ*.

3 Magahi *-waa*

Before discussing the semantics and syntax of *-waa*, a basic overview of its phonology is necessary. The suffix *-waa* attaches only to nouns, and the sequence forms a single prosodic unit. This can have effects on stress and vowel length, such as in the example below.

(16)

	book	book-*waa*
Romanization	*kitaab*	*kitabwaa*
IPA	[kɪ.t̪ɑːb]	[kɪ.t̪əb.wɑː]

Besides this interaction with stress and vowel length, *-waa* also has five allomorphs: *-waa, -aa, -(i)yaa, -(i)yãã, -maa* (Atreya & Sinha 2020). While there are general patterns that determine the use of each allomorph depending on the phonology of the root noun (e.g., *-yaa* occurs after nouns ending in *ii* and *-maa* occurs after nouns ending in a nasal consonant), several of them can be used in identical environments and there is a large amount of inter- and intra-speaker variation. For example, my consultants used three allomorphs for *-waa* with the word for 'book': *kitab-waa, kitab-aa, kitab-iyaa*. One speaker even used all three at different times.[5] Despite this slightly blurry set of facts, Atreya & Sinha and Alok agree that *-waa* is the general form, and I will gloss all these allomorphs as -WAA in the Magahi examples in this paper.

3.1 Semantics of *-waa*

I am taking definiteness to be the primary contribution of *-waa* since its use forces a definite interpretation. In particular, nouns suffixed with *-waa* cannot be interpreted as indefinites (17), generics (18), or kinds (19) (cf. Alok 2012: 46). Even if the indefinite in (17) is specific, *-waa* is disallowed.

(17) *(ek tho) bilai-(#yaa)*
one CLF cat-(#WAA)
'a/one cat'

(18) *chiṛai-(#waa) uṛa hai*
bird-(#WAA) fly AUX
'Birds fly.'

(19) *dainasor-(#waa) bilupt ho gelai*
dinosaur-(#WAA) extinct be went
'Dinosaurs are extinct.'

[5] See Atreya & Sinha (2020) for more information on the allomorphs of *-waa* and Alok (2022) for the possibilities of different allomorphs on the same noun.

Additionally, Löbner's (1985) diagnostic shows that *-waa* is not a demonstrative.

(20) #*laik-waa sutiit hai auu laik-waa na sutiit hai*
boy-WAA sleeping is and boy-WAA not sleeping is

'The boy is sleeping and the boy is not sleeping.' (Alok 2012: 25)

Nouns suffixed with *-waa* are number neutral, but a plural suffix *-(a)n* can be suffixed to nouns as well, in which case they must be interpreted as plural.

(21) *kutt-waa*
dog-WAA

'the dog(s)'

(22) *kutt-waa-n*
dog-WAA-PL

'the dogs'

However, I will not discuss number marking and will focus on singular definites.

3.1.1 Anaphoricity

Magahi *-waa*, like the Akan article and German strong article, is obligatory on anaphoric definites. Note that a demonstrative can optionally co-occur with *-waa*.

(23) *kal ham ek tho kutta dekhaliai. (uu) kutt-#(waa) bari sundar halai*
yesterday 1SG one CLF dog saw DEM dog-WAA very beautiful was

'Yesterday I saw a dog. The/that dog was very beautiful.'

The narrative sequence below, modeled after the example in Jenks (2018: 510), shows that *-waa* is obligatory for anaphoric definites regardless of syntactic position.[6]

(24) *class me e-go laṛaka auu laṛakii hai*
class in one-CLF boy and girl is

'There is a boy and a girl in class.'

a. *ham (uu) laṛak-#(waa) ke kal milaliai*
1SG DEM boy-WAA ACC yesterday met

'I met the/that boy yesterday.'

b. *ham (uu) laṛak-#(waa) ke khatir e-go upahaar le liye hai*
1SG DEM boy-WAA GEN for one-CLF gift take bring AUX

'I'm bringing a gift for the/that boy.'

[6]Additionally, see Alok (2012: 29-32) for arguments that *-waa* is not a topic marker.

c. *(uu) laṛak-#(waa) biis saal ke lago hai*
DEM boy-WAA twenty year GEN seem AUX

'The/that boy looks 20 years old.'

d. *hamra nãĩ lago hai ki (uu) laṛak-#(waa) bahoot interesting hai*
1SG.OBL NEG seem AUX COMP DEM boy-WAA very interesting is

'I don't think that the/that boy is very interesting.'

Finally, donkey definites also require *-waa* on the definite NP.

(25) a. *sabhe kisaan jekra paas gadha hai, uu gadha-#(waa) ke maaro hai*
every farmer REL.OBL near donkey is 3SG donkey-WAA ACC beat AUX

'Every farmer who has a donkey beats the donkey.'

b. *agar kisaan ke paas gadha hai, to uu gadha-#(waa) ke maaro hai*
if farmer GEN near donkey is then 3SG donkey-WAA ACC beat AUX

'If a farmer has a donkey, he beats the donkey.'

This further highlights the connection between *-waa* and familiarity (Jenks 2015b).

3.1.2 Non-uniqueness

In addition to familiarity, the incompatibility of *-waa* with larger situation uniqueness definites shows that it encodes non-uniqueness.[7]

(26) *chaand-(#waa) uuglai*
moon-(#WAA) rose

'The moon rose.'

(27) *suuraj-(#waa) puurab me uugo hai*
sun-(#WAA) east in rise AUX

'The sun rises in the east.'

[7]Simpson & Biswas (2016: 11) and Ushasi Banerjee (p.c.) report that definiteness marking with the classifier *Ta* in Bangla is possible for larger situation uniqueness definites such as 'moon' in episodic sentences but not generic ones. Similar facts are reported for Akan (Comfort Ahenkorah p.c.) and Cantonese (Ka-Fai Yip & Margaret Chui Yi Lee p.c.). I have not explored this fully in Magahi, but one of the three speakers I checked with, despite his initial judgement that it was infelicitious, said that *-waa* was acceptable with *chaand* 'moon' in (26). However, that speaker did not find *-waa* acceptable with *suuraj* 'sun' in (27) at all. More research is needed, but it is interesting that (26) has an episodic interpretation while (27) has a generic one.

(28) *amerika ke raaspati-(#yaa) paagal hai*
America GEN president-(#WAA) crazy is

'America's president is crazy.'

Superlatives, which are necessarily unique, are also incompatible with *-waa*.

(29) a. *pahala aadamii-(#yaa) chaand par*
first man-(#WAA) moon on

'the first man on the moon'

b. *duniya ke sab se tej aadamii-(#yaa) se milai ke mun hai*
world GEN all from smart man-(#WAA) from find PRT want AUX

'I want to meet the world's smartest man.'

Recently, however, it has been argued that certain classifier languages do not display a true unique/familiar dichotomy like the one in German. Instead, Yip et al. (2023) argue on the basis of Cantonese and Bangla that the lack of definite marking on larger situation uniqueness definites is not because the definites require non-uniqueness. Rather, Yip et al. argue that the bare nouns used in these instances behave like quasi-names, such as *Mom* in English. While Yip et al. present a convincing account of the Cantonese and Bangla data, I argue that this cannot extend to Magahi. In fact, Magahi *-waa* can also occur with proper and quasi-names, but only for people familiar to you of equal/lower social status.

(30) *ratan-maa*
Ratan-WAA

'Ratan' (referring to a friend or younger known person named Ratan)

(31) *bhai-waa*
brother-WAA

'brother' (referring to a younger brother) (Alok 2022: 1)

This behavior is not unusual. The use of definite elements with names is attested cross-linguistically e.g., in Greek, Maori (Anderson 2004). Additionally, Magahi has allocutive agreement that encodes honorificity in the clausal domain (Alok 2020, 2021; Alok & Baker 2022), so it is not entirely surprising that honorificity could be encoded in the nominal domain as well. However, while this shows that *-waa* is compatible with quasi-names, it could still be the case that larger situation uniqueness definites are treated as honorific/quasi-names of high status. To see that this is not the case, it is useful to note that *-waa* can be added to (quasi-)names of social superiors to show disrespect/contempt (Alok 2022). Nevertheless, in such cases, my consultants still consider *-waa* to be infelicitous on larger situation uniqueness definites.[8]

[8]I would like to thank an anonymous reviewer for suggesting to test this.

(32) *obaama ke baad ke halai raaspati? uu raaspati-(#yaa) buṛabak hai*
Obama GEN after who was president DEM president-(#WAA) idiot is

'Who was the president after Obama? That president is an idiot."

Yet, in a hypothetical scenario where our solar system has multiple suns or the Earth has multiple moons, my consultants say it would be felicitous to use *-waa* with *suuraj* 'sun' and *chaand* 'moon'. Thus, I maintain that non-uniqueness is a presupposition of *-waa*. I take the restrictions on the use of *-waa* with (quasi-)names to indicate that it also has a presupposition of non-honorificity (but see Alok 2022 who treats the non-honorificity of *-waa* as expressive content). It is this presupposition of non-honorificity that gives rise to the semantic effects described by Atreya & Sinha (2020); Alok (2022).

I will not attempt to give an account of the relationship here, but definiteness marking and non-honorificity appear to be connected in many other classifier languages as well. Jenks (2015a: 5) reports that in Thai high animate referents are expressed as bare nouns rather than bare classifier constructions, even when anaphoric. Simpson & Biswas (2016: 6) report similar data in Bangla and note that including the classifier indicates disrespect/contempt toward the referent. Finally, regarding names, Saul & Wilson (1980: 26) report that in Nùng classifiers may be used with proper names, but only for children. A possible explanation for these facts might come from the inverse relationship between marked semantic features and honorificity, explored by Wang (2023) for pronouns.

3.1.3 Weak Familiarity

So far, it appears that whenever German uses the strong article, Magahi uses *-waa* and whenever German uses the weak article, Magahi uses a bare noun. But, like the Akan article, *-waa* does not strictly require linguistic antecedence.

(33) **Context:** Ram and Rakesh are standing on the side of the road and see a dog on the other side. Ram says to Rakesh:

(uu) kutt-#(waa) saṛak paar karailai chaho hai
DEM dog-waa road across cross want AUX

'The/that dog wants to cross the road.'

Example (33) shows there is a contrast between Magahi *-waa* and the German strong article. Rather than separating the semantics of *-waa* from familiarity, I propose that the relevant notion of familiarity is slightly broader than linguistic antecedence. In particular, I adopt the definitions of *strong familiarity* and *weak familiarity* from Roberts (2003).

(34) **Strong Familiarity** (Roberts 2003: 304)
The NP has as antecedent a discourse referent introduced via the utterance of a (usually) preceding NP.

(35) **Weak Familiarity** (Roberts 2003: 304)

i. The entity referred to is perceptually accessible to the interlocutors.

ii. The entity referred to is globally familiar in the general culture or at least among the participants in the discourse, although not mentioned in the immediate discourse.

iii. Introduction of the NP's discourse referent is licensed solely by contextual existence entailments.

iv. Weak familiarity is guaranteed by giving a functional interpretation to the definite description (which function may have to be accommodated, with the intended argument(s) both familiar and highly salient (Bridging)).

The example in (33) is a case described by (35i.). Except for bridging, which will be mentioned in Section 3.1.4, the other examples of weak familiarity also require *-waa* in Magahi.

(36) **Context:** Ram and John are from the same town which has a single hospital that everyone knows about. Ram is not feeling well, so John tells him:

jaa aspatal-iyaa me dekhwaala
go hospital-WAA in examine

'Go and get check up in the hospital.' (Alok 2022: 5)

(37) a. *sabhe hotel ke kamar me e-go kitaab rakhal raho hai*
every hotel GEN room in one-CLF book kept PROG AUX

'A book is kept in every hotel room.'

b. *ii kamar-waa me kitab-aa tiivi-yaa ke niiche rakhal hai*
DEM room-WAA in book-WAA TV-WAA GEN below kept is

'In this room, the book is kept below the TV.'

Thus, *-waa* is a familiar definite marker that encompasses both strong and weak familiarity. The semantics I propose for *-waa* is largely the same as that of the German strong article.

(38) $[\![\text{-}waa]\!] = \lambda s_r \lambda P_{<e,st>} \lambda y : \exists! x\,[P(x)(s) \wedge x = y] \wedge \underline{\exists s'[s \leq s' \wedge |\{x \mid P(x)(s')\}| > 1]} \wedge \underline{\underline{\text{NHON}(x)}}.\iota x[P(x)(s) \wedge x = y]$

It has the additional presuppositions of non-uniqueness (single-underlined), as argued for in Akan by Owusu (2022), and non-honorificity (double-underlined) to capture the incompatibility of *-waa* with both larger situation uniqueness definites and (quasi-)names of people of higher social status. I assume that any referent that satisfies strong or weak familiarity (Roberts 2003) can introduce an anaphoric index as an argument for *-waa*.

3.1.4 Bridging

Magahi, unlike German, does not appear to distinguish between producer-product and part-whole bridging. I have not performed a full exploration of bridging in Magahi but, for the following two examples, speakers report *-waa* to be optional.

(39) *kamar-waa me dekhaliai. chat(-waa) bari ũũcha hai*
room-WAA in looked roof-WAA very high is

'I looked into the room. The roof is very high.'

(40) *kal e-go kitaab paṛhaliai. okar lekhak(-waa) samajhdaar hai*
yesterday one-CLF book read 3SG.GEN author-WAA smart is

'Yesterday I read a book. The author is very smart.'

The optionality of *-waa* here is a puzzle that I leave open for future research.

3.2 Syntax of *-waa*

This section argues that *-waa* can be a classifier that undergoes CLF to D movement or be base generated in D (Simpson 2005). In order to see how this analysis works, first we need to consider the usual use of classifiers in Magahi.

3.2.1 Magahi Classifiers

The majority of nouns in Magahi require a classifier to combine with numerals, though the classifier system is not very rich. There are two general classifiers *-go* and *tho* in free variation.

(41) {*e-go / ek tho*} *aam*
{one-CLF / one CLF} mango

'one dog'

(42) {*chaar-go / chaar tho*} *aadamii*
{four-CLF / four CLF} man

'four men'

The difference between them is that *-go* forms a closer prosodic unit with the numeral, as seen in the form of the numeral *ek* 'one' in (41). While *-go* and *tho* occur with the vast majority of nouns, some nouns require more contentful measure phrases (*massifiers* in the terminology of Cheng & Sybesma 1998).

(43) *ek muṭṭhi bhaat*
one handful rice

'a/one handful of rice'

(44) *du gilas paani*
two glass water

'two glasses of water'

Finally, there are two other classifiers which are only used with certain quantifiers.

(45) *baṛi menii aadamii*
lot CLF man
'lots of men' (Alok 2012: 47)

(46) *tanii sun aadamii*
little CLF man
'a few men' (Alok 2012: 47)

The few exceptions to the Magahi classifier system are related to time, such as *din* 'day', which do not combine with classifiers/measure phrases at all.

3.2.2 Classifiers in Definites

The question to be answered is how we know that *-waa* is related to the classifier position. For Kumar (2020), the evidence is that Magahi definites with *-waa* do not have the classifiers *-go* or *tho* and are very similar to bare classifier definites in other languages, especially closely related Bangla. Consider the following indefinite/definite pair in Bangla.[9]

(47) *ɛk ʈa boi*
one CLF book
'a/one book' (Dayal 2012: 204)

(48) *boi ʈa*
book CLF
'the book' (Dayal 2012: 204)

The Bangla definite in (48) looks very similar to a Magahi *-waa* definite, with the classifier seemingly occupying the same position as *-waa*. Note also that the classifier occurs pre-nominally in the indefinite but post-nominally, like *-waa*, in the definite.

Despite the similarities in a closely related language, so far there has been no evidence internal to Magahi that *-waa* behaves like the definite allomorph of the classifier. I argue that the best evidence comes from adjectives. Magahi adjectives in noun phrases with *-waa* must be suffixed with *-kaa* (masc.) or *-kii* (fem.), both of which I gloss as -KAA. This suffix is generally thought of as an allomorph of *-waa* that displays definiteness agreement on adjectives (Sinha 1966: 114). However, Kumar (2020, 2022) shows that this is actually a case of determiner spreading involving multiple DPs (Alexiadou 2014).[10] Additionally, despite its usual description as being definite, *-kaa* can occur in indefinite noun phrases.

(49) *paapaa, e-go baṛa-kaa baet lete aiba kaa*
papa one-CLF big-KAA bat bring come Q
'Father, bring me a big bat, won't you?' (Kumar 2022: 39)

Interestingly, after *-kaa* in indefinites, you can also get the classifier *-go*.

[9]The Bangla facts have been discussed much more, and the empirical picture is a lot more complicated than presented here. For accounts of classifiers, definites, and the Bangla noun phrase more generally, please see Bhattacharya (1999a,b); Dayal (2012, 2014); Chacón (2012); Biswas (2012, 2016); Syed (2015, 2016, 2017); Simpson & Biswas (2016); Syed & Simpson (2017).

[10]A determiner spreading analysis for these phrases also gives credence to the analysis of *-waa* as a determiner.

(50) *hamra e-go baṛa-kaa-{go / #waa} kutta chahi*
1SG one-CLF big-KAA-{CLF / #WAA} dog want

'I want a big dog.'

But in definite DPs only *-waa*, not *-go*, can appear after *-kaa*.

(51) a. *baṛa-ka-{waa / #go} kutt-waa hamra par bhãũk gelai*
big-KAA-{WAA / #CLF} dog-WAA 1SG.GEN on bark went

'The big dog barked at me.'

I take this as evidence that *-waa* is indeed a definite allomorph of the classifier. However, *-waa* can also combine with nouns that are incompatible with *-go* and *tho*.

(52) *baabujii thoṛe sun bhaat banalkai. bhat-waa ṭebal par hai*
dad little CLF made rice-WAA table on is

'Dad made a little rice. The rice is on the table.'

Din 'day', which does not occur with classifiers, can also occur with *-waa* (Prasad 2008: 189). The compatibility of *-waa* with such nouns means it serves as a general type of definiteness marker, which I take to be associated with the higher functional head D. Earlier, it was shown that *-waa* can also occur in CLF. To account for how it can be associated with both positions, I propose that for most definite nouns, *-waa* begins in CLF and moves to D. This eventually allowed for its reanalysis as a definite marker of category D leading to its compatibility with nouns such as *bhaat* 'rice' and *din* 'day' (cf. the analysis of Vietnamese definites in Simpson 2005). As evidence for CLF to D movement, consider the three forms of Magahi nouns reported by Grierson & Hoernle (1885): bare nouns or "the short form", nouns with *-waa* or "the long form", and nouns where *-waa* occurs twice or "the redundant form". The redundant forms of nouns provide support for CLF to D movement as they arguably involve pronunciation of both copies of *-waa*.[11] The forms for *baat* 'word' are given below.

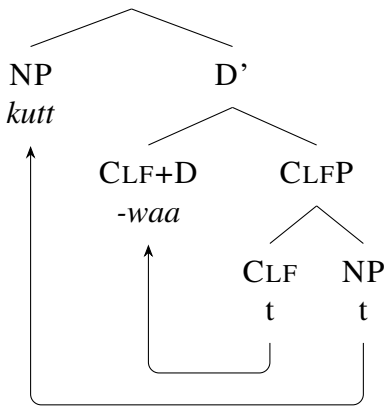
(53)

short form	**long form**	**redundant form**	
baat	*bat-iyaa*	*bat-iya-waa*	
word	word-WAA	word-WAA-WAA	(Grierson & Hoernle 1885: 16)

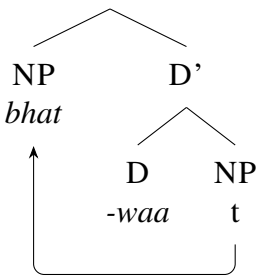
Thus, we have the following forms for *-waa* definites.

[11] I only consulted one Magahi speaker regarding redundant forms, but he did not accept them for any of the nouns I checked. Given this, it is possible that CLF to D movement of *-waa* is not synchronically active and *-waa* has been totally reanalyzed as category D. However, then the alternation between *-waa* and *-go* on adjectives would need further explanation. For this reason, I will keep the CLF to D movement analysis and assume that pronunciation of both copies of *-waa* is no longer allowed. However, I believe it is possible that the classifier use of *-waa* is restricted to just the adjectival examples, and an analysis where *-waa* is otherwise base-generated in D is a feasible alternative that would differ minimally from the present account.

(54)



(55)



A few features of the analysis in Schwarz (2009) are omitted in these structures for simplicity. For the structures in (54) and (55) there must also be a situation argument introduced with D and an index argument in a higher specifier of DP, above the SpecDP position that the NP occupies. In both DPs, the NP moves to SpecDP to check a [DEF] feature (cf. the NP-raising analysis of Bangla in Bhattacharya 1999a,b and its reformulation in terms of definiteness rather than specificity in Dayal 2012). The difference between the two is that for *kutt-waa* 'the dog', *-waa* is generated in CLF and moves to D, but for *bhat-waa* 'the rice', it is generated in D.

Interestingly, this process of a classifier becoming more determiner-like might also be happening for singular definites in Bangla. Consider the following example, where the indefinite expressions use the human classifier *jon*, but the definite expressions use the general classifier *Ta*.

(56) *laibreri-te Ek jon notun mEthor ar Ek jon gard rakha*
library-LOC one CLF$_{human}$ new janitor and one CLF$_{human}$ guard keep
*hoechhe. mEthor *(Ta) porisromi, kintu gard *(Ta) besh kuMRe.*
was janitor CLF hard-working but guard CLF quite lazy

'The library hired a new janitor and a new guard. The janitor is hard-working, but the guard is quite lazy.' (Simpson & Biswas 2016: 5)

Furthermore, use of the human classifier *jon* in the definite expressions is ungrammatical (Saurov Syed, p.c.), and *Ta* can also be used as a definite marker for nouns like 'rice', which it cannot combine with in indefinite constructions.

(57) *ma olpo bhaat ranna koreche. bhaat-Ta Tebil-er opor rakha ache.*
Mother some rice cook did Rice-CLF table-LOC on kept is

'Mom cooked some rice. The rice is on the table.' (Indira Das, p.c.)

This indicates that *Ta* can be used as a general marker of definiteness, like *-waa* in Magahi, and might not always be (exclusively) associated with the CLF position. However, there are other classifiers besides *jon* which can be used in both indefinites and bare classifier definites,[12] so the situation is not exactly the same as in Magahi and needs further research.

3.2.3 Arguments Against *-waa* as a Determiner

Alok (2012, 2022) presents three main arguments against the analysis of *-waa* as a determiner, based on the following: blocking, linear order, and numerals. Starting with blocking, Alok (2022) states if *-waa* is a determiner, the existence of definite bare nouns in Magahi is a Blocking Principle violation.

(58) **Blocking Principle** (Chierchia 1998: 360)
For any type shifting operation τ and any X:
$*\tau(X)$ if there is a determiner D such that for any X in its domain $D(X) = \tau(X)$.

However, even if bare noun definites require type-shifting via *iota*, this does not necessarily indicate that the blocking principle is being violated. As argued in Section 3.1, *-waa* has presuppositions of familiarity, non-uniqueness, and non-honorificity. Thus, the Blocking Principle will not be violated if type-shifting is used for the reference of an individual that fails to meet any of the presuppositional requirements of *-waa*, which is indeed the case for definite bare nouns in Magahi.

Turning to linear order, Alok (2012, 2022) notes that *-waa* occurs after the noun and adjectives are pre-nominal. Alok argues this is an issue if *-waa* heads a DP because because the way to resolve the linear order of *-waa* and the noun would be N to D movement. This would then mean that adjectives in noun phrases with *-waa* would have to occur after the N-*waa* sequence. However, the analysis in this paper derives the position of *-waa* relative to the noun by NP-raising. Thus, adjectives would still be expected to occur to the left of the noun, since they would be inside the NP during NP-raising.[13]

Finally, Alok argues that *-waa* is not a definite determiner because it occurs in indefinite numeral/quantifier expressions.

(59) *chaar go kitab-waa*
four CLF book-WAA
'four of the books' (Alok 2022: 16)

(60) *kuch kitab-waa*
some book-WAA
'some of the books' (Alok 2022: 16)

However the translations, taken directly from Alok (2022), show these are partitives and likely involve a more complex structure with two DPs (cf. the indefinites in 17).

To conclude, I will give one more piece of evidence for analyzing *-waa* as a determiner. In vocatives, *-waa* is disallowed.

[12] I would like to thank Andrew Simpson for pointing this out.

[13] This is consistent with Alok's (2012) analysis of adjectives as phrasal adjuncts inside NP rather than heads.

(61) **Context:** You see a boy across the street and want to call him.

laṛaka-(#waa)!
boy-#WAA

'Boy!'

This is consistent with observed differences between DPs and NPs (Longobardi 1994).

4 Conclusion

This paper has provided an overview of the Magahi nominal suffix *-waa*, with emphasis on the familiar/unique distinction proposed by Schwarz (2009). I have proposed, contrary to Alok (2012, 2022) and Atreya & Sinha (2020) but in line with Kumar (2020), that the primary contribution of *-waa* is definiteness.

Semantically, *-waa* is similar to the German strong article, but with additional presuppositions of non-uniqueness (Owusu 2022) and non-honorificity and a weaker requirement for familiarity. It is used for both strong and weak familiarity, as defined by Roberts (2003).

Syntactically, I have argued that *-waa* is the definite allomorph of the general classifier in Magahi. The primary evidence for this came from the alternation between *-waa* and *-go* on definite/indefinite adjectives. I have also proposed that besides undergoing CLF to D movement, *-waa* can be base generated in D (Simpson 2005) based on the redundant forms in Grierson & Hoernle (1885) and the compatibility of *-waa* with nouns that do not take the general classifiers. Crucially, I showed that *-waa* occurring in D does not violate the Blocking Principle (Chierchia 1998) or the order of elements in the Magahi noun phrase.

Acknowledgements

I would like to thank Ram Charitra Sharma, Ram Khelawan Sharma, Rakesh Sharma, Ram Kumar, and Kumar Brajesh for taking the time to share their Magahi judgements with me. I am also grateful to the audience at FASAL 14 at Stony Brook, as well as Maria Kouneli, Deepak Alok, Sreekar Raghotham, Shannon Bryant, Indira Das, Dorothy Ahn, Chandan Kumar, Mark Baker, and Ariela Ye for feedback and advice on various previous versions of this work.

References

Alexiadou, Artemis. 2014. *Multiple determiners and the structure of DPs*. John Benjamins Publishing Company.

Alok, Deepak. 2012. *A language without articles: the case of Magahi*: Jawaharlal Nehru University, New Delhi MA thesis.

Alok, Deepak. 2020. *Speaker and addressee in natural language: Honorificity, indexicality and their interaction in Magahi*: Rutgers, The State University of New Jersey dissertation.
Alok, Deepak. 2021. The morphosyntax of Magahi addressee agreement. *Syntax* 24(3). 263–296.
Alok, Deepak. 2022. The discourse particle -waa in Magahi nominal domain. *Discourse* .
Alok, Deepak & Mark Baker. 2022. Person and honorification: Features and interactions in Magahi. *Glossa: a journal of general linguistics* 7(1).
Anderson, John M. 2004. On the grammatical status of names. *Language* 80(3). 435–474.
Arkoh, Ruby & Lisa Matthewson. 2013. A familiar definite article in Akan. *Lingua* 123. 1–30.
Atreya, Lata & Sweta Sinha. 2020. Phonological and functional analysis of diminutive marker -waa in Magahi. *Dialectologia: revista electrònica* 43–57.
Bhattacharya, Tanmoy. 1999a. Specificity in the Bangla DP. *Yearbook of South Asian Languages and Linguistics* 2. 71–99.
Bhattacharya, Tanmoy. 1999b. *The structure of the Bangla DP*: University College London dissertation.
Biswas, Priyanka. 2012. Reanalyzing definiteness in Bangla. In *Annual Meeting of the Berkeley Linguistics Society*, 19–29.
Biswas, Priyanka. 2016. *Number marking and definiteness in Bangla*: University of Southern California dissertation.
Chacón, Dustin. 2012. Head movement in the Bangla DP. *Journal of South Asian Linguistics* 4(1). 3–24.
Cheng, Lisa Lai-Shen & Rint Sybesma. 1998. Yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua journal of Chinese studies* 28(3). 385–412.
Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30(4). 509–542.
Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural Language Semantics* 6(4). 339–405.
Clark, Herbert H. 1975. Bridging. In R.C. Schank & B.L. Nash-Webber (eds.), *Theoretical issues in natural language processing*, New York: Association for Computing Machinery.
Coppock, Elizabeth & David Beaver. 2015. Definiteness and determinacy. *Linguistics and Philosophy* 38. 377–435.
Dayal, Veneeta. 2012. Bangla classifiers: Mediating between kinds and objects. *Rivista di Linguistica* 24(2). 195–226.
Dayal, Veneeta. 2014. Bangla plural classifiers. *Language and Linguistics* 15(1). 47–87.
Dayal, Veneeta & Li Julie Jiang. 2023. The puzzle of anaphoric bare nouns in Mandarin: A counterpoint to Index! *Linguistic Inquiry* 54(1). 147–167.
Etxeberria, Urtzi & Anastasia Giannakidou. 2010. Contextual domain restriction and the definite determiner. In Isidora Stojanovic Francois Recanati & Neftali Villanueva (eds.), *Context-dependence, perspective and relativity*, 93–126. De Gruyter Mouton Berlin.

Etxeberria, Urtzi & Anastasia Giannakidou. 2019. Definiteness, partitivity, and domain restriction: A fresh look at definite reduplication. *Studies in Diversity Linguistics* (25).
Frege, Gottlob. 1892. On sense and reference. In P. Geach & M. Black (eds.), *Translations from the philosophical writings of Gottlob Frege*, 56–78.
Grierson, George & A.F. Hoernle. 1885. *A comparative dictionary of the Bihari language*. Calcutta, Bengal Secretariat Press.
Hawkins, John. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croon Helm.
Heim, Irene Roswitha. 1982. *The semantics of definite and indefinite noun phrases*: University of Massachusetts Amherst dissertation.
Jenks, Peter. 2015a. Determiners, bare nouns, and donkey sentences in numeral classifier languages. Presentation at the 89th Annual Meeting of the LSA.
Jenks, Peter. 2015b. Two kinds of definites in numeral classifier languages. In *Semantics and Linguistic Theory*, 103–124.
Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49(3). 501–536.
Kumar, Chandan. 2020. *Typology of nominal modifiers in major languages of Bihar: Magahi, Bhojpuri, Maithili, Angika & Bajjika*: Jawaharlal Nehru University dissertation.
Kumar, Chandan. 2022. Multiple determiners in Magahi: A case beyond agreement. *Indian Journal of Language and Linguistics* 3(1). 32–45.
Lahiri, Bornini. 2021. *The case system of Eastern Indo-Aryan languages: A typological overview*. Routledge India.
Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4(4). 279–326.
Longobardi, Giuseppe. 1994. Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry* 609–665.
Owusu, Augustina P. 2022. *Cross-categorial definiteness/familiarity*: Rutgers The State University of New Jersey dissertation.
Potts, Christopher. 2007. The expressive dimension. *Theoretical linguistics* 33(2). 165–198.
Prasad, Saryoo. 2008. *Magahī phonology: a descriptive study*. Concept Publishing Company.
Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26. 287–350.
Russell, Bertrand. 1905. On denoting. *Mind* 14(56). 479–493.
Saah, Kofi Korankye. 1994. *Studies in Akan syntax, acquisition, and sentence processing*: University of Ottawa dissertation.
Saul, Janice E. & Nancy F. Wilson. 1980. *Nung grammar*. Summer Institute of Linguistics, University of Texas at Austin.
Schwarz, Florian. 2009. *Two types of definites in natural language*: University of Massachusetts Amherst dissertation.
Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and Linguistics Compass* 7(10). 534–559.

Schwarz, Florian. 2019. Weak vs. strong definite articles: Meaning and form across languages. *Studies in Diversity Linguistics* (25).

Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asia. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford Handbook of Comparative Syntax*, 806–838. New York: Oxford University Press.

Simpson, Andrew & Priyanka Biswas. 2016. Bare nominals, classifiers, and the representation of definiteness in Bangla. *Linguistic Analysis* 40(3-4). 167–198.

Sinha, A.C. 1966. *Phonology and morphology of a Magahi dialect: A thesis submitted to the University of Poona*. Poona (Deccan College Postgradute and Research Inst.).

Syed, Saurov. 2015. Focus-movement within the Bangla DP. In *Proceedings of WCCFL 32*, 332–341.

Syed, Saurov. 2016. Decomposing definiteness: Arguments for a split D-domain in Bangla. In *Proceedings of WCCFL 33*, 390–397.

Syed, Saurov. 2017. *The structure of noun phrases in Bengali: What it tells us about phases and the universality of DPs*: University of Southern California dissertation.

Syed, Saurov & Andrew Simpson. 2017. On the DP/NP status of nominal projections in Bangla: Consequences for the theory of phases. *Glossa: a journal of general linguistics* 2(1).

Verma, Sheela. 2003. Magahi. In Danesh Jain & George Cardona (eds.), *The Indo-Aryan Languages*, chap. 13. Routledge.

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

Yip, Ka-Fai, Ushasi Banerjee & Margaret Chui Yi Lee. 2023. Are there "weak" definites in bare classifier languages? In *Semantics and Linguistic Theory*, 253–275.

Processing of Relative Clauses in Malayalam

JAYAKRISHNAN SUBRAMONY, *Independent scholar (jsubramo@usc.edu)*

ABSTRACT

A study of the processing of relative clauses may offer insight into how a range of processes involving structural manipulation by way of movement might be realized in the mind. In this paper, I elucidate a self-paced reading experiment that investigates the processing of relative clauses in Malayalam. I use singly embedded relative clauses and counterbalance each item by varying the position of the RC in the sentence, and the gapping (subject or object) from the clause. I observe a slight preference for object relative clauses, and explain my results using an expectation based model.

1 Introduction

Models of processing account differently the manner of processing of relative clauses across languages. However, the general trend is that cross-linguistically subject extracted relative clauses are processed faster – due to the distance of extraction being lesser. Memory-based models and expectation-based models have different predictions for languages depending on how the head noun is placed in RCs. The two types of organization are noun-first and noun-last, depending on the position of the clausal modifier. English is a noun-first type of language:

(1) The man [whom I saw — in London] ...

On the other hand, languages such as Japanese and Chinese are noun-last:

(2) Japanese

[watashi-wa rondon-ni — mita] otoko ...
[I-TOP London-LOC — see.PST] man

'The man whom I saw in London...'

Memory-based models account for the ease of processing of subject extracted relative clauses in languages such as English, whereas their predictions with regards to noun-final languages such as Japanese go against the observations. Expectation-based models do account for a weak advantage for subject relative clauses in noun-final languages (O'Grady 2011, Levy 2008).

Malayalam is a Dravidian language that is spoken in the state of Kerala in India. With over 35 million native speakers, it is one among the 22 scheduled languages in the Indian constitution. It is a default head-final, agglutinative SOV language. It also allows for scrambling. It has a vocabulary heavily dependent on Sanskrit. The language lacks overt subject-verb agreement, a feature unique among the major Dravidian languages.

Malayalam relative clauses are of the noun-final type. It uses a modifier-like construction, known as a participial construction, in their formation, where a canonical relativizer is absent. Instead, they are formed by the suffixation of the participial morpheme *-a* to the verb which is inflected for time reference. This same participial morpheme is found in modifier structures such as adjectives in the language.

(3) Malayalam

a. e_i raavaNan-e konna raaman$_i$
EC raavaNan.ACC kill.PFV raaman.NOM
'Raaman who killed Raavanan'

b. raaman e_i konna raavaNan$_i$
raaman EC kill.PFV raavaNan.NOM
'Raavanan who Raman killed'

In the above examples, 3a is called the *subject (extracted)* RC, as the head noun is coindexed with the subject position empty category of the relative clause, and 3b is the *object (extracted)* RC, as the head noun coindexes with the empty category in the object position of the RC.

Mathew (2005) shows that participial morphemes in Malayalam precede the head noun, and leave a gap in the relative clause participial construction. Hima (2017) treats the participial as a determiner-like entity (cf. the Semitic relative clause analysis in Ouhalla 2004). Ishizuka (2005) proposes that in Japanese – where there is no relative pronoun, the RC structure does not involve movement but rather has a null pronoun in the position of where the gap should be, coindexed with the head noun. This follows the analysis in Pesetsky (2004), where the valuation of the interpretable unvalued feature of the participial is provided by the head noun. Ishizuka's account also provides a way to account for the processing of relative clauses in Japanese.

2 Accounts of Processing

I look at the processing of relative clauses using two processing models—one, a memory-based model, and the other, an expectation- or frequency-based model of processing.

2.1 Dependency Locality Theory

The Dependency Locality Theory (Gibson & Hickok 1993, Gibson 2000, and Gibson & Warren 2004), or DLT henceforth, which successfully explains the processing of pronominals in SVO languages, is a memory- or integration-based account of sentence processing. Comprehension is facilitated by the construction of a distance-based metric. The informational sources are integrated at each moment, and the interpretation is constrained by the available computational resources. DLT is explained using human computational resources that depend on the distance between the two elements that need to be integrated.

The computational resources required include the integration cost and the memory cost. The head and the constituents are connected, and the measure of the new discourse elements that intervene between them, is given by the energy units (EU) as the integration cost. The memory cost to keep the incomplete segments in memory, enabling the parser to keep track of the incomplete dependencies, is known as the memory units (MU). The sum of EU and MU is the total processing cost. Gibson's assumption (Gibson & Hickok 1993) is that the parser adopts an active filler strategy where the filler is assigned as soon as possible – at any plausible gap position.

In DLT, the number of discourse elements that need to be parsed is proportional to the integration cost. It also obeys the minimal attachment hypothesis – the incoming material is attached using the fewest nodes possible, as long as the parse is consistent with the wellformedness rules of the language.

According to DLT, the more the distance between the gap (marked as e, for empty category) and the filler, the integration costs are higher, and hence the processing time would be higher. The active filler strategy and minimal attachment hypothesis together imply that in a head-final language such as Malayalam, a filler-gap structure such as a relative clause would have an object advantage. That is, an object extracted RC would be processed faster compared to a subject extracted RC.

2.2 Surprisal Theory

Surprisal (Hale 2001, Levy 2008) quantifies the amount of new information conveyed by a word in context. The surprisal at word w_i is formalized as the negative log probability of observing word w_i given that words $w_1 ... w_{i-1}$ have already been processed. Processing difficulty is proportional to the amount of new information that needs to be processed, and it is not locality based, but rather is a parallel processing theory. Surprisal does not ascribe to any one grammar formalism that is used as a processing mechanism and instead uses frequency and expectation—it is a psycholinguistic analogue to mathematics and statistics.

It predicts a reversal of the locality-based difficulty patterns in syntactically constrained contexts. The integration of knowledge is incremental in resolving syntactic ambiguity. Expectation-based models account for a weak advantage in processing of subject relative clauses in noun-final languages.

3 Literature Survey

There is a plethora of existing work done on the processing of relative clauses in Germanic languages including English, and Romance languages. More recent work in processing has been done in Chinese, Japanese and Korean as well.

3.1 Noun-initial languages

A corpora-based experiment (Reali & Christiansen 2007) showed that *pronominal* object relative clauses took less time to be processed than *pronominal* subject relative clauses.

This was due to a difference in their frequency in speech. They also conducted a self-paced reading experiment to disconfirm a structure-based account that predicts a universal preference for subject relative clauses. Integration accounts cannot explain this preference since pronouns (overt or null) are not treated as new discourse elements – this is also the case with filler-gap pairs. Filler-gap pairs do not contribute to a processing difficulty as they are not new discourse elements. The experiment showed that pronominal object relative clauses are processed faster. This can be explained with theories such as similarity-based interference (Bever 1975, Gordon et al. 2001). Due to the dissimilarity of the head noun phrase with the personal pronouns, the interference in gap-filling is lesser. This accounts for a decreased interference during processing, as opposed to non-pronominal DPs. In addition, the preference of object position pronominal relative clauses can satisfactorily be explained with a frequency-based account.

In an ERP study in German by Mecklinger et al. (1995), the relative clauses were varied on syntactic and semantic dimensions, to examine the asymmetry in response to object and subject extracted relative clauses. There was a bias against the object, which was explained with the active-filler strategy—when gaps were encountered in the sentences, the main clause NP was assigned to the gap. Rohde & Horton (2014) observed that in English, implicit-causal constructions had a higher preference for relative clauses that attached to the lower argument. In Russian, Levy et al. (2013) proposed to integrate both the memory-based and expectation-based accounts to explain the differences in online sentence processing due to different sentential ordering, which is possible in Russian due to case marking.

3.2 Noun-final languages

Similar experiments have been conducted in noun-final Chinese, which has a processing advantage on subject relative clauses, whose analysis lean towards an expectation-based account of processing of relative clauses. Earlier studies accounted only for subject extracted subject relatives and object-extracted object relatives, wherein the object relative clauses were seen to be processed faster (Jäger 2015). Xu et al. (2019) reported online self-paced reading tasks in Chinese which show a marked preference for ORCs over SRCs, explained with the DLT model. They bring to question the claim that SRCs are processed faster cross-linguistically (assumed by the structural distance hypothesis), and highlight a need for crosslinguistic research. Wu et al. (2012) observed a facilitation with animate subjects and inanimate objects as heads, but a switch in animacy makes SRCs faster to process. Hsiao & Gibson (2003) found that for canonically ordered relative clauses, a memory-based account predicts that subject relative clauses would be processed slower, but they did observe the same for doubly-embedded relative clauses as well. Gibson & Wu (2013) observed statistically significant differences in the reduced processing speed of subject relative clauses, in Chinese relative clauses with a disambiguating context. Jäger et al. (2015) noted that a surprisal-based account for a sentence completion task does account for the faster processing of subject relative clauses over object relative clauses in

Chinese. In addition, Carreiras et al. (2010) also found an object RC advantage in Basque, and explained it using DLT.

Kwon et al. (2006) show that there is a processing advantage for subject gaps over object gaps in relative clauses and adjunct clauses in Korean, extending the cross-linguistic advantage in subject processing, and casts doubts on the notion of syntactically determined structural distance; as well as on the nature of the gap in relative clauses (trace vs. null pronominal bounded by a null operator) – with existing data not resolving the dichotomy.

Prideaux (1982) studied the processing of Japanese relative clauses and used both the closure strategy, wherein the parser prefers to close the node, once the end of a particular unit is encountered; and the normal form strategy, where the form that is presented to the parser is assumed to be the canonical form of the material to be parsed. This account using the closure strategy showed that object position RCs are processed faster, in opposition to the standard assumptions that lean towards a faster processing of subject RCs, due to the extracted element closing the matrix clause. The normal form account however shows the difference in processing speed exists due to the word order, and object extracted relative clauses were hence easier to parse in Japanese. However, the theories were in contrast with later observations that showed trends similar to English.

MacDonald & Montag (2009) conducted elicitation tasks in English and Japanese with native speakers. They showed that in the presence of an inanimate noun, the relative clause structures in English were more often passivized than in Japanese. Animate nouns were equally passivized in both languages. They posited that passivization for inanimate nouns occur more in English due to a higher frequency of passive formation in the language. In addition, they report that it could be due to the differential amounts of priming by the experimental task itself.

Ishizuka (2005) work on the processing of relative clauses in Japanese speaks of the different processing models and how some account for Japanese relative clauses, whereas others do not. Firstly, the relative clause in Japanese is accounted for with a null pronoun in the position of the gap noun.

(4) Japanese

a. e_i uma-o ketta $roba_i$-ga shinda
EC horse-ACC kick.PST mule-NOM die.PST
'The mule that kicked the horse died (Subject RC)'

b. uma-ga e_i ketta $roba_i$-ga shinda
horse-NOM EC kick.PST mule-NOM die.PST
'The mule that the horse kicked died (Object RC)'

Following a DLT account, subject RCs should be harder to process. The integration cost is higher for subject extracted RCs as there were more intervening elements. However, the results were in contrast, and were explained by the depth of embedding model (O'Grady (1997)), where the number of nodes measure the distance traversed. The Object RC pronominal is more deeply embedded as compared to the subject RC pronominal,

making it harder to access. This explained the experimental results. In addition, a temporary reading of an object RC as the main clause and backtracking to reanalyze it as an RC causes an increase in reading time. The study also noted that case matching conditions were processed significantly faster than case clashing conditions of extraction and position of RCs. Mansbridge & Tamaoka (2019), redid the experiment, but with an eye-tracking task, to obtain similar results.

4 Experiment

Since the universality of subject advantage in RC processing has been contested, working it out from a crosslinguistic perspective is essential. Malayalam sets a good stage for this experimentation. According to DLT, when object position object RCs are considered in Malayalam, encountering a second noun (in a sentence like 5b below) would increase the processing time at that point due to the unexpectedness of a second nominal in the nominative case. In the case of a subject position ORC (5a below), the position of the relative verb still leaves information to be desired about the object of the RC, increasing processing time.

In the case of subject RCs, overall surprisal could be equal but individual levels after each word is encountered may be lesser. This may be explained by how subjects often are the topics of sentences, and one expects more information regarding the subject, thus decreasing the amount of new information, by virtue of expecting new information. However, I also want to look into whether a memory-based parsing model is able to explain this difference as well.

4.1 Hypothesis and Prediction

Hypothesis: Object relative clauses in Malayalam are processed faster than subject relative clauses, and an expectation-based model would align with, and thereby explain, the observed results. However, subject RCs, due to their universal trends of being processed faster, will also be looked into from an integration-based perspective.

Prediction: The reading time measure for object extracted relative clauses would be lower as compared to that of subject extracted relative clauses. Overall comprehension of the sentence will be ORC biased, with markedness of structure playing a pivotal role in determining surprisal. This means, in a default-SOV language such as Malayalam, an order OS or VO order would be marked, hence increase the level of surprisal. This however can also clarify what can occur in the next region. For example, with an Object extracted ORC, surprisal is initially high as two nominals follow one another, but this resolves the structure as an RC, hence the RT would show a decreasing trend after this juncture.

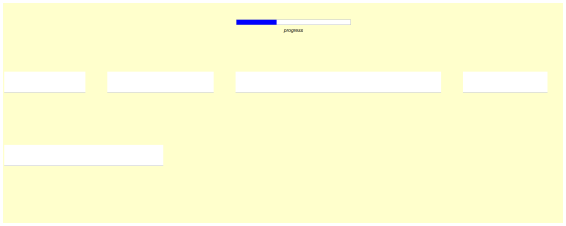


Figure 1: Sentence without the words is shown first

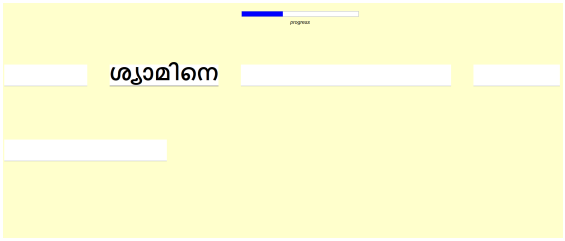


Figure 2: On pressing the SPACE BAR, the first word appears (not shown), and on pressing it again, it vanishes and the second word appears

4.2 Methodology

The task was a self-paced reading, moving window task. The stimuli were presented to the participants on PCIbex. Word(s) or phrases were shown on the screen on pressing the SPACE BAR. On a press, the reading time (the time taken from display of word to pressing the SPACE BAR, in milliseconds) would be recorded, and the word would disappear to show the next word in the sentence. Blank white rectangles indicating the whole sentence was provided for the participants at the start of each item. On the pressing SPACE BAR, the participant could move forward through the sentence regions. The masking was a modified dashed-sentence paradigm (Kush & Dillon 2021). The target sentences were presented in the Malayalam script.

Each experiment started with four unrelated practice trials for the participants to familiarize themselves with the procedure. Each item was followed by a comprehension question that asked for the correct answer, with two options given as full sentences. The participants could select the correct option by pressing the F- or the J- keys on their devices. The participants were given the choice of taking a short break every 16 items.

4.3 Task

16 sentences were shown to the participants, with 20 fillers and 40 sentences of other tasks (also acting as fillers), all counterbalanced. All experimental items were canonical SOV ordered sentences, with transitive verbs both in the RC and the matrix clause. The transitive verbs selected for a nominative and an accusative noun each, with the thematic roles nearly canonical Agent-Patient, for uniformity, and *the nouns used were all proper names*. All the sentences are hence appositive relative clauses. This was done because Malayalam does not allow for inanimate nouns to be morphologically accusative.

Two factors were controlled for the experiment viz:

1. Position of the RC in the sentences (subject vs. object) - to account for different configurations of RC occurrences.

2. Position of the gap of the noun from the RC (subject gap or subject extracted RC vs. object extracted RC)

A sample of the items with the forward slash indicating the breakdown of the regions of presentation is given below:

(5) a. RC in the Subject Position with an Object gap (SPOE)
darshana/ kaLi-ppi-cca/ Vanaja/ Ashaye protsaahippi-ccu
NPROP play-CAUS-PFV.PRT NPROP NPROP-ACC encourage-PFV
'Vanaja who Darshana played with encouraged Asha.'

b. RC in the Object Position with an Object gap (OPOE)
darshana/ Vanaja/ kaLi-ppi-cca/ Ashaye protsaahippi-ccu.
NPROP NPROP play-CAUS-PFV.PRT NPROP-ACC encourage-PFV
'Darshana encouraged Asha who Vanaja played with.'

c. RC in the Object Position with a Subject gap (OPSE)
darshana/ Vanaja-ye/ kaLi-ppi-cca/ Ashaye
NPROP NPROP-ACC play-CAUS-PFV.PRT NPROP-ACC
protsaahippi-ccu.
encourage-PFV
'Darshana encouraged Asha who played with Vanaja.'

d. RC in the Subject Position with a Subject gap (SPSE)
darshana-ye/ kaLi-ppi-cca/ Vanaja/ Ashaye
NPROP-ACC play-CAUS-PFV.PRT NPROP NPROP-ACC
protsaahippi-ccu.
encourage-PFV
'Vanaja who played with Darshana encouraged Asha.'

Counterbalancing was done as seen above. Only one type of each item was presented to each participant. The items were broken down into separate regions as can be seen in the example above, after each word, and the space bar had to be pressed to continue to the next region.

Comprehension questions for all the above sentences was to choose the more correct option (randomized in presentation) from below:

a. darshana aasha-ye protsaahippi-ccu
Darshana aasha-ACC encourage-PFV
'Darshana encouraged Asha.'

b. vanaja aasha-ye protsaahippi-ccu
Vanaja aasha-ACC encourage-PFV
'Vanaja encouraged Asha.'

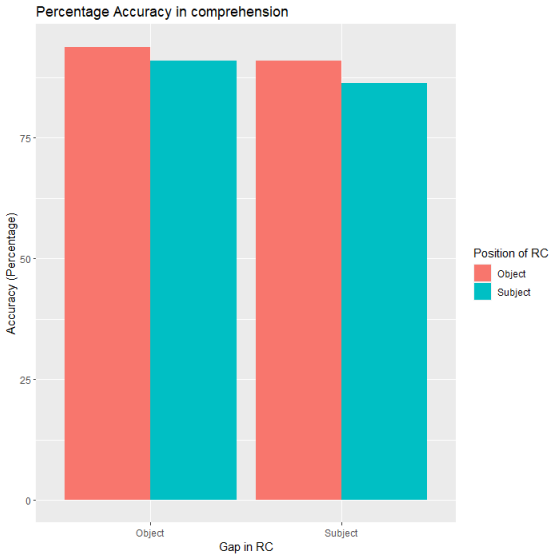


Figure 3: Graph with the mean accuracy in the Comprehension Task for each type of sentence

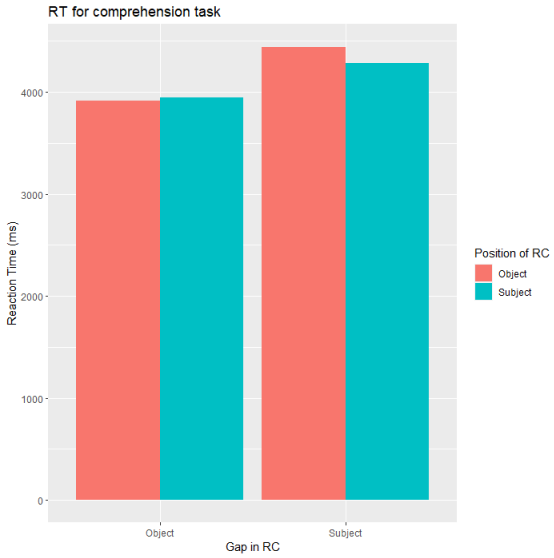


Figure 4: Graph showing the mean RT in the Comprehension Task for each type of RC sentence

One of the experimental design elements chosen was to present the accusative noun of the main clause with its corresponding verb. This design choice, in retrospect, shall be modified in future work.

4.4 Participants

Participants (n=53, male =18, $\geq$ 75% responses between 200-2500ms, and $\geq$ 67% accuracy for all items including fillers) read 4 sentences of each crossed factors (subject and object position vs. subject and object extraction), and the conditions were counterbalanced across the participants.

The participants fell within an age range of 18-53 years with mean age: 27.3 $\pm$ 8.5 years were recruited online. They were provided with a compensation of 25INR for their participation.

5 Results and Discussion

The data obtained was sorted first. The accuracy percentages were calculated, and the average RT for both the comprehension task, as well as of each region, was plotted against the corresponding region. Following this, I also performed a statistical analysis on R (2 factor repeated measures ANOVA), to account for statistically significant data.

These graphs (Figures 3 and 4) point to a higher accuracy in the comprehension tasks where the gap is in the object position, despite the time to complete the comprehension task being lower for these items. In addition, the lowest accuracy is obtained for subject RC,

with a gap in the subject position. This observation indicates a possibility that object RCs are easier to process due to the noun-final structure of Malayalam RCs.

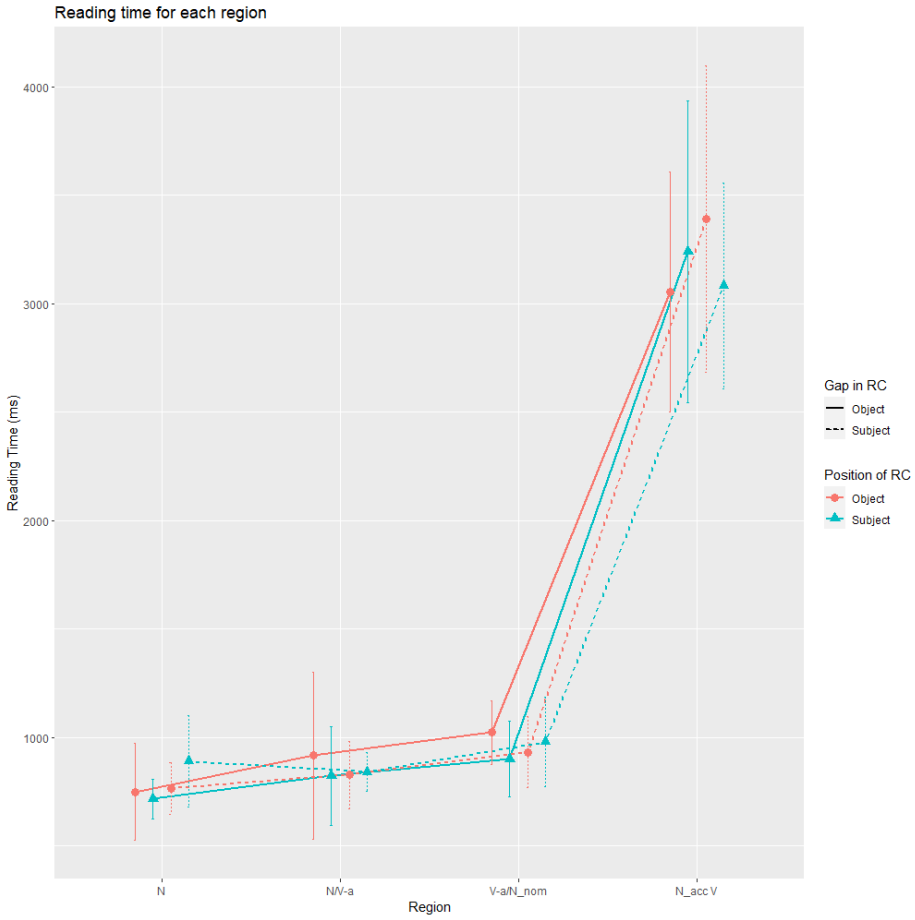


Figure 5: Reading time for each region: for sentences with the RC in the object position, the second region is the noun, and the third region is the verb (*V-a*), and for sentences with the RC in the subject position, the second region is the verb *V-a*

The graph (Figure 5) of the reading time for each region in milliseconds indicates a lot of interesting observations:

In the first region, the average reading time for the noun is the highest in the case of a Subject RC in the subject position. This can be attributed to the noun being accusative. The canonical order of sentences in Malayalam is SOV, hence encountering an accusative is an unexpected occurrence.

In the second region, it can be noticed that the highest RT is for sentences with an object gap RC in the object position (cf. 5b). It is not usual for sentences to have two nouns in the nominative to follow one another, without any intervening particle that shows conjunction or disjunction. This can be the reason for this observation.

In the third region, the RT for sentences with an object gap RC in the object position is still the highest. This can be explained by how the sentence has so far had two nouns in the nominative case. The verb is required, but the sentence remains unresolved.

In the final region, where all elements are put together, and the dependencies are exhausted, the trend is quite interesting. The highest RT is for sentences with a subject gap RC in the object position of the sentence. Sentences with the gap in the RC in a position congruent to the position of the RC modified noun take the least time (RC gap is in the object, and the RC modifies the object noun - or subject gap RC modifying the subject).

Figures 6 and 7 are the plots of the RTs for each element (noun or verb) in the relative clause, classified according to the type of relative clause (Gap) and the position of the RC, i.e., whether the RC modifies the noun or the verb of the matrix clause.

The plot of the reading times of RC nouns show a higher RT for the object RC mod-

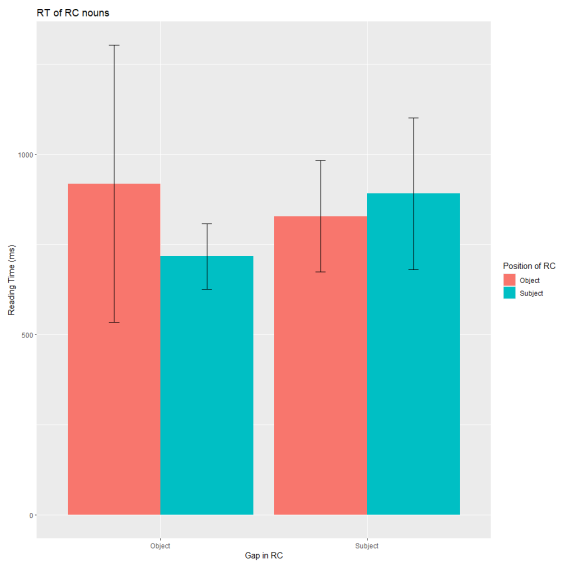


Figure 6: Mean Reading times of the nouns within the RC

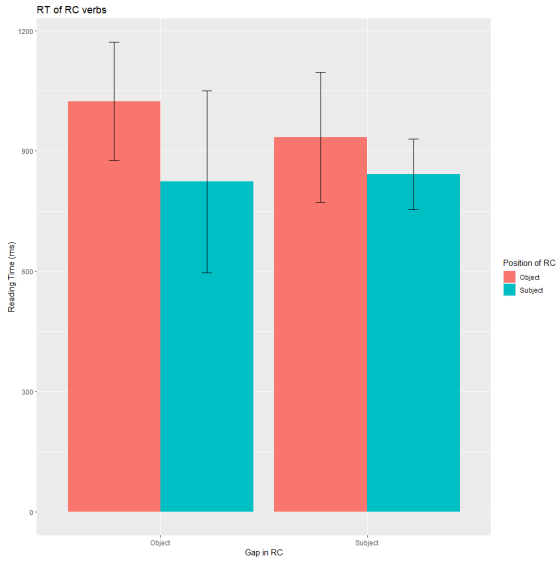


Figure 7: Mean Reading times of the RC participial verbs

ifying the object. This is attributable to the sentence being of the form (5b), as the RC nominative noun follows another nominative noun. It is also high for the subject gap RC modifying a subject, as the noun is accusative, and occurs as the first element of the sentence.

The plot of the reading times of the RC verbs also has an increase for the structure in (5b). The verb follows two nouns in the nominative case; hence it resolves one of the nouns' relations. This increases the RT as the other noun is still unresolved. As for the sentences with a subject gap in the RC with an object position, the time is higher as the structure expects

In addition to plotting the reading times, a two-factor repeated measures ANOVA was performed for RC nouns and RC verbs. The reading times are statistically significant in the following contexts:

1. on RC nouns across the type of gapping in the RC: $p = 0.048$
2. on RC verbs, across the position of the RC modifier (modifying the subject vs. the object): $p = 0.03$
3. on RC verbs, across the interaction of the gap as well as the position of the modifier: $p = 0.036$

In addition to the experimental data, the sentential items were also analyzed for the surprisal of each region. This has been plotted in Figure 8.

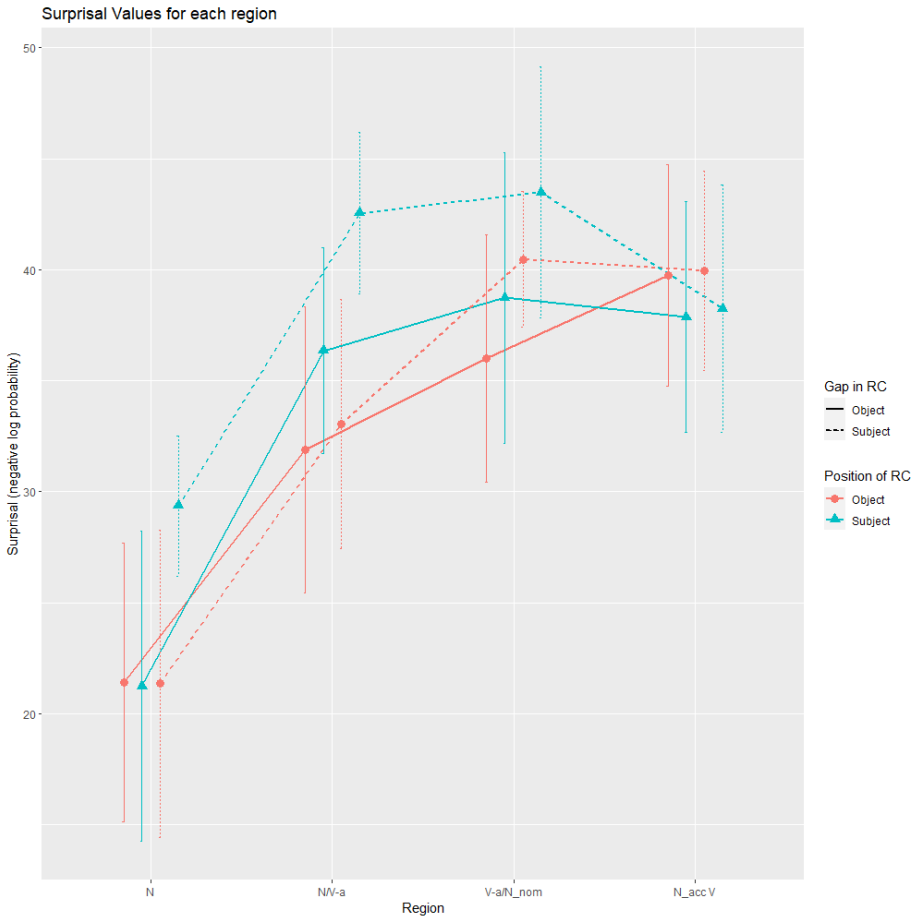


Figure 8: Surprisal at each region: for sentences with the RC in the object position, the second region is the noun, and the third region is the verb (*V-a*), and for sentences with the RC in the subject position, the second region is the verb *V-a*

The surprisal was calculated for the dataset using the *L3Cube Pune Malayalam Bert LLM*, with an incremental LM scorer. The surprisal analysis was performed with the minicons library in Python (Misra 2022). The surprisal analysis suggests that object gapped RCs (solid blue and red lines) are more oriented to expectations than subject gapped RCs (dashed blue and red lines). With regards to the position of the RCs, subject modifying RCs (both blue lines) tend to take more time initially, but it balances out in the end. This could be due to the relativized verb being encountered in the second region instead of later.

The surprisal data however shows an increase in the middle region instead of at the end of the sentence. The opposite observation is true for the experimental data. This discrepancy could be due to how in the case of absolute expectation values are more primed by the canonical structures, and non-canonical observations tend to increase the surprisal; whereas sentence termination comes with a decreased surprisal. In the case of experiments, however, there is an increase at the end due to the resolution of all dependencies, that occurs only when the matrix verb is read.

Gapping in RC	Position of Modifier	RT (ms)	Surprisal
Subject	Subject	5797.618	401.62
	Object	5919.849	383.09
Object	Subject	5683.16	354.8
	Object	5747.632	351.05

Table 1: Total RT and the total surprisal for each type of sentence

An analysis of the total RT and the total surprisal of each type of sentences have values as tabulated in Table 1, which shows a reduction in RT and surprisal in object gapped RCs over subject gapped RCs.

Analysing the memory load on each sentence type, it may be observed that the memory load is directly proportional to the number of unintegrated (with respect to integration with verbs) elements, at any point. From the items 5a, 5b, 5c and 5d, the memory load is highest in the RCverb for 5b, with two nominative nouns, followed by 5c. And with respect to the matrix verbs, all appear to have equal memory load (an RC modified proper noun and another noun) - with possible difference on where the modification occurs (on the subject or the object). This however does not translate directly, as the object modifier RCs have similar RTs at the matrix verb.

5.1 Summary of results

The analyses above show that the processing of Malayalam relative clauses seem to have a preference to Object RCs over Subject RCs during comprehension, and this can be explained with an expectation-based model, over a memory-based model. However, the relative clauses themselves seem to have a very slight subject preference, when the noun and the verb are being read. The work needs to be further refined, and more accurate methods employed, for more conclusive evidence.

6 Future Directions

This experiment offers evidence that supports an object RC preference in processing of Malayalam relative clauses. A theoretical analysis of RCs must follow to understand this observation, in addition to supplementing this with other behavioural tasks such as eye-tracking experiments and elicited production experiments. More importantly, tasks where a relative clause may be expected even before encounter, can provide much better evidence to the observations outlined above.

Acknowledgements

I would like to thank Prof. Shruti Sircar (English and Foreign Languages University, Hyderabad), my advisor during my master's dissertation, and Prof. Brian Dillon (University of Massachusetts, Amherst), who offered critical guidance for the task. Their mentorship and advice has played a huge part in the completion of this work. I also thank FASAL for selecting this work as a flash talk, and the comments from the three anonymous reviewers, that I have incorporated in the body of the text.

References

Bever, Thomas G. 1975. Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: Holistic and analytic. *Annals of the New York Academy of Sciences* 263(1). 251–262.

Carreiras, Manuel, Jon Andoni Duñabeitia, Marta Vergara, Irene De La Cruz-Pavía & Itziar Laka. 2010. Subject relative clauses are not universally easier to process: Evidence from basque. *Cognition* 115(1). 79–92.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain/MIT Press* .

Gibson, Edward & Gregory Hickok. 1993. Sentence processing with empty categories. *Language and Cognitive Processes* 8(2). 147–161.

Gibson, Edward & Tessa Warren. 2004. Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax* 7(1). 55–78.

Gibson, Edward & H-H Iris Wu. 2013. Processing chinese relative clauses in context. *Language and Cognitive Processes* 28(1-2). 125–155.

Gordon, Peter C, Randall Hendrick & Marcus Johnson. 2001. Memory interference during language processing. *Journal of experimental psychology: learning, memory, and cognition* 27(6). 1411.

Hale, John. 2001. A probabilistic earley parser as a psycholinguistic model. In *Second meeting of the north american chapter of the association for computational linguistics*, .

Hima, S. 2017. Relative clauses and/or participials in malayalam. *Strength for Today and Bright Hope for Tomorrow Volume 17: 7 July 2017 ISSN 1930-2940* 352. 133.

Hsiao, Franny & Edward Gibson. 2003. Processing relative clauses in chinese. *Cognition* 90(1). 3–27.

Ishizuka, Tomoko. 2005. Processing relative clauses in japanese. *UCLA Working papers in Linguistics* 13. 135–157.

Jäger, Lena, Zhong Chen, Qiang Li, Chien-Jer Charles Lin & Shravan Vasishth. 2015. The subject-relative advantage in chinese: Evidence for expectation-based processing. *Journal of Memory and Language* 79. 97–120.

Jäger, Lena Ann. 2015. *Working memory and prediction in human sentence parsing: cross-linguistic evidence from anaphoric dependencies and relative clauses*: Universität Potsdam dissertation.

Kush, Dave & Brian Dillon. 2021. Principle b constrains the processing of cataphora: Evidence for syntactic and discourse predictions. *Journal of Memory and Language* 120. 104254.

Kwon, Nayoung, Maria Polinsky & Robert Kluender. 2006. Subject preference in korean. In *Proceedings of the 25th west coast conference on formal linguistics*, 1–14.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177.

Levy, Roger, Evelina Fedorenko & Edward Gibson. 2013. The syntactic complexity of russian relative clauses. *Journal of memory and language* 69(4). 461–495.

MacDonald, Maryellen & Jessica Montag. 2009. Word order doesn't matter: Relative clause production in english and japanese. In *Proceedings of the annual meeting of the cognitive science society*, vol. 31 31, .

Mansbridge, Michael P & Katsuo Tamaoka. 2019. Ambiguity in japanese relative clause processing. *Journal of Japanese Linguistics* 35(1). 75–136.

Mathew, Rosmin. 2005. Simply malayalam participials. *CONSOLE XIV* .

Mecklinger, Axel, Herbert Schriefers, Karsten Steinhauer & Angela D Friederici. 1995. Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory & Cognition* 23(4). 477–494.

Misra, Kanishka. 2022. minicons: Enabling flexible behavioral and representational analyses of transformer language models. *arXiv preprint arXiv:2203.13112* .

Ouhalla, Jamal. 2004. Semitic relatives. *Linguistic inquiry* 35(2). 288–300.

O'Grady, William. 1997. Syntactic development. *University of Chicago* .

O'Grady, William. 2011. Relative clauses. *The Acquisition of Relative Clauses* .

Pesetsky, David. 2004. Tense, case, and the nature of syntactic categories.

Prideaux, Gary D. 1982. The processing of japanese relative clauses1. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 27(1). 23–30.

Reali, Florencia & Morten H Christiansen. 2007. Processing of relative clauses is made easier by frequency of occurrence. *Journal of memory and language* 57(1). 1–23.

Rohde, Hannah & William S Horton. 2014. Anticipatory looks reveal expectations about discourse relations. *Cognition* 133(3). 667–691.

Wu, Fuyun, Elsi Kaiser & Elaine Andersen. 2012. Animacy effects in chinese relative clause processing. *Language and Cognitive Processes* 27(10). 1489–1524.

Xu, Kunyu, Jeng-Ren Duann, Daisy L Hung & Denise H Wu. 2019. Preference for object relative clauses in chinese sentence comprehension: Evidence from online self-paced reading time. *Frontiers in Psychology* 10. 2210.

An Agree-based Analysis of Nominal Agreement: Evidence from Hindi-Urdu

MADHUSMITHA VENKATESAN, *Indian Institute of Technology Delhi*

ABSTRACT

There are two different ways of understanding DP-internal agreement: Either as a result of an independent feature sharing mechanism, or as a consequence of the same Agree-based mechanism that explains verbal agreement. This paper makes an empirical argument in favour of an Agree-based account of nominal agreement: the operation Agree explains the diverse empirical facts found in nominal agreement, and it is sensitive to the structure of the various modificational relations that exist within the DP. The empirical domain is provided by Hindi-Urdu where adjectives and possession indicating morphemes agree with the head noun of the DP. By exploring these agreement relations in detail, this paper presents a case for analysing nominal agreement as arising out of Agree - the same operation that is understood to underlie verbal agreement.

1 Introduction

Within the generative tradition, there is a prevalent tendency to separate agreement on verbs from agreement on other functional heads (Baker, 2008). The former has received considerable attention with multiple models of agreement mechanisms being constantly developed and refined, while the latter has often been relegated to the peripheries with the label 'Concord' attached to it.

This paper contests the view that nominal agreement, *i.e.* agreement on functional heads inside the DP, is executed by a separate operation. The stance adopted here is that agreement on adjectives, determiners and other functional heads in the DP should be explained using the operation Agree. Empirical support for this claim comes from Hindi-Urdu, an agreement rich Indo-Aryan language. The chief rationale for this proposal stems from the failure of existing feature percolation accounts of Concord to explain some of the agreement patterns found inside the DP.

Using empirical evidence from Hindi-Urdu, an agreement rich Indo-Aryan language, it demonstrates that agreement on adjectives, determiners and other functional heads in the DP can be executed by the operation Agree. Additionally, the paper also presents that existing accounts of Concord fall short in explaining for these facts.

The paper is organized as follows: Section 2 introduces nominal agreement in Hindi-Urdu, and lays out the patterns of agreement that will crucially inform this analysis. Section 3 explains how existing accounts of agreement deal with such cases of nominal agreement. Section 4 takes up the current proposal and demonstrates how the existing account fails whereas the Agree-based account explains all the facts laid out. Section 5 provides a summary and concludes the argument.

2 Agreement in the Nominal Domain

It is a well known fact that there are several elements in the nominal domain, or the DP, that display the features of the head noun in the DP. This instance of agreement, often referred to as Concord, is not one of full phi feature agreement; there is only number and gender agreement. Person agreement is not a characteristic of this relation. This paper will specifically focus on gender agreement inside the DP. Hindi-Urdu has grammatical gender; all nouns, regardless of their animacy status, are assigned an inherent gender value that is either [MASC] or [FEM][1]. This value is then derivationally seen on certain functional heads. The agreeing elements in the DP includes determiners, adjectives, possession markers, among others. The following examples are from Hindi-Urdu, where adjectives (1), possessive determiners (2) and possession markers (3)[2] agree with the gender value of the noun they modify.

(1) a. nay-aa ghar
new-MSg house.MSg
'new house'

b. nay-ii ghaDi
new-FSg watch.FSg
'new watch'

(2) a. mer-aa ghar
my-MSg house.MSg
'my house'

b. mer-ii ghaDi
my-FSg watch.FSg
'my watch'

(3) a. kaanc k-aa ghar
glass **PSP-MSg house.MSg**
'glass house'

b. kaanc k-ii ghaDi
glass PSP-FSg watch.FSg
'glass watch'

The patterns recorded in (1), (2) and (3) constitute nominal agreement, and will be the prime focus of this paper. To elaborate, this paper will look into the mechanisms by which such agreement patterns are generated, and propose that these patterns do fall under the orbit of the operation Agree.

[1]For ease of exposition, all instances of [MASC] agreement are shown in bold typeface and all instances of [FEM] agreement are shown with underlined text

[2]Some possession relations in Hindi-Urdu are indicated by post-positions. This is glossed as PSP in all the forthcoming examples.

3 Conventional Approaches to Nominal Agreement

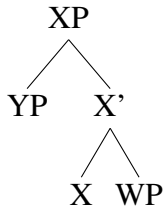
Arguments against Agree in the DP come from two major sources: Giusti's (2008) account of adjectival agreement in Bantu and Romance languages, and Norris's (2017a, b) account of nominal agreement in Estonian. In this section, we shall discusses both of those.

3.1 Giusti (2008)

Giusti (2008) develops an analysis for nominal agreement which is fundamentally based on the assumption that agreement on verbs is executed by a mechanism different from agreement on adjectives. At the outset, Giusti clarifies that the former involves a specific probing operation, matching under Agree and ultimately movement of the matched features. Adjectival agreement, or Concord, according to this account arises when a modifier with uninterpretable phi features is merged into the structure.

In this account, it is also argued that feature sharing between the head and its modifiers is done as a result of the feature of the head percolating to the modifiers, which are merged as specifiers of the head. This is illustrated below in (4), where X is a head and YP its modifier. The features of X are expected to percolate to YP by virtue of the Spec-Head configuration that they are in. There is no probing on the part of the modifier.

(4)

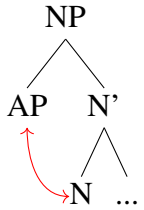


Using examples such as (5) from Italian, Giusti posits that features are transferred from a head to its specifiers. In (5), the determiner 'the', and the adjectives 'Italian' and 'beautiful' agree with the [FEM.PL] features of 'friends'.

(5) le belle amiche italiane di Maria
the.fem.pl beautiful.fem.pl friends.fem.pl Italian.fem.pl of Mary
'Mary's beautiful Italian friends' Italian(Giusti, 2008)

Giusti explains these agreement patterns as a result of feature sharing under a spec-head configuration, when the AP is merged as the specifier of NP, as shown below in (6)

(6)



Thus, according to this account, the structural proximity between the modifier and the head is not considered as a factor of the resulting nominal agreement. Features are expected to be shared between them simply by virtue of the modifier being located in the specifier position of the head in question.

However, with novel empirical data from Hindi-Urdu, it becomes apparent that such an account of feature sharing fails to generate several patterns of agreement found in natural language, thus forcing us to reject it in favour of a more nuanced approach.

3.2 Norris (2017)

Norris (2017a,b) also supports the idea that nominal agreement should be treated distinctly from verbal agreement. In his account of nominal agreement in Estonian, Norris presents some conceptual arguments against equating adjectival and verbal agreement. The main objection for treating the two as the same stems from the fact that while verbal agreement is expressed only once in a sentence (on the verb), adjectival agreement can occur multiple times in the same DP. Secondly, agreement within the DP occurs across a variety of categories - adjectives, adverbs, determiners etc. - and this diversity cannot be accounted for by Agree. Thirdly, verbal agreement has an established connection with structural case; nominal agreement does not have any such correlation, and therefore the two must be treated differently.

It is here that I depart from Norris's account; while these differences between nominal and verbal agreement are impossible to ignore and must be used to keep the two distinct, they do not imply that the underlying mechanism between the two be entirely different. It is certainly plausible for agreement on verbs be executed by the same underlying operation, and yet for the two to manifest differently in the derivation. Thus, I adopt the stance that the operation Agree can derive patterns of nominal and verbal agreement without obviating the aforementioned conceptual distinctions between the two.

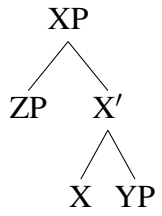
Further, this paper goes on to demonstrate that certain patterns can only be derived by an Agree-based account. This presents a strong albeit purely empirical argument in favor of understanding nominal agreement as a result of the same mechanism that has served explanations of verbal agreement so well.

4 The Current Proposal

In this work, I propose that instances of agreement such as the cases in (1), (2) and (3), which are traditionally explained as instances of Concord, should be in fact reanalysed as falling out from the operation Agree. Such an approach is not unprecedented; Toosarvandani & Van Urk (2014) have demosntrated for Zazaki, an Iranian language that an Agree-based account is the ideal one for explaining all the agreement patterns obtained in the language. Similarly, Carstens (2001) also presents the merits of an Agree based approach to analyse agreement on adjectives and possessors, using empirical support from Bantu languages.

The model of Agree adopted in this paper is referred to as Bidirectional Agree (Baker, 2008). According to this model, A probe bearing unvalued case or phi features can either C-Command or be C-Commanded by a goal that bears the matching valued features. Essentially, the relative hierarchical location of the probe and the goal does not matter as long as one of them C-Commands the other; Agree can happen in a configuration where either the probe is located higher than the goal, or the goal is located higher than the probe. This is illustrated below in (7), where the probe in X can potentially agree with two goals: one that it C-Commands (YP) and one that it is C-Commanded by (ZP). Between these two options, factors such as locality (closest out of the two) will play a role in determining ultimately which one of the two the probe agrees with.

(7)



In the following subsections, I show that this model of Agree can successfully derive all the patterns outlined above in Hindi-Urdu. This will be demonstrated by considering four domains of agreement within the DP, namely (i) Simplex adjectives (ii) Complex adjectives (iii) Possession indicators (iv) Participial modifiers inside the DP. Each of these domains will be explored individually in the following subsections.

4.1 Simplex Adjectives

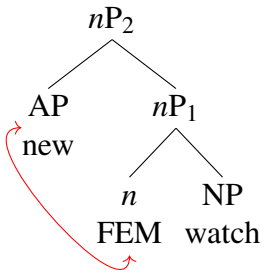
The first domain of agreement to be considered is the one where there is one adjective modifying a noun, such as the examples in (1) repeated here as (8). In (8a), the adjective 'new' agrees with the [MASC] feature of 'house' and in (8b) it agrees with the [FEM] feature of 'watch'.

(8) a. nay-aa ghar
new-MSg house.MSg
'new house'

b. nay-ii ghaDi
new-FSg watch.FSg
'new watch'

A structural representation of (8b) is given in (9), where the adjective 'new' is hosted as an adjunct specifier of the *n*P that houses the gender feature of the noun.

(9)



The feature percolation account propounded by Norris and Giusti would account for such cases by positing that the [FEM] feature of the noun percolates upwards to the adjective, as features are capable of percolating from head to specifier. An Agree account, on the other hand, would derive this agreement as a result of a probing exercise by the adjective. The adjective containing a probe for phi features extends a search for valuation. The closest item with matching valued features is the NP with its phi features placed at *n*. Agreement on the adjective can then be traced back to the probe-goal relation between the two.

Essentially, what we see is that in the case of simplex adjectives such as (8) above, both accounts are equally adequate at explaining the facts. In fact, a cursory evaluation of the two, the feature percolation account could appear as the more simple and natural one, and therefore the more appealing option of the two. Consequently, the feature percolation account gained widespread acceptance within the theory.

While this is perfectly logical, I go on to show that when we increase the complexity of the DP internal modifiers, the feature percolation account fails to hold up. In the following subsections, this shall be demonstrated by incrementally increasing the complexity of the DP. The first step would be to add another layer of modification. Then comes the domain of possession markers and then finally both the theories will be put to test with the introduction of a participle verb into the DP.

4.2 Complex Adjectives

This section is about what happens when we add another layer of modification in the DP, creating a complex adjective. What we now have are two modifers in the same DP, as illustrated in (10) below. The presence of two modifiers could lead to two distinct interpretations of the DP: one where both the modifiers iteratively modify the head noun (10a) , and another where the lower (and linearly second) modifier describes the head noun and the higher (and linearly first) modifier qualifies the lower modifier (10b).

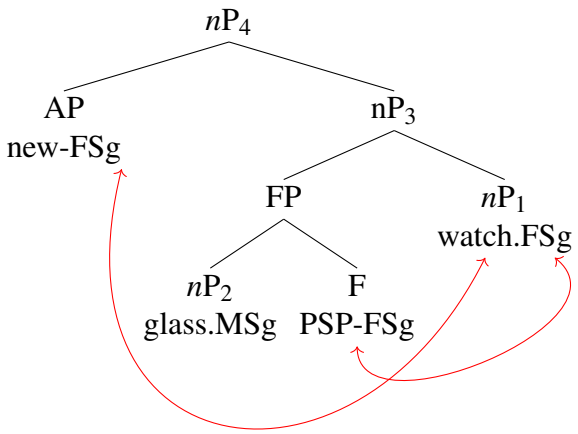
(10) a. nay-ii kaanc k-ii ghaDi
new-FSg glass.M PSP-FSg watch.FSg
'new watch which is made of glass'

b. nay-e kaanc k-ii ghaDi
new-M.OBL glass.M PSP-FSg watch.FSg
'watch which is made of new glass'

The two interpretations listed above have correlates in agreement patterns too. In (10a), where the head noun 'watch' is iteratively modified by both 'new' and 'of glass', both of these functional heads agree with the [FEM] value of 'watch'. Similarly, in (10b) where 'of glass' modifies 'watch' and 'new' modifies 'glass', the PSP morpheme in 'of glass' agrees with the [FEM] value of 'watch' and 'new' agrees with the [MASC] value of 'glass'.

Thus, the structural representation of the two are distinct. (10a) is represented as (11) below, where AP as well as PSP agree with the head noun 'watch'.

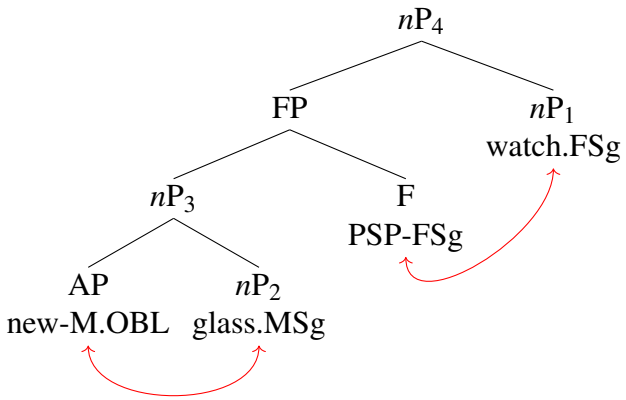
(11)



Once again, both, the feature percolation account as well as the Agree-based account can explain this pattern. According to the former, the gender values of the noun 'watch' will be percolated to its specifiers, AP and FP. This results in the obtained agreement patterns. According to the latter, the probes in F and AP find nP$_1$ as their closest goal, and therefore agree with it.

However, the picture is slightly different in the case of the alternative arrangement (10b), where two different agreement patterns exist in the same DP. (10b) is structurally represented as (12), where the adjective 'new' modifies and agrees with 'glass' and only PSP on 'glass' agrees with the head noun 'watch'.

(12)



The feature percolation account falters here, as it does not predict this pattern at all. According to standard percolation methods, features of a noun simply percolate to its specifier. The specifier here is FP. Thus, the percolation account expects all the elements internal to FP to agree with the head noun 'watch'. However, that is not what happens. There is an independent agreement relation taking place between two internal constituents of FP (AP and $n\text{P}_2$), which is beyond the ambit of the feature percolation account. An Agree-based account, on the other hand, can successfully account for both the agreement patterns obtained here. As expected by the Agree-based account, each probe would agree with the closest goal that it either C-Commands, or is C-Commanded by. With that in mind, we see that the closest goal to the probe in AP is $n\text{P}_2$ 'glass' and the closest goal to the probe in F is $n\text{P}_1$ 'watch', and thus, we get the appropriate agreement patterns. A feature percolation account is unable to capture such nuances, as we shall further demonstrate by introducing possessors into the equation.

4.3 Introducing Possessors

We shall now consider how the two accounts fare with the addition of another element in the DP: possessors. In Hindi-Urdu, possessive determiners agree with the phi features of the possessum, as illustrated in (2) above. Consider the examples below, where the possessor is 'my' and the possessum is 'watch'. In (13a), all the modifiers iteratively modify the head noun 'watch' and also exhibit agreement with its [FEM] value. In (13b), the possessive determiner and PSP modify 'watch' and agree with it, whereas 'new' qualifies 'glass' and agrees with it. The final pattern (13c) is an ungrammatical one, where the possessive determiner agrees with 'glass' while the other modifiers agree with 'watch'.

(13) a. mer-ii nay-ii kaanc k-ii ghaDi
my-FSg new-FSg glass.MSg PSP-FSg watch.FSg
'my new glass watch' (my watch which is new and made of glass)

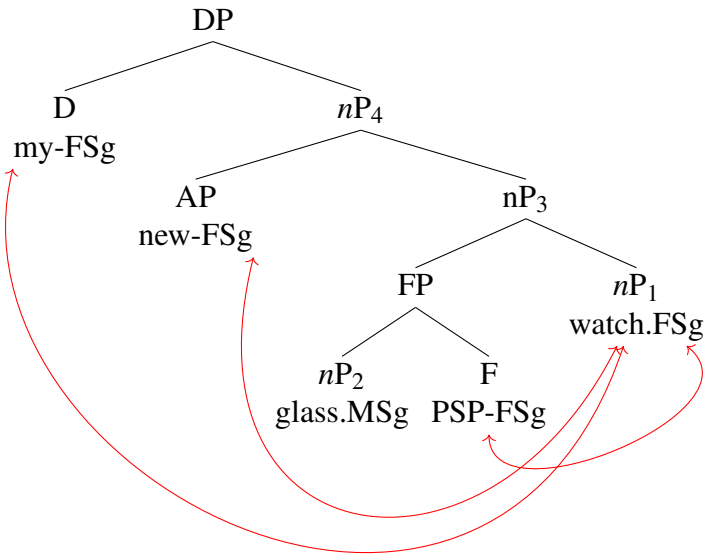
b. mer-ii nay-e kaanc k-ii ghaDi
my-FSg **new-M.OBL** **glass.MSg** PSP-FSg watch.FSg
'my watch made of new glass'

c. *mer-e nay-ii kaanc k-ii ghaDi
my-M.OBL new-FSg **glass.MSg** PSP-FSg watch.FSg
'(Intended: new watch made of my glass)'

A cursory look at these facts suggests that there is nuance in this agreement system which cannot be captured efficiently by a simple percolation account. The presence of such variation suggests that structural factors must be at play, and we shall probe further into each of them.

The iterative agreement pattern in (13a) is represented as (14) below, where the possessive determiner, adjective and PSP, all agree with the [FEM] value of the head noun 'watch.'

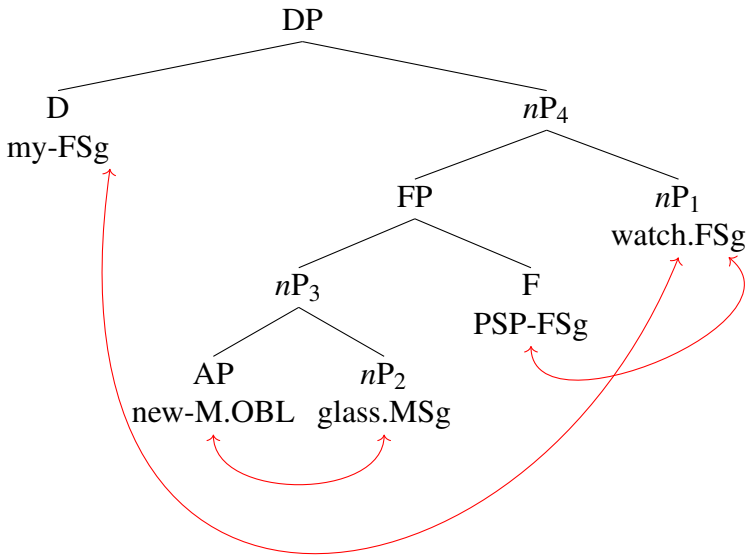
(14)



This pattern can be explained equally well by both, the feature percolation and the Agree-based accounts. According to the former, the features of the noun percolate to all of its modifiers, including the possessive determiner. The Agree-based account posits that there are probes in all the functional heads, and that agreement happens with the closest C-Commanding *n*P, which is 'watch' when the structure is built bottom up. This particular pattern is not helpful in evaluating between the two accounts.

However, the choice between the two becomes immediately clear when we consider the agreement patterns in (13b). This is structurally represented as (15): there are two distinct agreement relations at play. 'new' qualifies glass, whereas 'my' and PSP modify 'watch', and this correlates with the agreement patterns obtained.

(15)



A feature percolation account would expect the features of the head noun to percolate to all the modifiers inside the DP - the determiner, adjective as well as PSP. However, as outlined in (12), there is an independent agreement relation between AP and nP$_2$, which is distinct from agreement with the head noun. Thus, (15) is another instance where the feature percolation account is inadequate while explaining agreement patterns inside the DP.

The Agree account then emerges as the more suitable alternative: According to this approach, the probe in AP locates nP$_2$ in its closest search domain. For the other probes, namely D and PSP, the closest available nP is 'watch', and they agree with its [FEM] value. These two agreement patterns have the ancillary benefit of capturing the modificational relations in the DP too; 'new' modifies and agrees with 'glass', and 'my' and PSP modify and agree with 'watch'.

(13c), which is an ungrammatical sentence, can also be explained by an Agree-based account. Under the given structural configuration, the probe in D will never find 'glass' as its closest goal. Thus, [MASC] agreement on D will never be obtained, as predicted correctly by the Agree based account. This is another instance that a feature percolation account would not be able to foresee and preempt.

Essentially, adopting an Agree-based account makes it possible to formally capture the relation between modificational and agreement patterns, and this is done using the fundamental principle of Closest C-Command. We do not have to invoke additional conditions or constraints unique to DP internal constituents to explain these facts. In the next subsection, I present the case of yet another agreement relation inside the DP, and compare the two accounts in those terms.

4.4 Participial Modifiers

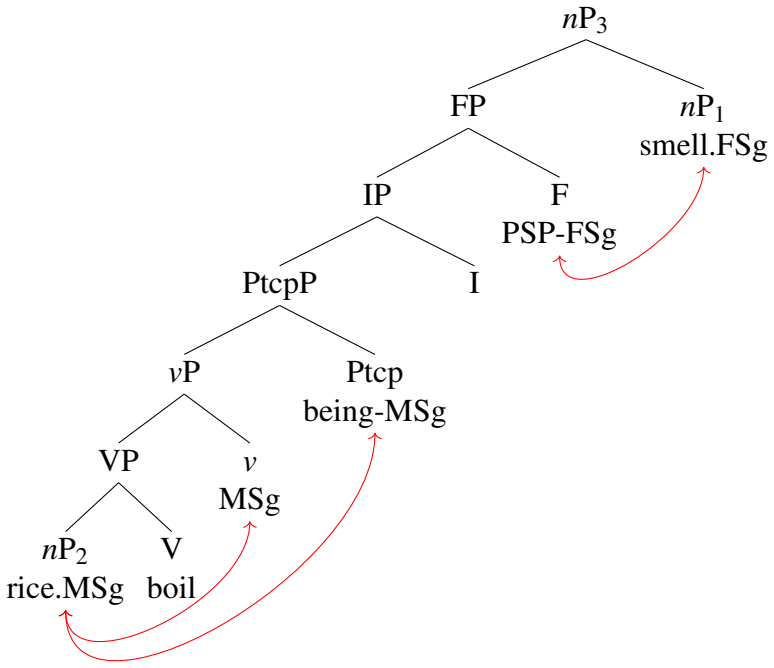
We have already seen in Sections 4.1, 4.2 and 4.3 that the feature percolation account fails to explain the existence of more than one agreement pattern in a DP. An Agree-based account that takes into consideration structural relations correctly predicts these facts. In continuation with that, another piece of evidence in support of an Agree-based account comes from DPs containing a non-finite participle, such as (16a) and (16b) below. In (16a), there are two agreement relations inside the DP. The participial form of the verb 'boil' and the non finite AUX 'being' agree with 'rice' whereas PSP agrees with 'smell'. (16b), where all the agreeing heads exhibit [FEM] agreement, is ungrammatical.

(16) a. ubalt-e hu-e caawal k-ii khushboo
boil.PTCP-M.OBL being-M.OBL rice.MSg PSP-FSg smell.FSg
'the smell of rice being boiled'

b. *ubalt-ii hu-ii caawal k-ii khushboo
boil.PTCP-FSg being-FSg **rice.MSg** PSP-FSg smell.FSg
(Intended: the smell of rice being boiled)

The structural representation of (16a) is given below in (17), where the probe in PSP agrees with the head noun 'smell', whereas the probes in *v* and Ptcp agree with 'rice'.

(17)



An Agree-based account can efficiently explain these patterns; in a bottom up approach to structure building, the closest goals for the probes in Ptcp and *v* is nP_2 inside the IP. The closest goal for the probe in F is nP_1. These agreement patterns are also reflective

of the modificational relations inside nP_3: the elements inside the IP modify the object of the verb 'rice', and FP is a modifier of the head noun of the *n*P 'smell'. Once again, we see that the Agree-based account equips us to find structural correlates for the different modificational relations in a domain. Essentially, the Agree-based account thus manages to offer a uniform explanation for different agreement relations because it is sensitive to the underlying structural relations, a lacuna in the percolation approach.

There is some cross-linguistic evidence coming from the same domain of participial agreement. Polinsky (2016) presents DPs such as (18) from Archi. Nouns in Archi are specified with a noun class marker, which is then reflected in verbal agreement. (18) contains two agreement patterns within it: there are two probes in the participial form of the verb 'bake'; one of them agrees with the object of 'bake' (bread = (noun class III)) and the other agrees with the head noun 'smell' (smell = noun class IV).

(18) Archi

x:alli b-aˆcar-t:u-t di
bread.III.SG.ABS III.SG-bake-IPFV-ATTR-IV.SG smell.IV.SG.ABS

'the smell of bread being baked'

In this instance, agreement between the object 'bread' and the participle 'bake' is analysed as an anomaly, because it is not expected under the feature percolation approach. However, when we turn to the Agree-based account, we have a way to explain these patterns sufficiently: the probe in 'bake' finds 'bread' in its C-command domain and agrees with it. The other instance of agreement in (18), between the same verb and 'smell', can be explained as 'smell' being in the C-command domain of the entire verbal complex.

This goes on to show yet another instance where the Agree-based acccount can capture the agreement patterns effectively with the help of a single underlying mechanism: Probe-Goal relations under closest C-command.

5 Conclusion

To summarise, we see multiple instances of agreement patterns inside the DP where the feature percolation account fails to hold up because it cannot explain complex agreement patterns. The Agree-based account prevails because it is sensitive to underlying structural relations and can provide sufficient motivation for each instance of agreement relation between two items. This point became evident as we considered cases of complex adjective agreement, possessor agreement and participial agreement inside the DP. An Agree-based account rooted in structural relations such as C-Command was able to provide explanations for all the agreement patterns outlined above.

Having an Agree-based account for DP internal agreement is also beneficial to a general theory of agreement, as it lets us explain nominal and verbal agreement using the same mechanism, without having to invoke special constraints for the former, thus reducing the overall theoretical machinery needed to explain agreement in natural language.

Acknowledgements

I would like to thank the attendees of FASAL-14 for their insightful questions and comments which both, challenged and contributed to this analysis greatly.

References

Baker, Mark C. 2008. *The syntax of agreement and concord*, vol. 115. Cambridge University Press.

Carstens, Vicki. 2001. Multiple agreement and case deletion: Against φ-incompleteness. *Syntax* 4(3). 147–163.

Giusti, Giuliana. 2008. Agreement and concord in nominal expressions. *The Bantu-Romance Connection* 201–237.

Norris, Mark. 2017a. Description and analyses of nominal concord (pt i). *Language and Linguistics Compass* 11(11). e12266.

Norris, Mark. 2017b. Description and analyses of nominal concord (pt ii). *Language and Linguistics Compass* 11(11). e12267.

Polinsky, Maria. 2016. Agreement in Archi from a minimalist perspective. *Archi: Complexities of agreement in cross-theoretical perspective* 184–232.

Toosarvandani, Maziar & Coppe Van Urk. 2014. The syntax of nominal concord: What ezafe in Zazaki shows us. In *Proceedings of NELS*, vol. 43, 221–234.