

Proceedings of  
**(Formal) Approaches to South Asian Languages 13**

Held in 2023 at the University of Michigan, Ann Arbor, MI, USA

Edited by Samopriya Basu & Aaditya Kulkarni  
June 2024

## Preface

The 13<sup>th</sup> Conference on (Formal) Approaches to South Asian Languages was hosted by the Department of Linguistics at the University of Michigan, Ann Arbor, between the 31<sup>st</sup> of March and the 2<sup>nd</sup> of April, 2023. The mode of the conference was *hybrid*, giving participants who could not be physically present at the in-person venue the opportunity to attend online. This proceedings volume collects articles stemming from eleven of the thirty-odd talks of various lengths given at the conference, and one additional paper which was originally scheduled to be presented but the authors could not make it at the time and yet were generous enough to submit it in full for publication here.

The papers the reader will find in this volume cover a diverse spectrum of topics around the theoretical analysis of South Asian languages. Represented are studies in language contact (KARTIK-NARAYAN, VENKATESAN), morphology (AGARWAL, DUTTA-BANERJEE, SYED-LEE), syntax (DEES, HALE-KISSOCK, ISHIKAWA-YOSHIDA, SCHMURA, SURESH) and semantics (BANERJEE-BANERJEE-BHATT, KIDWAI-SOBOLAK) — the year’s special theme of “language contact and multilingualism” shining through some of them. Furthermore, the languages represented in these papers showcase the inherent diversity in South Asian linguistics almost as much as the topics being analyzed — chronologically and genealogically — ranging from Tamil diglossia and Sanskrit to the modern vernaculars of Hindi-Urdu, Bengali, Kodava, Kannada and more. ISHIKAWA & YOSHIDA, notably, take a typological approach treating a few different languages comparatively.

It is no secret that organizing an event of this magnitude featuring participants from several countries and across two modes (physical and virtual) is no easy feat, and we as editors express our heartfelt thanks to Dr. Savithry Namboodiripad and the rest of the organizing team at the University of Michigan and beyond for accomplishing this complex task without any glitch. We also thank the conference participants and presenters, and especially the authors of the papers submitted to the present proceedings. We certainly enjoyed reading through them ourselves and are confident they will serve as valuable scholarly resource to the South Asian linguistics community, inducing further inquiry and research. Finally, we would like to apologize for the delay of a couple of months in publication of this volume. The fault is ours and ours alone, and we can only hope that the quality of linguistics from our contributing colleagues contained in here will somewhat distract from the shortcomings of the editors.

— Samopriya Basu & Aaditya Kulkarni

## Contents

- **Dependent dative case in Hindi-Urdu** 1  
Hashmita AGARWAL
- **Egocentric questions: The view from Bangla and Hindi-Urdu** 17  
Arka BANERJEE, Kousani BANERJEE, Rajesh BHATT
- **Kannada through the lens of the NP/DP parameter** 37  
Joshua DEES
- **'-Te/-e' marker as a plural in Bangla** 56  
Chandni DUTTA, Sayantani BANERJEE
- **The internal syntax of *iva*-clauses in Vedic Sanskrit** 75  
Mark HALE, Madelyn KISSOCK
- **Relative clause constructions in New Indo-Aryan languages:  
Hierarchies of macro roles** 83  
Sakura ISHIKAWA, Shigeki YOSHIDA
- **Language contact and sound change: Reasons for mutual  
unintelligibility between formal and colloquial registers of Tamil** 103  
Anushri KARTIK-NARAYAN
- **The unique functionality of Urdu light verb *jaa* and Voice head  
variation** 107  
Sana KIDWAI, Frances SOBOLAK
- **A preliminary description of *vanthu* in Spoken Tamil** 129  
Suganya Rajendran SCHMURA
- **DOM in Kodava takk: A complex interaction among multiple factors** 143  
Saurov SYED, Catherine LEE
- **Processing of stacked NPs in embedded sentences in Malayalam** 162  
Revathi SURESH
- **Gender agreement in a Tamil-Hindi bilingual situation: The role  
of feature valuation** 183  
Madhusmitha VENKATESAN

# Dependent dative case in Hindi-Urdu

HASHMITA AGARWAL, *University of California Los Angeles*

## ABSTRACT

While dative case has traditionally been analysed as a case assigned to a DP by a head (Chomsky 1981, 1986; Woolford 2001, 2006), Baker & Vinokurova (2010) and Baker (2015) have argued that dative case in Sakha is a dependent case in the sense of Marantz (1991). Following Baker & Vinokurova (2010)'s analysis of Sakha, this paper proposes a dependent case analysis of dative case in Hindi-Urdu, based on crucial evidence from the causativised ingestive construction. This account is novel support for the view that dative case may be a dependent case in some languages.

## 1 Introduction

'Dative case' most commonly refers to the case assigned to the indirect object in ditransitives, but a diverse range of arguments crosslinguistically are marked with what is called 'dative case' in a given language, and these cases demonstrate distinct morphosyntactic behaviours.

For example, dative is termed a syntactically inactive inherent case in Modern Greek, since it is tied to the theta role of a goal in ditransitives, and does not disappear in passives or unaccusatives. (Chomsky 1986; Anagnostopoulou 2003; Alexiadou et al. 2014). In Japanese, dative case is suggested to be a structural case on goals in ditransitives, while it is inherent in monotransitives—since it alternates with nominative only in ditransitives (Woolford 2006; Fukuda 2007; Ishizuka 2010; Alexiadou et al. 2010). Crucially, under all of these traditional analyses, dative case is assigned by some functional head to some very proximal DP. (Chomsky 1981, 1986; Woolford 2001, 2006).

In contrast to the traditional view of dative as a functional head case, Baker & Vinokurova (2010) and Baker (2015) have proposed that in some languages like Sakha, dative case is a 'dependent case' that surfaces on a DP when it is in a particular configuration with regard to other DPs and elements in the structure. In particular, they propose that whenever two DPs unvalued for case are in a c-command relationship in the same local domain, one of the DPs is assigned dependent case. Baker & Vinokurova (2010) and Baker (2015) build on Marantz (1991)'s disjunctive case hierarchy, in which they name 'dependent case'—also based on the relative positioning of DPs in the structure—to be one of the possible modalities of case assignment.

In this paper, I argue that dative case in Hindi-Urdu is a dependent case in all its instances like in Sakha—due to similarities in the environments dative case appears in in both languages, as well as parallel morphosyntactic behaviour under passivisation. The dependent dative case rule I argue for is stated in (1).

- (1) **DATIVE CASE RULE:** If DP<sub>1</sub> c-commands DP<sub>2</sub> in the complement of *v*, assign dative to DP<sub>1</sub>

The characterisation of all dative case in Hindi-Urdu as a dependent case lends support to the view that along with being a functional head case of some kind in different languages (or across different constructions in the same language), dative case may also be a dependent case in some languages.

The structure of this paper is as follows: I outline the distribution of dative case in Hindi-Urdu, followed by syntactic diagnostics to distinguish dative case from accusative case in Hindi-Urdu, which are both realised as *-ko*. Then, I argue through the causativised ingestive construction that dative case cannot be a functional head case due to the lack of an assigning head. Lastly, I propose that dative is a dependent case assigned via the rule in (1), and show that only a dependent case analysis of the dative in Hindi-Urdu sufficiently captures its distribution.

## 2 Distribution of dative case

In ditransitives in Hindi-Urdu, the indirect object invariably receives dative case (2), which is syncretic with accusative *-ko*, (3).

- (2) *Dative on indirect object*  
 miina=ne ṭiina\*(=**ko**) kitaab di-i  
 Mina=ERG Tina\*(=DAT) book give-PFV  
 ‘Mina gave Tina a/the book’

- (3) *Accusative on direct object*  
 miina=ne fuul(=**ko**) dekh-aa  
 Mina=ERG flower(=ACC) see-PFV  
 ‘Mina saw a/(the) flower’

Dative *-ko* and accusative *-ko* may also coappear in the same clause (4), as Bhatt & Anagnostopoulou (1996) have shown. Note that the shifting of the direct object over the indirect object in (4)—which is required for the former to be marked accusative in ditransitives—is discussed in detail in Bhatt & Anagnostopoulou (1996), and will not be addressed further here, since it is independent of dative case assignment.

- (4) *Dative and accusative case in ditransitives*  
 miina=ne kitaab=**ko**<sub>j</sub> ṭiina=**ko** \_\_\_<sub>j</sub> di-yaa  
 Mina=ERG book=ACC<sub>j</sub> Tina=DAT \_\_\_<sub>j</sub> give-PFV  
 ‘Mina gave Tina the book’

In addition to goals in ditransitives, dative case also appears on experiencer arguments in experiencer constructions, (5) (Davison 2004).

- (5) a. *Experiencer construction*  
 mona\*(=ko) bukhaar hai  
 Mona\*(=DAT) fever AUX.PRES  
 ‘Mona has a fever’
- b. *Experiencer construction*  
 mona\*(=ko) siima dikh-ii  
 Mona\*(=DAT) Sima appear-PFV  
 ‘Mona saw Sima’ (Lit.: ‘Sima appeared to Mona’)

Ingestives—like in (6a)—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 (6b).<sup>1</sup> *Salma* is the added causer argument in (6b), while *kutte* ‘dog’ becomes the causee. Importantly, dative case is found in causativised ingestives—the causee DP *kutte* ‘dog’ in (6b), which is sandwiched between *Salma* and *seb* ‘apple’ is marked dative.

- (6) a. *Ingestive*  
 kutte=ne seb khaa-yaa  
 dog=ERG apple eat-PFV  
 ‘The dog ate an apple’
- b. *Causativised ingestive*  
 salma=ne kutte\*(=ko) seb khil-aa-yaa  
 Salma=ERG dog\*(=DAT) apple eat-CAUS-PFV  
 ‘Salma fed the dog an apple’ (Lit.: ‘Salma made the dog eat an apple’)

Now that I have specified the environments in which dative case occurs in Hindi-Urdu, I will explore diagnostics that help distinguish between dative and accusative *-ko*.

## 2.1 Dative *-ko* vs. accusative *-ko*

Dative *-ko* is syncretic with the accusative/DOM *-ko* on direct objects, but they are syntactically distinct cases (Mohanani 1994; Davison 2014, pace Kalin 2014). Evidence for the distinctness of dative *-ko* and accusative *-ko* comes from dative case always being obligatory—including in passives and on inanimates—in contrast with the often optional accusative case. As the ditransitive examples in (7) show, dative *-ko* is obligatory on the indirect object (as well as on experiencers (5) and in causativised ingestives (6b)), but accusative *-ko* is optional on many direct objects (7b).<sup>2</sup>

<sup>1</sup>The term ‘ingestives’ is due to how verbs of this class semantically denote consumption in some form—whether literal or abstract—across many Indic languages (Masica 1976). Causativised ingestives are unique in that the causee argument is marked dative. Some other verbs that belong to the ingestive class include *pii* ‘drink’, *dekh* ‘see’, *sikh* ‘learn’ etc. See Masica (1976) and Bhatt & Embick (2017) for more on ingestives.

<sup>2</sup>The form of the nominal that accusative *-ko* attaches to in (7b) and further examples is slightly simplified for ease of reading. The direct object in (7b) can be realised as either case-unmarked *chuuhaa* or accusative

- (7) a. *Obligatory dative -ko, no accusative -ko*  
 miina=ne **billii\*(=ko)** chuuhaa di-yaa  
 Mina=ERG cat\*(=DAT) mouse give-PFV  
 ‘Mina gave the cat a mouse’
- b. *Obligatory dative -ko, optional accusative -ko*  
 miina=ne chuuhe(=ko)<sub>j</sub> **billii\*(=ko)** \_\_\_<sub>j</sub> di-yaa  
 Mina=ERG mouse(=ACC)<sub>j</sub> cat\*(=DAT) \_\_\_<sub>j</sub> give-PFV  
 ‘Mina gave the cat a/the mouse’

Unlike with accusative case—which may or may not be obligatory depending on the specificity and animacy (among other factors) of the object (8)—a DP’s semantic properties have no influence on the obligatoriness of dative case. While (7a) already showed the obligatoriness of dative case on a goal DP whose referent is animate, (9) shows the obligatoriness of dative case even on an inanimate goal.

- (8) a. *Obligatorily accusative direct object*  
 miina=ne **tiina\*(=ko)** dekh-aa  
 Mina=ERG Tina\*(=ACC) see-PFV  
 ‘Mina saw Tina’
- b. *Optionally accusative direct object*  
 miina=ne fuul(=ko) dekh-aa  
 Mina=ERG flower(=ACC) see-PFV  
 ‘Mina saw a(/the) flower’
- (9) *Obligatory dative -ko on inanimate objects*  
 miina=ne fuul\*(=ko) paanii di-yaa  
 Mina=ERG flower\*(=DAT) water give-PFV  
 ‘Mina watered the flower’ (Lit: ‘Mina gave water to the flower’)

In passives of transitives,<sup>3</sup> accusative *-ko* becomes optional on a direct object (10a), even in cases where it was obligatory in active voice (c.f. (8a)).<sup>4</sup> When an indirect object is passivized, dative *-ko* remains obligatory on the indirect object, (10b).

*chuuhe=ko*. The variability in the final vowel of the nominal stem has no bearing on the argument presented.

<sup>3</sup>Note that while the experiencer construction discussed in §2 is also transitive, it is not discussed here because it cannot be passivized/made into an active impersonal construction in a similar way as the constructions in (10).

<sup>4</sup>The change in agreement on the verb and auxiliary when *tiinaa* is case-unmarked vs. accusative only has to do with case in that the  $\phi$ -probe in Hindi-Urdu is *case-discriminating* (Bobaljik 2008; Preminger 2014; Agarwal 2022). I will not delve into the  $\phi$ -agreement facts in Hindi-Urdu any further in this paper, since they are independent of dative case assignment.

- (10) a. *Passivised DO, optional accusative -ko*  
 ṭinaa=ko dekh-aa ga-yaa / ṭinaa dekh-ii ga-yii  
 Tina=ACC see-PFV PASS-PFV / Tina see-PFV PASS-PFV  
 ‘Tina was seen’
- b. *Passivised IO, obligatory dative -ko*  
 fuul\*(=ko) paanii di-yaa ga-yaa  
 flower\*(=DAT) water give-PFV PASS-PFV  
 ‘The flower was watered’ (Lit: ‘The flower was given water’)

It is clear from (7-10) that although dative and accusative case in HU are both realised as *-ko*, they are structurally different cases, and I will treat them as such throughout the paper.

## 2.2 Identifying *-ko* in causativised ingestives

Since the evidence for dependent dative case in Hindi-Urdu will ultimately come from causativised ingestives like (11b), I will briefly demonstrate that the *-ko* on the causee in these constructions is indeed dative, and not accusative. The causativised ingestive construction in (6b) is repeated below as (11).

- (11) *Causativised ingestive*  
 salma=ne kutte\*(=**ko**) seb khil-aa-yaa  
 Salma=ERG dog\*(=DAT) apple eat-CAUS-PFV  
 ‘Salma fed the dog an apple’ (Lit: ‘Salma made the dog eat an apple’)

Bhatt & Embick (2017) have argued that the causee *kutte* ‘dog’ is obligatorily dative—and not accusative—in this construction. I will briefly outline some of their arguments for *-ko* on *kutte* ‘dog’ being dative.

We saw in (8b) that the presence of accusative case on an object has to do with its specificity and animacy, while dative case in (9) showed no such optionality. (12) shows that the case on the causee in ingestives is indeed dative, because even a nonspecific inanimate causee is obligatorily *-ko*-marked.<sup>5</sup>

- (12) *Causativised ingestive with nonspecific inanimate causee*  
 salma=ne koi ek guriyaa\*(=**ko**) saariii pehen-aa-yii  
 Salma=ERG some one doll\*(=DAT) sari wear-CAUS-PFV  
 ‘Salma dressed one of the dolls in a sari’ (Lit: ‘Salma made one of the dolls wear a sari’)

More evidence in favour of *-ko* on the causee in causativised ingestives being dative comes from passivising the ingestive causative. Recall from (10a) that accusative *-ko* becomes optional in passives on all arguments, regardless of their specificity and animacy, while dative *-ko* remains obligatory on all arguments in passives (10b). Passivising (11) and

<sup>5</sup>A specific indefinite reading can be resisted in (12) by felicitously following up the statement in the example with ‘but I do not know which doll’.



(12)—as in (13)—requires the obligatory retention of *-ko* on the causee in both instances, again indicating that this *-ko* is in fact dative.

(13) a. *Passive of (11)*

kutte\*(=**ko**) seb khil-aa-yaa ga-yaa  
dog\*(=**DAT**) apple eat-CAUS-PFV PASS-PFV  
'The dog was fed an apple' (Lit: 'The dog was made to eat an apple')

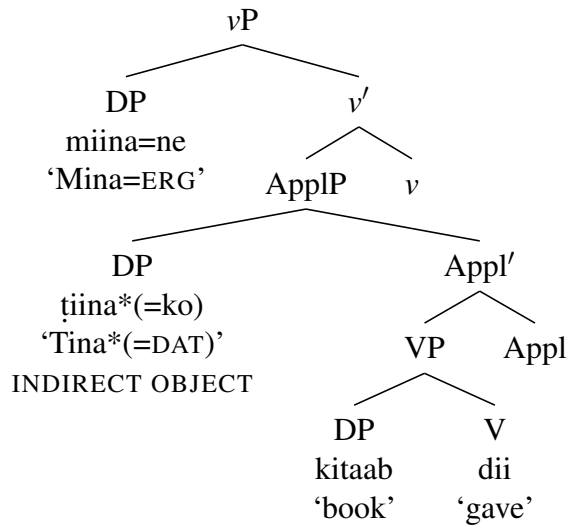
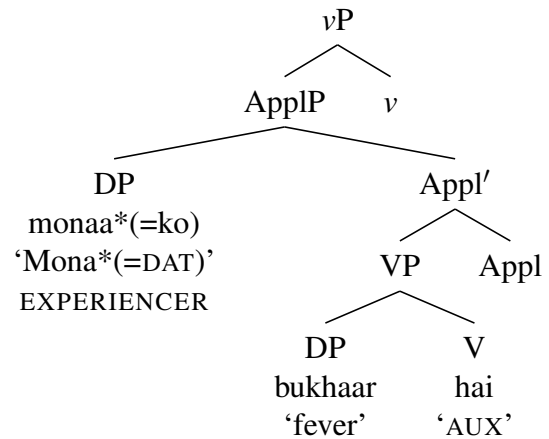
b. *Passive of (12)*

koi ek guṛiyaa\*(=**ko**) saari pehen-aa-yii ga-yii  
some one doll\*(=**DAT**) sari wear-CAUS-PFV PASS-PFV  
'One of the dolls was dressed in a sari' (Lit: 'One of the dolls was made to wear a sari')

(12) and (13) are clear arguments for dative *-ko* on causees in causativised ingestives. I now turn to possible analyses of dative case assignment in Hindi-Urdu, keeping in mind the instances of dative case discussed so far.

### 3 Theories of dative case assignment in Hindi-Urdu

Before turning to theories of dative case assignment, I will briefly introduce the assumptions I make about the location of dative DPs in the structure. Larson (1988) and Pykkänen (2008) have argued that indirect objects are introduced in the specifier of an applicative head. Based on their proposals, I adopt the structure in (14) for ditransitives, as illustrated for (2). Davison (2004) has proposed that experiencer arguments are merged in the same position as goals—Spec,VP in her terms and Spec,AppIP when transposed to this account. In her view, the lower DP in the experiencer construction is in Spec,VP much like the direct object in (di)transitives, so the AppIP structure in (14) is applicable to experiencer constructions as well, (15).

(14) *Ditransitive structure for (2)*(15) *Experiencer structure for (5)*

Once we adopt these fairly common assumptions, the ditransitive and experiencer examples so far are compatible with two analyses of dative case assignment in Hindi-Urdu. First, a functional head case analysis, where Appl assigns dative case to its specifier, akin to what Kalin (2014) has proposed. Ditransitives and experiencers are also compatible with a dependent dative case analysis, where dative is assigned to the higher of two DPs in the complement of  $v$ , as Baker & Vinokurova (2010) have argued for dative case in Sakha.

I argue that all instances of dative case in Hindi-Urdu are dependent case. The dependent case analysis neatly captures the generalisation that a dative-marked DP is always in the environment of a lower DP within a  $vP$  domain, as exemplified by every instance of dative case in this paper so far. Crucial evidence for dependent dative case in Hindi-Urdu comes from causativised ingestives like (6b), where the causee argument is exceptionally marked dative. I will show in §3.1 that dative case cannot be a functional head case in this construction, and thus must be a dependent case, detailed in §3.2.

### 3.1 Dative is not a functional head case

I now turn to the structure of the causativised ingestive, and show that dative case on causees in this construction cannot be a functional head case. I will refer to the causativised ingestive as just 'causatives' going forth, both for the sake of simplicity, and to reflect how the causativised ingestive mirrors causative constructions in other languages.

Recall once again the simple ingestive construction in (6a), repeated in (16a), There is a dedicated agent of eating, *kutte* 'dog', which is marked ergative like other external arguments in transitive perfective clauses. The structure of transitive ingestives is then parallel to the structure of other (non-experiencer) standard transitives, with the external argument in Spec, $vP$ . When the ingestive is causativised as in (16b), *kutte* 'dog' is now dative instead (and the causer is ergative). The puzzle, then, is the source of dative case in the causative.

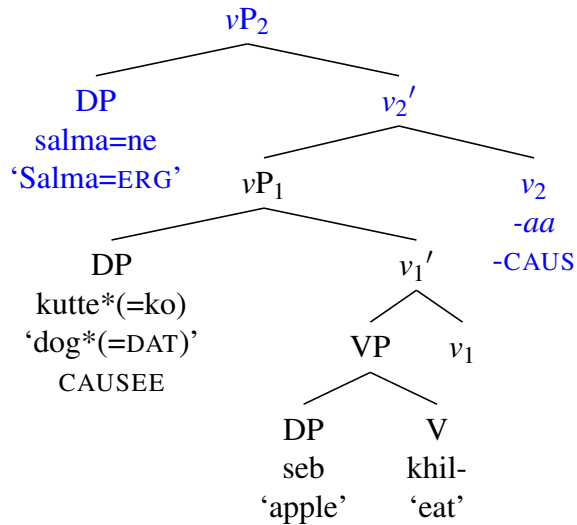
- (16) a. *Ingestive*  
 kutte=ne seb khaa-yaa  
 dog=ERG apple eat-PFV  
 ‘The dog ate an apple’
- b. *Causativised ingestive*  
 salma=ne kutte\*(=ko) seb khil-aa-yaa  
 Salma=ERG dog\*(=DAT) apple eat-CAUS-PFV  
 ‘Salma fed the dog an apple’ (Lit: ‘Salma made the dog eat an apple’)

In Japanese causatives (17), the same pattern is obtained as in (16), where the non-dative external argument in the simple transitive becomes dative once embedded under a causative shell (Harley 2008). This pattern of dative case on the causee in a causativised transitive clause is also replicated in Sakha (Baker & Vinokurova 2010). Akin to *-aa* in the verbal complex in (16b), both Japanese and Sakha also have a dedicated causative morpheme in the verbal complex in the causative construction.

- (17) a. *Simple transitive in Japanese*  
 hanako=ga pizza=o tabe-ta  
 Hanako=NOM pizza=ACC eat-PST  
 ‘Hanako ate pizza’ (p.c. Katsuya Wakabayashi)
- b. *Causative*  
 taro=ga hanako\*(=ni) pizza=o tabe-sase-ta  
 Taro=NOM Hanako\*(=DAT) pizza=ACC eat-CAUS-PST  
 ‘Taro made Hanako eat pizza/fed Hanako pizza’ (Harley 2008; p.c. Katsuya Wakabayashi)

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

(18) *Causative structure*



As far as functional head case theory goes, there are two contenders for dative case assigners in the causative: First,  $v_1$ , which introduces *kutte* ‘dog’ in both the ingestive in (16a) and its causative in (16b). Second, the causative head  $v_2$ , which embeds  $vP_1$  and introduces the causer. I will now argue that neither of these heads assign dative case in the causative in Hindi-Urdu.

The argument against  $v_1$  assigning dative case in the causative is simple— $v_1$  does not assign dative case to *kutte* ‘dog’ in the simple ingestive in (16a), so it cannot assign dative case to *kutte* ‘dog’ in the causative in (16b). As seen in (19), the simple ingestive is ungrammatical with a dative subject in place of an ergative subject, so  $v_1$  is not a dative-assigning head in the ingestive. Then, under functional head case theory, the only remaining dative case assigner is  $v_2$ , the causative head.

(19) *Ingestive with dative subject*

\**kutte=ko seb khaa-yaa*  
 dog=DAT apple eat-PFV  
 Intended: ‘The dog ate an apple’

To argue against  $v_2$  assigning dative case, I show that this causative head does not assign dative case in other configurations where it is found. Take for example the unergative in (20a), and its causativised counterpart in (20b).

- (20) a. *Unergative*  
 kutta daur̥ rahaa hai  
 dog run PROG AUX.PRES  
 ‘The dog is running (around)’
- b. *Causativised unergative*  
 salma kutte(=ko) daur̥-aa rahii hai  
 Salma dog(=ACC) run-CAUS PROG AUX.PRES  
 ‘Salma is making a/the dog run’

Notably, the causee in the causative in (20b) is not dative. It is either case-unmarked or accusative, but not dative, as evidenced by the optionality of *-ko*. Recall from (8)-(10) that any instance of optional *-ko* is accusative, since dative *-ko* is always obligatory, even on nonspecific inanimate DPs and in passives. Passivising (20b) corroborates that *-ko* on *kutta* ‘dog’ is accusative, and not dative.

- (21) *Passive of causativised unergative*  
 kutta / kutte=ko daur̥-aa-yaa jaa rahaa hai  
 dog / dog=ACC run-CAUS-PFV PASS PROG AUX  
 ‘The dog is being made to run.’

Since the causee in causativised unergatives is accusative, not dative, the causative head *-aa*—represented as  $v_2$  in (18)—also cannot assign dative case. Then, no head is available to assign dative case to the causee in causatives of transitive ingestives like (16b) or (12), and dative case in Hindi-Urdu cannot be a functional head case. I will now move to on showing that every instance of dative case in this paper can in fact be modelled as a dependent case.

### 3.2 Dative case is a dependent case

While functional head case theory can account for the distribution of dative case in ditransitives and experiencers, it cannot in causatives like (16b) or (12), as shown in §3.1. Still, the uniform behaviour of dative case in ditransitives, experiencers, and causatives with regard to non-optionality and retention under passivisation suggests that a uniform analysis of dative case assignment in all of the aforementioned constructions is warranted. As claimed earlier in §3, dative case is only ever found on a DP when there is a lower DP in the same domain. Consider again the ditransitive, experiencer, and causative examples in (22), with the lower DP underlined.

- (22) a. *Ditransitive*  
 miina=ne tiina\*(=ko) kitaab di-i  
 Mina=ERG Tina\*(=DAT) book give-PFV  
 ‘Mina gave Tina a/the book’
- b. *Experiencer construction*  
 mona\*(=ko) bukhaar hai  
 Mona\*(=DAT) fever AUX.PRES  
 ‘Mona has a fever’
- c. *Causativised ingestive*  
 salma=ne kutte\*(=ko) seb khil-aa-yaa  
 Salma=ERG dog\*(=DAT) apple eat-CAUS-PFV  
 ‘Salma fed the dog an apple’

The fact that dative case only appears in environments where there is a proximal lower DP is the clearest evidence in favour of a dependent case analysis of dative case in Hindi-Urdu. When a lower DP is absent, like in the causativised unergative in (20b), dative case is absent too. However, a lower proximal DP is only necessary, not sufficient to condition dative case on a DP, as demonstrated by simple transitive clauses like (3) and (6a), repeated as (23), where the higher argument is non-dative.

- (23) a. *No dative on higher DP*  
 miina{=ne/\*=ko} fuul(=ko) dekh-aa  
 Mina{=ERG/\*=DAT} flower(=ACC) see-PFV  
 ‘Mina saw a(/the) flower’
- b. *No dative on higher DP*  
 kutte{=ne/\*=ko} seb khaa-yaa  
 dog{=ERG/\*=DAT} apple eat-PFV  
 ‘The dog ate an apple’

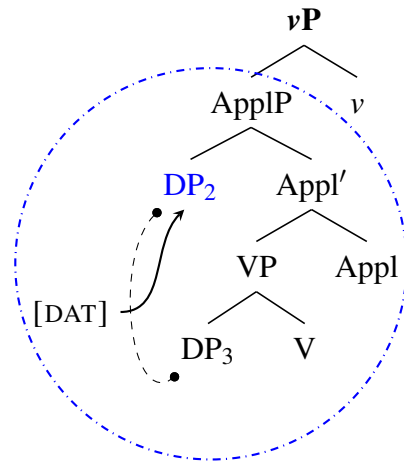
I claim that the missing ingredient for a complete dependent case analysis of the Hindi-Urdu dative is higher *v*P structure. As suggested in the trees for (22) in (14) and (18), the dative DP is always dominated by *v*P. Then, just like Baker & Vinokurova (2010) have proposed for Sakha, dative case comes out to be the case that appears on the higher of two DPs in the complement of *v* in Hindi-Urdu as well. A formal rule for dependent dative case assignment in Hindi-Urdu is proposed in (24).

- (24) **DATIVE CASE RULE:** If DP<sub>1</sub> c-commands DP<sub>2</sub> in the complement of *v*, assign dative to DP<sub>1</sub>

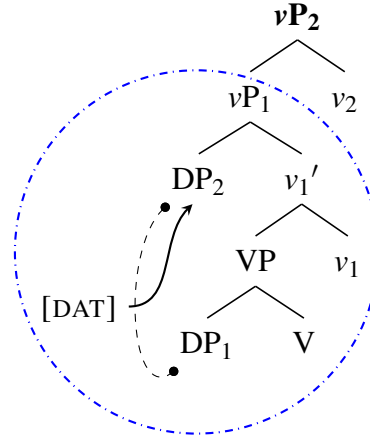
(24) derives the distribution of dative case in all of the prototypical examples in (22). The dative case rule in (24) is schematised for ditransitives/experiencers in (24) and for causatives in (24).<sup>6</sup>

<sup>6</sup>Recall from §3 that experiencer arguments and goals in ditransitives both merge in Spec,AppIP.

(25) a. *Dative case assignment in ditransitives/experiencers*

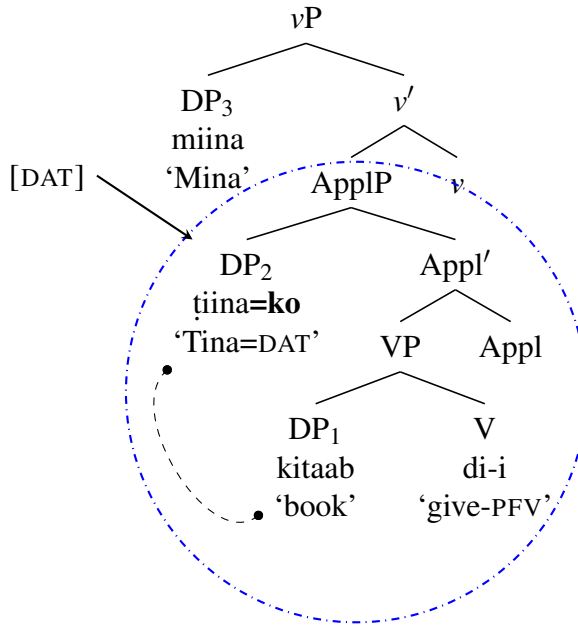


b. *Dative case assignment in causatives*

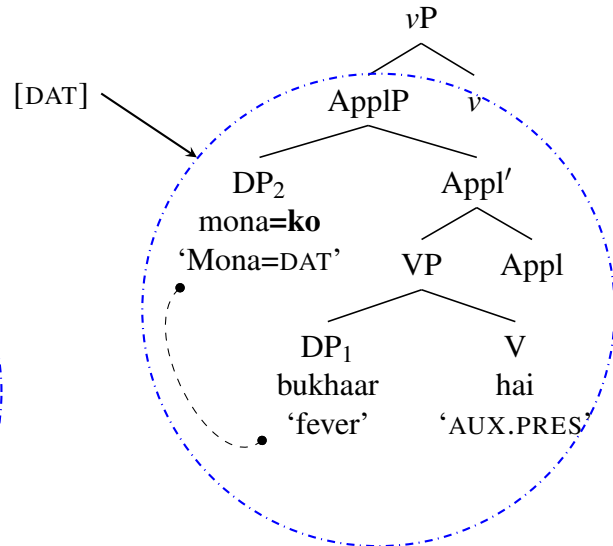


For the sake of clarity, I will briefly spell out the process of dative case assignment in each of the examples in (22). The ditransitive example in (22a) is illustrated as (26). The experiencer example in (22b) is illustrated as (15).

(26) *Dative case assignment in (22a)*



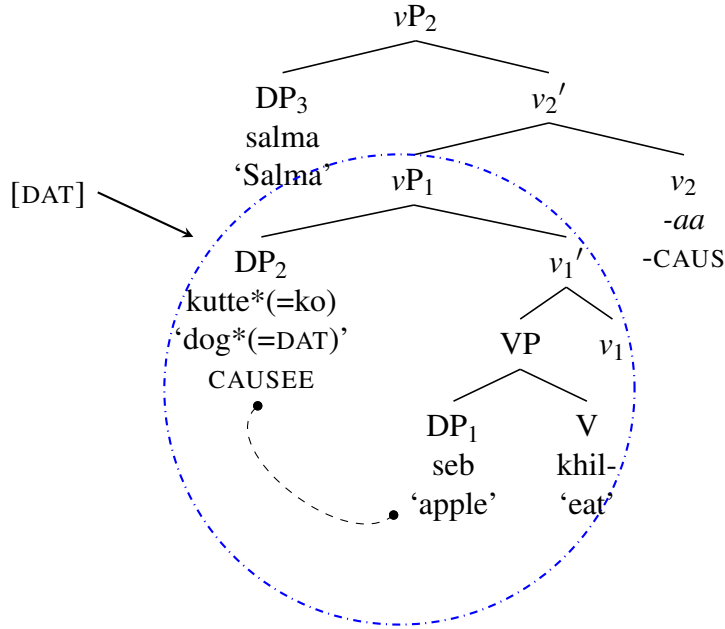
(27) *Dative case assignment in (22b)*



Lastly, the causative example in (22c) is sketched out in (28).<sup>7</sup>

<sup>7</sup>An astute reader might wonder if it is possible to have a causativised ditransitive verb with two adjacent dative arguments—the causee and the goal—due to the rule in (24) applying twice. Unfortunately, I have not found any ditransitive verbs that can be causativised to create such a configuration.

(28) *Dative case assignment in (22c)*



While functional head case theory could only derive the distribution of dative case in ditransitives and experiencers, the dependent dative case rule in (24) singlehandedly derives the distribution of dative case in ditransitives, experiencers, as well as causatives.

(24) also importantly derives the obligatory preservation of dative case in passives like (10b) and (13), since both conditioners of dative case—namely a lower DP and higher  $v$ P structure—are retained under passivisation. For example, in (10b), repeated as (29), *fuul* ‘flower’ receives dative case due to its lower case competitor *paanii* ‘water’, and due to both the theme and passivised goal being embedded under  $v$ , even in the absence of the external argument.

(29) *Passivised goal in ditransitive*

*fuul\**(=ko)    *paanii di-yaa ga-yaa*  
 flower\*(=DAT) water give-PFV PASS-PFV  
 ‘The flower was watered’ (Lit: ‘The flower was given water’)

The dative case facts in causativised ingestives also extend to another kind of causative construction, the indirect causative (30), where the causative suffix is *-vaa* (Bhatt & Embick 2017). Assuming the  $v$ P housing the causee ‘dog’ is dominated by another  $v$ P in (30) just like in the direct causative, the rule in (24) also derives the distribution of dative case in indirect causatives. *kutte* ‘dog’ is dative in (30a) due to the lower DP *seb* ‘apple’, while the same DP in (30b) is case-unmarked or accusative, since the description of the dative rule is not met in the absence of a lower DP.



- (30) a. *Indirect transitive causative*  
 salma=ne kutte\*(=**ko**) seb khil-vaa-yaa  
 Salma=ERG dog\*(=**DAT**) apple eat-CAUS.INDR-PFV  
 ‘Salma had the dog eat an apple’ (someone other than Salma made the dog eat the apple)
- b. *Indirect intransitive causative*  
 salma=ne kutte\*(=**ko**) daur-vaa-yaa  
 Salma=ERG dog\*(=**ACC**) run-CAUS.INDR-PFV  
 ‘Salma had the dog run’ (someone other than Salma made the dog run)

Note that in configurations where dative case is found in Hindi-Urdu, the lower DP does not have to be case-unmarked. Even if accusative case assignment precedes dative case assignment in examples like (7), given again as (31), *chuuhe* ‘mouse’ still invariably conditions dative case on *billii* ‘cat’. Baker (2015) has proposed that in some languages, DPs that have already received case can nonetheless trigger dependent case on another DP, and the direct object in (31) is a possible example.

- (31) *Accusative-marked case competitor for dative case*  
 miina=ne chuuhe(=**ko**)<sub>j</sub> **billii**\*(=**ko**) \_\_\_<sub>j</sub> di-yaa  
 Mina=ERG mouse\*(=**ACC**)<sub>j</sub> cat\*(=**DAT**) \_\_\_<sub>j</sub> give-PFV  
 ‘Mina gave the cat a/the mouse’

Further evidence that overtly case-marked DPs in Hindi-Urdu may still trigger dependent dative case on a higher DP in *vP* is given in (32), where the dative DP *kutte* ‘dog’ has an instrumental case competitor.

- (32) *Instrumental-marked case competitor for dative case*  
 salma=ne kutte\*(=**ko**) billii-se mil-aa-yaa  
 Salma=ERG dog\*(=**DAT**) cat-INST meet-CAUS-PFV  
 ‘Salma introduced the dog to a cat’ (Lit: ‘Salma made the dog meet a cat’)

To recapitulate, the dependent dative case rule in (24) derives the distribution of dative case in ditransitives, experiencers, direct and indirect causatives, and passives. Then, only dependent case theory offers an explanatorily adequate account of all dative case in Hindi-Urdu, since functional head theory could not explain the source of dative case in causatives.

#### 4 Conclusion

I have shown in this paper that dative case in Hindi-Urdu is a dependent case. Dative case—which is found on goals in ditransitives, on experiencers, and on causees in causatives—demonstrates similar morphosyntactic behaviour in all of these constructions, in that it is retained under passivisation and is never optional. Yet, dative case in Hindi-Urdu resists characterisation as a functional head case due to the lack of an assigning head in the causative. Analysing the dative as a dependent case in Hindi-Urdu instead derives

its appearance in ditransitives, experiencers, causatives, and passives (of ditransitives and causatives) alike. Dative case in Hindi-Urdu being a dependent case solidifies Baker & Vinokurova (2010); Baker (2015)'s addition of dependent dative case to the typology of dative cases—where dative may be characterised as an inherent or structural head case (Zaenen et al. 1985; Chomsky 1986; Woolford 2006). The dependent case analysis of the Hindi-Urdu dative also lends support to the view that some dependent case rules allow a case competitor that has already been valued for case.

## Acknowledgements

Many thanks to Stefan Keine, Ethan Poole, Anoop Mahajan, Rajesh Bhatt, Michelle Yuan, and the audiences at UCLA and FASAL-13 for helpful discussion and comments.

## References

- Agarwal, Hashmita. 2022. *Phases are Read-Only*: University of California, Los Angeles MA thesis.
- Alexiadou, Artemis, Elena Anagnostopoulou & Christina Sevdali. 2010. Patterns of dative-nominative alternations. *Proceedings of the North Eastern Linguistics Society (NELS)* 41.
- Alexiadou, Artemis, Elena Anagnostopoulou & Christina Sevdali. 2014. Opaque and transparent datives, and how they behave in passives. *The Journal of Comparative Germanic Linguistics* 17. 1–34.
- Anagnostopoulou, Elena. 2003. *The syntax of ditransitives: Evidence from clitics*. Walter de Gruyter.
- Baker, Mark. 2015. *Case*. Cambridge University Press.
- Baker, Mark C & Nadya Vinokurova. 2010. Two modalities of case assignment: Case in Sakha. *Natural Language & Linguistic Theory* 28. 593–642.
- 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.
- Bobaljik, Jonathan David. 2008. Missing persons: A case study in morphological universals. *The Linguistic Review* (25(1-2)). 203–230.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.
- 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. 137–164.

- Fukuda, Shin. 2007. Object case and event type: Accusative-dative object case alternation in Japanese. In *Annual meeting of the Berkeley Linguistics Society*, vol. 33 1, 165–176.
- Harley, Heidi. 2008. On the causative construction. In Shigeru Miyagawa & Mamoru Saito (eds.), *The handbook of Japanese linguistics*, 20–53. Oxford University Press.
- Ishizuka, Tomoko. 2010. *Toward a unified analysis of passives in Japanese: A cartographic minimalist approach*. University of California, Los Angeles.
- Kalin, Laura. 2014. *Aspect and argument licensing in Neo-Aramaic*: University of California, Los Angeles dissertation.
- Larson, Richard K. 1988. Scope and comparatives. *Linguistics and Philosophy* 11. 1–26.
- Marantz, Alec. 1991. Case and licensing. In *Proceedings of the Eighth Eastern States Conference on Linguistics (ESCOL'91)*, 234–253. Ohio State University.
- Masica, Colin P. 1976. *Defining a linguistic area: South Asia*. Orient Blackswan.
- Mohanan, Tara. 1994. *Argument structure in Hindi*. Center for the Study of Language (CSLI).
- Preminger, Omer. 2014. *Agreement and its failures*. MIT press.
- Pylkkänen, Liina. 2008. *Introducing arguments*. MIT Press.
- Woolford, Ellen. 2001. Case patterns. In G. Legendre, J. Grimshaw & S. Vikner (eds.), *Optimality theoretic syntax*, 509–543. MIT Press.
- Woolford, Ellen. 2006. Lexical case, inherent case, and argument structure. *Linguistic Inquiry* 37. 111–130.
- Zaenen, Annie, Joan Maling & Höskuldur Thráinsson. 1985. Case and grammatical functions: The Icelandic passive. *Natural Language & Linguistic Theory* 3. 441–483.

# Egocentric questions: The view from Bangla and Hindi-Urdu

ARKA BANERJEE, *WBNUJS, Kolkata*

KOUSANI BANERJEE, *EFLU, Hyderabad*

RAJESH BHATT, *UMass Amherst*

## ABSTRACT

In Bangla and Hindi-Urdu, we find a kind of question that is grammatically restricted to being about a particular event. This kind of a question consists of a demonstrative pronoun followed by a plain question. We refer to such questions as egocentric questions and to the demonstrative pronoun they contain as the egocentric pronoun. The egocentric pronoun picks out an event and the question is about this event. Since the speaker and the hearer need to pick out the event the question is about, such questions cannot be used in a state of speaker ignorance. This differentiates them from plain questions where speaker ignorance is the default. We show that various properties of egocentric questions follow from the need to be able to assign a reference to the egocentric pronoun and from the nature of access the speaker has to the event that the egocentric pronoun picks out.

## 1 Introduction

This paper will address a phenomenon in Bangla (Bengali) and Hindi-Urdu where demonstratives can appear together with a question, however in a non-canonical manner. Here, the term ‘non-canonical’ is used to refer to the fact that these demonstratives are not assigned any theta role. Consider the following:

### (1) Bangla

- a. *e tumi ki korcho?*  
this you what do.PROG.PRS.2  
‘What is this that you are doing?’
- b. *e ami kothay elam?*  
this I where come.PST.1  
‘What is the place where I came?’

### (2) Hindi-Urdu

- a. *yeh tum kya: kar rahe ho?*  
this you what do PROG.MPL be.PRS.2PL  
‘What is this that you are doing?’
- b. *yeh ham kahã: a: gaye hẽ?*  
this we where come GO.MPL be.PRS.3PL  
‘What is this place where we have come?’

In the above examples, it is noteworthy that no theta role is assigned to the demonstrative *elyeh* ‘this’. The issue of this point becomes clearer when we look at the use of *elyeh* in the following examples:

(3) Bangla

- a. *e* (*lok*) *kothay thake?*

this person where stay.PRS.3

‘Where does this person stay?’

Context: the speaker has met Riya’s friend Ravi, and the speaker is asking Riya where Ravi stays (pointing towards Ravi).

- b. *e* (*meye*) (*ki*) *bhalo?*

this girl PQP good

‘Is this (girl) good?’

Context: Riya is showing a picture of her school group to the speaker, and (s)he points towards a girl in that picture asking if that girl is good.

(4) Hindi-Urdu

- a. *yeh* (*lar̥ki:*) *kahā: rah-ti: hε?*

this girl.F where stay-IMPFV.F be.PRS.3SG

‘Where does this girl live?’

Context: Riya is showing a picture of her school group to the speaker, and the speaker points towards a girl in the picture and asks the above question.

- b. (*kya:*) *yeh* (*lar̥ki:*) *mehnatī: hε?*

PQP this girl hardworking be.PRS.3SG

‘Is this (girl) hardworking?’

Context: Riya is showing a picture of her school group to the speaker, and (s)he points towards a girl in that picture asking if that girl is hardworking.

For the last two pairs of examples, from (3a) to (4b), it is clear that the demonstrative *elyeh* either gets a theta role (when it appears as a pronoun, without a nominal complement) or is part of a nominal that gets a theta role (when it appears as a determiner). The distinction between the canonical usage of the demonstrative *elyeh* and its non-canonical usage can be brought out further by making the demonstrative take a nominal complement as in (5a-6b).

(5) Bangla

- a. *\*e kaj tumi ki korcho?*

this work you what do.PROG.PRS.2

Intended: ‘What is this work that you are doing?’

- b. *\*e jaygay ami kothay elam?*

this place I where come.PST.1

Intended: ‘What is this place where I came?’

(6) Hindi-Urdu

- a. \**yeh ka:m tum kya: kar rahe ho?*  
this work you.PL what do PROG.MPL be.PRS.2PL  
Intended: ‘What is this work that you are doing?’
- b. \**yeh jagah ham kahā: a: gaye hē?*  
this place we where come GO.PFV.MPL be.PRS.3PL  
Intended: ‘What is this place where we have come?’

In these cases the canonical interpretation is not available as there are too many DPs and too few theta roles. For example in (6a), the main verb assigns a theta role to its object *kya:* ‘what’, leaving the DP *yeh ka:m* ‘this work’ without a theta role. The same is the case with Bangla examples in (5a)-(5b). This is unsurprising. What is more surprising is that the non-canonical interpretation is also unavailable. We learn therefore that in the non-canonical usage, the demonstrative needs to function as a pronoun i.e. without a nominal complement.

In this paper, we will attempt to identify what role this non-argumental, demonstrative pronoun *e/yeh* plays in examples like (1a-2b), its semantic contribution(s), and the syntactic restrictions on the demonstrative in the non-canonical reading. Hereafter, throughout this paper, we will refer to these non-canonical questions as *egocentric* questions and we will refer to the non-argumental, demonstrative pronoun *e/yeh* as an *egocentric pronoun*.

## 2 Clause-type sensitivity of the egocentric pronoun

In this section, we explore what kinds of clauses the egocentric pronoun can appear in. As noted earlier in (1) and (2), the egocentric pronoun can appear with constituent questions. There seems to be no restriction concerning what the questioned constituent is. We have already seen examples where the questioned constituent is *ki/kya:* ‘what’ (1a, 2a), *kothay/kahā:* ‘where’ (1b, 2b). It also appears with *ke/kaun* ‘who’, *kibhabe/kaise* ‘how’, *kokhon/kab* ‘when’, *kanol/kyō* ‘why’, and *koto/kitna:* ‘how much’. Consider the following:

(7) Bangla

- |   |   |
|---|---|
| a. <i>e Ram kokhon elo?</i><br>this Ram when come.PST.3<br>‘When is it that Ram came?’                | c. <i>e tui eta kano korli?</i><br>this you it why do.PST.2<br>‘Why is it that you did it?’                       |
| b. <i>e ke esheche bari-te?</i><br>this who come.PRF.PRS.3<br>home.LOC<br>‘Who is it that came home?’ | d. <i>e tui koto khacchish?</i><br>this you how-much eat.PROG.PRS.2<br>‘How much food is it that you are eating?’ |

(8) Hindi

- a. *yeh Ram kab a:-ya:?*  
this Ram when come-PFV.MSG  
'When is it that Ram came?'
- b. *yeh tum aisa: ka:m kyō: kar*  
this you such thing why do  
*rahe ho?*  
PROG.MPL be.PRS.2PL  
'Why is it that you are doing such a thing?'
- c. *yeh kaun a: raha:*  
this who come PROG.MSG  
*hε?*  
be.PRS.3SG  
'Who is it that is coming?'
- d. *yeh tum kitna: kha:na: kha:*  
this you how-much food eat  
*rahe ho?*  
PROG.MPL be.PRS.2PL  
'How much food you are eating!'

The egocentric pronoun is also compatible with questions with reduplicated *wh*-words (9a,10a) as well as with multiple *wh* questions (9b, 10b).

(9) Bangla

- a. *e tui kake kake merechish?*  
this you whom whom hit.PROG.PRS.2  
'Who are the people that you have hit?'
- b. *e tui kake ki bol-e boshli?*  
this you whom what say-CNV sit.PST.2  
'Who is x and what is y s.t. you said y to x?'

(10) Hindi-Urdu

- a. *yeh tum=ne kis kis=ko pi:t diya:?*  
this you=ERG whom whom=DOM beat give.PFV.DEF  
'Who are all these people that you have hit?/What a wide range of people you have hit'
- b. *yeh tum kis=se kya: kah baiṭh-e?*  
this you who=INS what say SIT-PFV.MPL  
'Who is x and what is y s.t. you ended up saying y to x?'

Apart from the above constructions, the egocentric pronoun also appears in polar (11a, 12a) and alternative questions (11b, 12b).

(11) Bangla

- a. *e tui (ki) Rishi-ke boi-ta dili?*  
this you PQP Rishi-DAT book-CL give.PST.2  
'Is it Rishi that you gave the book to?'
- b. *e tui mach kheli na mangsho?*  
this you fish ate NA meat  
'What is it between fish and meat you ate?'

(12) Hindi-Urdu

- a. *yeh tum=ne (kya:) Ravi-ko kita:b di: thi:?*  
this you=ERG QP Ravi-DAT book give.PFV.F be.PST.F.SG  
'Is it Ravi that you gave the book to?'
- b. *yeh tum papi:ta: kha: rahe ho ki aam?*  
this you papaya eat PROG.MSG be.PRS.2 THAT mango  
'Are you eating papaya or mango?'

The above survey of questions that the egocentric pronoun can appear in allows us to rule out analyses which directly connect the proximal demonstrative with the *wh*-word and the resulting question gives us the identity of the object that the demonstratives point to. Here is what such an analysis might look like:

(13) **it, who** *does Joan like*  
'Who is it that Joan likes?'

The exact details are not important because we can rule out such analyses on two grounds. The demonstrative *e/yeh* is not used for pointing to degrees/amounts - for that *eto/itna:* would be used, see (7d/8d). Furthermore in polar questions and alternative questions, there is no *wh*-phrase that could be associated with the proximal demonstrative and yet as we have seen in (11) and (12), the egocentric pronoun is possible in such questions.

## 2.1 Compatibility with rhetorical questions

The egocentric questions we have seen so far have been information seeking questions. But egocentric questions need not always be information seeking. They can also be used in rhetorical contexts as well (Caponigro & Sprouse, 2007; Biezma & Rawlins, 2017). Consider the following examples where the egocentric question is used to convey a reproach.

- (14) Context: I am watching you talk to Mahesh.
- a. Bangla  
*e tui kibhabe kotha bolchish Mahesh-er shathe?*  
this you how.manner talk.PROG.PRS.2 Mahesh-GEN with  
'How are you talking to Mahesh?'
- b. Hindi-Urdu  
*yeh tum Mahesh=se kaise ba:t kar rahe ho?*  
this you Mahesh=INS how talk do PROG.MPL be.PRS.2  
'How are you talking to Mahesh?'

This question does not ask for information; instead, the speaker uses it to convey to the hearer that the speaker thinks the manner in which the hearer is speaking to Mahesh is inappropriate.



(15) Context: You and I can see and hear Mina talking to someone. We know who she is talking to.

a. Bangla

*e Mina kar shathe kotha bolche?*  
this Mina who with talk.PROG.PRS.3  
'Who is it that Mina is talking to?'

b. Hindi-Urdu

*yeh Mina kis=se ba:t kar rahi: he?*  
this Mina who=INS talk do PROG.F be.PRS.3SG  
'Who is it that Mina is talking to?'

Here too, the question is not asking for a canonical answer as the speaker can see who Mina is talking to. Instead, it is used to convey that there is something inappropriate about Mina talking to this person. We wish to make two additional points here. The first is that these questions are not limited to being rhetorical. If we change the context, the questions become information seeking.

(16) Context: I am watching you talk to Mahesh using strange sounds and gestures.

a. Bangla

*e tui kibhabe kotha bolchish Mahesh-er shathe?*  
this you how.manner talk.PROG.PRS.2 Mahesh-GEN with  
'How are you talking to Mahesh?'

b. Hindi-Urdu

*yeh tum Mahesh=se kaise ba:t kar rahe ho?*  
this you Mahesh=INS how talk do PROG.MPL be.PRS.2  
'How are you talking to Mahesh?'

Response: Ah! It's this new code we've been trying out!

A similar adjustment of the context allows (15) to function as an information seeking question. The switch from a rhetorical use to an information seeking one is accompanied by a corresponding shift in the prosody.

The second point is more tricky. We have shown that egocentric questions can function both as rhetorical questions and information seeking questions. But this is also true of canonical questions! So what is special about egocentric questions in this regard? We contend that there is something about the form of egocentric questions that they lend themselves very easily to being used in rhetorical questions and exclamation, in comparison to normal questions. Once we present our semantic proposal for egocentric questions, we will return to this aspect of their behavior in §4.3.

We have now shown that egocentric *e/yeh* can occur with all types of questions. Next, we turn to whether it can appear in other clause types.

## 2.2 Only with questions

The tendency of egocentric questions toward being used as rhetorical questions raises the query, whether the egocentric pronoun can be used in declaratives as well. It turns out that it can only be used with questions, as already seen, and with *wh*-exclamatives (17a/17b). The occurrence of the egocentric pronoun with *wh*-exclamatives does not pose a challenge for its tendency to appear with questions because *wh*-exclamatives can be given a question-based analysis (cf. Banerjee, 2022, for Bangla).

(17) Context: I am surprised to see what a beautiful picture you have drawn.

a. Bangla

(*e*) *tui ki darun ekta chobi enkechish!*  
this you what marvellous one.CL painting draw.PRF.PRS.3  
'What a marvellous painting you have drawn!'

b. Hindi-Urdu

(*yeh*) *tum=ne kya: sundar tasvi:r bana:-yi: he!*  
this you=ERG what beautiful picture make-PFV.F be.PRS.3SG  
'What a beautiful picture you have made!'

However, the egocentric pronoun can occur neither in declaratives (18a/18b) nor in what Rett (2008) calls *propositional exclamations* (18c).

(18) a. Bangla declaratives:

\**e tui Mina-r sathe kotha bolchish*  
this you Mina-GEN with talk.PROG.PRS.2  
Lit.: '\*This you are talking to Mina'.

b. Hindi-Urdu declaratives:

\**yeh tum Mina=se ba:t kar rahe ho*  
this you Mina-WITH talk do PROG.MPL be.PRS.3SG  
Lit.: '\*This you are talking to Mina'.

c. Bangla propositional exclamations:

(*e*) *ajke darun ranna hoyeche!*  
this today marvelous cooking happen.PRF.PRS.3  
Lit.: '\*This today marvelous cooking happened.'

The egocentric pronoun is also not compatible with imperatives.

(19) a. Bangla

(*e*) *apnara shobai pichon-er dorja diye bero-ben.*  
this you all back-GEN door through exit-FUT.IMP.HON  
'\*This exit through the back door!'

- b. Hindi-Urdu  
 (\***yeh**) *pi:che=se ba:har ja:-o.*  
 this back-FROM outside go-IMP  
 ‘\*This go out through the back!’

The crux of the above discussion is that the egocentric pronoun is sensitive to clause types. It is possible only in question contexts or in those contexts that are derived from questions (for example, *wh*-exclamatives), and nowhere else.

### 3 A semantics for egocentric questions

Our initial intuition for building up the semantic contribution of these egocentric questions is that the egocentric pronoun (*e/yeh*) in an egocentric question is a pronominal element and its referent should be identified by both the speaker and the hearer successfully. The following section outlines an intuition about the semantics of the egocentric questions.

#### 3.1 Informal intuition

We propose that in egocentric questions, the egocentric pronoun (*e/yeh*) picks out an event. Let us call it QEvent. We assume that the contribution of the egocentric question is to restrict the question to QEvent. Unlike ordinary questions, which do not directly restrict the event variables of their answers, egocentric questions are questions about a particular event. The syntax of an egocentric question consists of the the egocentric pronoun in a high position in the left periphery followed by the CP that corresponds to the question.

(20) Egocentric Questions:

- a. Syntax:  
 $[_{TopP} \mathbf{e/yeh} [_{CP} \text{Question}]]$
- b. Semantics:  
 $[[\mathbf{e/yeh}_i \text{ Who is Mina talking to?}]] =$   
 $\{[\text{Mina is talking to } x \text{ in QEvent}]: x \text{ is a person}\}, \text{ where } g(i) = \text{QEvent}$

Ordinary questions, presumably, lack the topic layer and thus correspond to just the CP portion.

(21) Plain Question:

- a. Syntax:  
 a plain CP
- b. Semantics:  
 $[[\text{Who is Mina talking to?}]] =$   
 $\{\exists e [\text{Mina is talking to } x \text{ in } e]: x \text{ is a person}\}$

It is possible that in a language like English, ordinary questions might correspond to two structures – a plain question and an egocentric question with a silent pronoun in place of the egocentric pronoun.

One prediction of this implementation is that the event that *e/yeh* picks out has to match the question event. Suppose *e/yeh* picks out an event of Mina dancing, then this event cannot be an event in which Mina is talking to someone, assuming the events to be minimal.

### 3.2 The position of the egocentric pronoun in syntax

The syntax that we have proposed is inspired by Rizzi (1997). Rizzi notes that in matrix clauses in Italian, a topic must precede a *wh*-phrase.

(22) (Rizzi, 1997, page 289, ex. 13)

a. \*Wh-Top

\**A chi, il premio Nobel, lo daranno?*  
to whom the prize Nobel it give.Fut

Intended: ‘Who will they give the Nobel Prize to?’

b. Top-Wh

*Il premio Nobel, a chi lo daranno?*  
the prize Nobel to whom it give.Fut

‘Who will they give the Nobel Prize to?’

The same restriction applies to egocentric questions. A *wh*-phrase in an egocentric question may not precede the egocentric pronoun.

(23) a. Bangla

\**tumi ki e korcho?*  
you what this do.PROG.PRS.2

Intended: ‘What is this that you are doing?’

b. Hindi-Urdu

\**tum kya: yeh kar rahe ho?*  
you what this do PROG.MPL be.PRS.2

‘What is this that you are doing?’

There is, of course, a big difference between Italian on the one hand and Bangla and Hindi-Urdu on the other – Italian has overt and obligatory *wh*-movement to [Spec, CP], Bangla and Hindi-Urdu do not.<sup>1</sup> Given the absence of overt *wh*-movement to [Spec, CP] in Bangla and Hindi-Urdu, the facts in (23) could be seen as unsurprising. However, a deeper examination of the ordering restrictions that obtain in egocentric questions reveals that the kind of restrictions that Rizzi noted in Italian are relevant in Bangla and Hindi-Urdu as well.

<sup>1</sup>This point is not uncontroversial – see Simpson & Bhattacharya (2003) who argue that Bangla has obligatory and overt *wh*-movement to [Spec, CP] followed by movement of the following argumental material to a position above [Spec, CP]. If, as we argue, the egocentric pronoun is located higher than [Spec, CP], the ordering restriction can be made to follow from their proposal with the following restriction – non-*wh* material can move past the egocentric pronoun but the *wh*-phrase, trapped in [Spec, CP], cannot.

The default position for egocentric *e/yeh* is initial and this is the order we have seen so far. But in fact, egocentric *e/yeh* can appear in a non-sentence-initial position as long as it precedes the *wh*-phrase(s) of the clause in which it appears.<sup>2</sup> Consider the following:

- (24) a. Bangla  
 (✓*e*) *tumi* (✓*e*) *ki* (×*e*) *korcho* (×*e*)?  
 this you this what this do.PROG.PRS.2 this  
 ‘What is this that you are doing?’
- b. Hindi-Urdu  
 (✓*yeh*) *tum* (✓*yeh*) *kya:* (×*yeh*) *kar rahe* *ho* (×*yeh*)?  
 this you this what this do PROG.MPL be.PRS.2 this  
 ‘What is this that you are doing?’

How might we derive the acceptable orders in (24), where the egocentric pronoun is not in an initial position? Both Bangla and Hindi-Urdu allow for scrambling and we can reasonably propose that material that precedes the egocentric pronoun gets there via clause-local scrambling. But then why can’t *wh*-phrases scramble past the egocentric pronoun – if they could, the unacceptable orders in (24) would not be so. We cannot appeal to scrambling not being available to *wh*-phrases. They can in fact be scrambled, just not past the egocentric pronoun. Consider the following example where the indirect object can be scrambled past the subject (25b) but not past the egocentric pronoun (25c).

- (25) a. *yeh* *Ram kaun=se larke=ko pra:iz de raha:* *he?*  
 this Ram which boy=DAT prize give PROG.MSG be.3SG  
 ‘Who is this boy to whom Ram is giving the prize?’

---

<sup>2</sup>We noted earlier that egocentric questions can also be based on polar questions and alternative questions, which lack argumental *wh*-XPs though they may contain the Polar Question Particle *kya:*. We find with these that (1) the default order has the egocentric pronoun in sentence-initial position, (2) non-focused material may precede the egocentric pronoun but focused material cannot, and (3) a PQP if present must follow the egocentric pronoun.

- i. Hindi-Urdu polar questions
- a. *yeh* (*kya:*) *tum* (*kya:*) [<sub>F</sub>*Mina=se*] *ba:t kar rahe* *ho?*  
 this PQP you PQP Mina=INS talk do PROG.MPL be.PRS.2  
 ‘Are you talking to Mina?’
- b. \**kya:* *yeh* *tum* [<sub>F</sub>*Mina=se*] *ba:t kar rahe* *ho?*  
 PQP this you Mina=INS talk do PROG.MPL be.PRS.2  
 ‘Are you talking to Mina?’
- c. *tum* *yeh* (*kya:*) [<sub>F</sub>*Mina=se*] *ba:t kar rahe* *ho?*  
 you this PQP Mina=INS talk do PROG.MPL be.PRS.2  
 ‘Are you talking to Mina?’
- d. \**tum* [<sub>F</sub>*Mina=se*] *yeh* *ba:t kar rahe* *ho?*  
 you Mina=INS this talk do PROG.MPL be.PRS.2  
 ‘Are you talking to Mina?’

- b. *yeh kaun=se larke=ko Ram pra:iz de raha: hε?*  
 this which boy=DAT Ram prize give PROG.MSG be.3SG  
 ‘Who is this boy to whom Ram is giving the prize?’
- c. *\*kaun=se larke=ko yeh Ram pra:iz de raha: hε?*  
 which boy=DAT this Ram prize give PROG.MSG be.3SG  
 intended: ‘Who is this boy to whom Ram is giving the prize?’

So why can non-*wh*-phrases scramble past the egocentric pronoun but not *wh*-phrases? Let’s spell out the proposal in Rizzi (1997) a little further. Rizzi (1997) argues for the following hierarchy for Italian, which has topic positions above and below the landing site of *wh*-movement.<sup>3</sup>

$$(26) \quad [_{TopP} [_{CP} \text{wh-XP} [_{C'} C[+Q] [_{TopP} \dots]]]]$$

We assume that the egocentric pronoun occupies the high topic position in Bangla and Hindi-Urdu. *wh*-phrases in these languages do not obligatorily and overtly move to [Spec, CP]; moreover, these languages being head-final, there is no visible reflex of the C[+Q] head on the left periphery.

$$(27) \quad \text{Bangla/Hindi-Urdu:}$$

$$[_{TopP} \text{e/yeh} [_{CP} \dots \text{wh-XP} \dots C[+Q]]]$$

A *wh*-XP must appear in the scope of a C[+Q] and the projection in which the egocentric pronoun appears needs to be higher than the question CP. This is possible when the egocentric pronoun precedes the *wh*-XP as in (27). But when the *wh*-XP precedes the egocentric pronoun, it means the question CP is higher than the egocentric pronoun’s TopP. This runs afoul of our initial proposal about the location of the the egocentric pronoun in a high topic position. The location of a non-*wh*-XP is uninformative about the scope of the question CP and hence a non-*wh*-XP that precedes the egocentric pronoun does not create a problem.

We conclude that the egocentric pronoun is generated in a high topic position which is followed by a question CP. Any such sequence is technically grammatical. Whether such a sequence is actually acceptable depends upon whether the speaker and the hearer can together use the egocentric pronoun to individuate an event that the associated question can be posed of. These two components: the individuation of the event and the fact that the questions are posed of this event are what make egocentric questions differ from normal questions meaning-wise. We turn to them in the following sections.

#### 4 How egocentric questions differ from normal questions and why

We need to identify the limitations on the application of egocentric questions in order to pinpoint the function of the egocentric pronoun. We accomplish this by contrasting ego-

<sup>3</sup>The evidence for a lower position comes from embedded questions where the *wh*-phrase can precede or follow a topic. The \*Wh-Top order in (22) follows from obligatory verb movement to C in matrix questions.

centric questions with the corresponding plain questions. We see that egocentric questions are only felicitous in a subset of the contexts where plain questions are felicitous.

(28) Context: Riya returned home late at night. Riya's partner believes that Riya is cheating on him, and she was with Ravi all this time (though he has no proof of it). Riya's partner asks:

a. Bangla

(#e) *tui kar shathe chili?*

this you who.GEN with were

'Who is it that you were with?'

b. Hindi-Urdu

(#yeh) *tum kis=ke sa:th thi:?*

this you who=GEN with be.PST.F

'Who is it that you were with?'

(29) Context: I am aware that you go out every evening. However, I do not possess the concrete proof that you went out anywhere yesterday. The day after, I ask you:

a. Bangla

(#e) *tui gotokal kothay giyechili?*

this you yesterday where go.PRF.PST.2

'Where have you gone to yesterday?'

b. Hindi-Urdu

(#yeh) *tum kal kahā: gaye the?*

this you yesterday where go.PFV.MPL be.PST.MPL

'Where had you gone yesterday?'

(30) Context: According to what I know, you go out every evening. I even witnessed you heading out yesterday. I thus ask you the next day:

a. Bangla

(e) *tui gotokal kothay giyechili?*

this you yesterday where go.PRF.PST.2

'Where have you gone to yesterday?'

b. Hindi-Urdu

(yeh) *tum kal kahā: gaye the?*

this you yesterday where go.PFV.MPL be.PST.MPL

'Where had you gone yesterday?'

#### 4.1 Questions about a particular event

The set of data constructed above highlights the fact that the use of the egocentric pronoun in question situations is only acceptable if we have proof of some sort regarding the specific incident that the question is concerning. We cannot utter an egocentric question in a mere belief context as seen in (28) which describes a scenario where Riya's partner would conceivably accept the possibility that Riya and Ravi go out together. However, there is no specific event or occasion in his mind, to which he can refer while asking Riya's whereabouts using the egocentric *e/yeh*. To put it differently, these egocentric questions limit the inquiry to an event picked out by the demonstrative; if such an event is not available, they are deemed infelicitous. Now, if we modify the scenario slightly and suppose that the speaker, Riya's partner, just checked Ravi's status on WhatsApp which displays their hangout photos from that night, an egocentric question in (28) would be felicitous, as he now has a specific event to refer to while uttering the egocentric question. Due to the lack of a specific event that the inquiry can be associated with, the egocentric question is likewise infelicitous in the context of (29). In contrast, uttering an egocentric question is perfectly acceptable in a situation like (30), because the speaker has a particular event in his/her mind to which (s)he is referring while asking the question in (30). Though we have certain limitations while uttering egocentric questions, a straightforward question is acceptable in all these cases.

One might however think that egocentric questions are felicitous only in cases of direct evidence as shown in (30). Such is not the case. Look at the scenario in (31) where the speaker lacks direct proof, but the egocentric question is still admissible because there is some indirect evidence.

(31) Context: During the meeting, Mina misbehaved with her employee Ram. After learning about Mina's actions at the meeting, the speaker asks the following questions to her:

a. Bangla

(*e*) *tui kemon byabohar korli Ram-er shathe?*

this you how behaviour do.PST.2 Ram-GEN with

'How did you behave with Ram?'

b. Hindi-Urdu

(*yeh*) *tum=ne Ram=ke sa:th kaisa: bartaav kiya:?*

this you=ERG Ram=GEN with how behaviour do.PFV.MSG

'How did you behave with Ram?'

To put it briefly, an egocentric question needs some question event, as mentioned in §3.1, to be present in the very first place, which the speaker of the egocentric question needs to be able to pick out with the egocentric pronoun. The question event needs to be an actual event. The following example shows that future events do not count as actual events. To



get to an actual event, we need more. Consider a context where you were supposed to go to Delhi next week, a plan that we are both aware of. In such a context, an egocentric question is not ok. However, if you fall terribly sick the day before your trip and I know this, an egocentric question becomes ok. But speaker knowledge is key – if I do not know about your illness, we are back to square one and an egocentric question is out. In all three situations, however, the plain question is felicitous.

- (32) Case 1: no sickness: egocentric #  
 Case 2: sickness, Speaker knows of sickness: egocentric ok  
 Case 3: sickness, Speaker does not know of sickness: egocentric #

a. Bangla

(e) *tui kibhabe dilli jabi?*  
 this you how Delhi go.FUT.2  
 ‘How will you go to Delhi?’

b. Hindi-Urdu

(yeh) *tum dilli kaise ja:-oge?*  
 this you Delhi how go.FUT.M.2  
 ‘How will you go to Delhi?’

#### 4.2 The role of the speaker and the hearer

Since a question involves both the speaker and the hearer, it is important to identify the role of the hearer in these question situations. (33) depicts a scenario where only the speaker has access to the event to which the egocentric question is referring. In such a case, the egocentric question is not felicitous.

- (33) Context: At the meeting, Mina acted inappropriately towards her employee Ram. After observing Mina’s actions, the speaker decides to ask Ravi about the situation. However, the speaker does not know that not only did Ravi not attend the meeting, he also didn’t hear anything about its outcome. The speaker asks Ravi:

a. Bangla

(#e) *Mina kemon byabohar korlo Ram-er shathe?*  
 this Mina how behaviour do.PST.3 Ram-GEN with  
 ‘How did Mina behave with Ram?’

b. Hindi-Urdu

(#yeh) *Mina=ne Ram=ke sa:th kaisa: bartaav kiya:?*  
 this you=ERG Ram=GEN with how behaviour do.PFV.MSG  
 ‘How did Mina behave with Ram?’

Let’s start with the plain question in (33), which is not infelicitous. However in the given context, Ravi is not in a position to answer this question and he would probably respond by indicating that he doesn’t know. But the egocentric question is infelicitous as Ravi cannot identify the event that the egocentric pronoun is supposed to pick out.

We see that it is insufficient to have only the speaker’s access to the question scenario. In order to understand the hearer’s function, we must take into account circumstances in

which the hearer—that is, the individual to whom the question is addressed—does not participate in the question situation. This is what we accomplished in (33). In (30) and (31) where the hearer is a part of the question situation, (s)he always has access to the event which the egocentric question is aiming at.

The following table lists the favorable circumstances for asking egocentric questions. These questions are referred to in this study as ‘egocentric’ questions because they are primarily speaker-oriented, meaning they are expressed from the speaker’s point of view; that is, they can only be asked if the speaker possesses some evidence, either direct or indirect, regarding the question scenario. Note though that as we have just seen, the hearer also needs to be able to identify the intended referent of the egocentric pronoun.

Situations	Status	Sp/Addr
DIRECT EVIDENCE:	✓	Both
INDIRECT EVIDENCE:	✓	Both
BELIEF ONLY:	✗	Sp but not Addr
KNOWLEDGE W/O EVIDENCE:	✗	Sp but not Addr

Table 1: Situations where egocentric questions are acceptable

### 4.3 Partial versus complete Access: information seeking versus rhetorical

Earlier in §2.1 and §2.2, we reported that this egocentric *e/yeh* is also flexible with rhetorical and *wh*-exclamative instances. We’ve already mentioned the importance of the speaker/hearer access to the question situation which distinguishes canonical questions from egocentric questions. Now understanding the type of access and information available to the speaker/addressee about a question situation helps in identifying what type of egocentric question is being uttered i.e., whether it is an information seeking egocentric question, or an exclamative/reproach scenario.

We get a rhetorical question or an exclamation if the speaker/hearer has complete access and knowledge of the scenario under consideration, as seen in (14/15/17) respectively. It becomes an information-seeking question when the available data is incomplete (16). In (14/15/17), the speaker has both complete access and full information about the question situation, whereas in (16) though the speaker has access to the question situation, the information available to the speaker is insufficient or incomplete.

This way of characterizing the distribution of information seeking versus rhetorical readings in egocentric questions also gives us a way of explaining the slippery intuition that there is something about the form of egocentric questions that lends them to being more easily rhetorical/exclamative. We believe this intuition comes from the fact that because of their form, egocentric questions can never be used in a state of total ignorance about the question event. This is in contrast to plain questions, where ignorance about the question

event (beyond an assumption that there is such an event) could even be seen as a default state of affairs, though partial information and full information is also possible.

## 5 Unsolved mysteries

### 5.1 Proximal and distal demonstratives

The egocentric pronoun, *e/yeh* ‘this’, that we have seen so far in this paper, is the proximal demonstrative pronoun. Apart from the proximal demonstrative *e*, there are two other demonstratives in Bangla, which has a tripartite demonstrative system: <sup>4</sup> distal (e.g. *o*), and anaphoric (e.g. *se*) (Guha, 2020). We note that the distal and anaphoric demonstrative are not freely compatible with egocentric questions (34) but that there are egocentric questions where the distal demonstrative *o* is ok (35).

- (34) Context: You and I are classmates. I heard that yesterday in class you solved a mathematics puzzle in a very unique way. Today, I asked you the following:

*e/\*o/\*se tui kal onko-ta kibhabe solve korechili?*  
 this/\*that/\*that you yesterday mathematics-CL how solve do.PRF.PST.3

‘What was the way you used solving the mathematics puzzle?’

- (35) *o tui ki korli?*  
 DIST.DEM you what do.PST.2

‘What was that you did?’

Demonstratives in Bangla combine with classifiers (Dayal, 2012, 2014) but *e* with classifier *-ta* does not freely appear in egocentric questions.

- (36) *e-ta kinbo.*  
 PROX.DEM-CL buy.FUT.1  
 ‘I/we will buy it.’

- (37) Context: You and I are classmates. I heard that yesterday in class you solved a mathematics puzzle in a very unique way. Today, I asked you the following:

*e(\*-ta) tui kal onko-ta kibhabe solve korechili?*  
 this-CL you yesterday mathematics-CL how solve do.PRF.PST.3

Intended: ‘What was the way you used solving the mathematics puzzle?’

---

<sup>4</sup>Hindi-Urdu only has a bipartite system, contrasting the proximal demonstrative *yeh* with the distal *dvo*. Replacing the proximal demonstrative *yeh* with the distal *vo* in an egocentric question leads to strong deviance. Classifiers play a minor role in Hindi-Urdu and do not combine with demonstratives.

- i. *#vo tum kya: kar rahe ho?*  
 that you what do PROG.MPL be.PRS.2  
 Intended: ‘What are you doing?’



‘Which boy did you not praise at all?’

b. Hindi-Urdu

*tum=ne kis laṛke=ki bilkul=bhii ta:ri:f nahĩ: ki?*  
you=ERG which boy=GEN.F at.all=EVEN praise NEG do.PFV.F

‘Which boy did you not praise at all?’

(42) Context: At a party, you notice everyone but one particular boy.

a. Bangla

*tui kon chele-ta-ke ektu-o patta dili na?*  
you which boy-CL-ACC little-EVEN interest give.PST.2 NEG

‘To which boy did you not show any signs of interest?’

b. Hindi-Urdu

*tum=ne kis laṛke=ko bilkul=bhii la:in nahĩ: di?*  
you=ERG which boy=GEN.F at.all=EVEN line NEG give.PFV.F

‘To which boy did you not show any signs of interest?’

We find that the egocentric variants of the above are good but only in a more restricted context.

(43) Context: a situation where you are talking to a boy whom everyone praises and the conversation involves you berating this boy.

a. Bangla

*e tui kon chele-ta-r ektu-o proshongsha korli na?*  
this you which boy-CL-GEN little-EVEN praise do.PST.2 NEG

‘Who was this boy who you did not praise at all i.e. who you berated?’

b. Hindi-Urdu

*yeh tum=ne kis laṛke=ki: bilkul=bhi: ta:ri:f nahĩ: ki?*  
this you=ERG which boy=GEN.F at.all=EVEN praise NEG do.PFV.F

‘Who was this boy who you did not praise at all i.e. who you berated?’

(44) Context: a situation which involves you and a boy whom almost everyone has a crush on but you are totally ignoring.

a. Bangla

*e tui kon chele-ta-ke ektu-o patta dili na?*  
this you which boy-CL-ACC little-EVEN interest give.PST.2 NEG

‘Who was this boy towards whom you did not show any signs of interest?’

b. Hindi-Urdu

*yeh tum=ne kis laṛke=ko bilkul=bhi: la:in nahĩ: di:?*  
this you=ERG which boy=GEN.F at.all=EVEN line NEG give.PFV.F

‘Who was this boy towards whom you did not show any signs of interest?’

In contrast to sentential negation on which there seem to be restrictions in egocentric questions, narrow-scope negation is freely possible. Consider the following data:

(45) a. Bangla

*e tui amake na janiye kothay geli?*  
 this you me NEG inform.CP where go.PST.2  
 ‘Where is it that you went without informing me?’

b. Hindi-Urdu

*yeh tum mujhe bina: bata:ye kahā: cale gaye?*  
 this you me.DAT without inform.CP where go.PFV.MPL go.PFV.MPL.2  
 ‘Where is it that you went without informing me?’

Likewise while there are restrictions on sentential negation, the corresponding ‘negative’ situations which are not formally negated are fully unrestricted.

(46) Bangla

a. sentential negation: restricted

*e tui kake shomman dekhali na?*  
 this you whom respect show.PST.2 NEG

Intended: ‘Who was this person who you did not show respect?’

✓Context: uttered in a context where you surprisingly did not show respect to a delegate who everyone respects.

b. negative predicate: unrestricted

*e tui kake oshomman korli?*  
 this you whom disrespect do.PST.2

‘Who was this person who you disrespected?’

✓Context: uttered in a context where you surprisingly did not show respect to a delegate who everyone respects.

(47) Hindi-Urdu

a. sentential negation: restricted

*yeh tum kis=ki: ta:ri:f nahī: kar rahe the?*  
 this you who=GEN.F praise.f NEG do PROG.MPL be.PST.MPL

‘Who was this person who you were not praising at all i.e. who you were berating?’

✓Context: uttered in a situation where you are talking to a boy whom everyone praises but the conversation involves you berating this boy.

b. negative predicate: unrestricted

*yeh tum kis=ki: bura:i: kar rahe the?*  
 this you who=GEN.F praise.f do PROG.MPL be.PST.MPL

‘Who was this person who you were dispraising?’

✓Context: uttered in a situation where you are talking to a boy whom everyone praises but the conversation involves you berating this boy.

We leave the question of why sentential negation has a limited distribution in egocentric questions for future work.

## 6 Acknowledgements

We thank Diti Bhadra, Miriam Butt, Regine Eckardt, and Sreekar Raghotham for their comments as well as the reviewers and audience of the (f)ASAL 13 conference. We also extend our thanks to the reviewers of the WCCFL 41 conference. Special thanks to Maria Biezma and Ashwini Deo for meeting with us to discuss these ideas at length. All errors are our own.

## References

- Banerjee, Kousani. 2022. Wh-exclamatives call for a question semantics: The view from Bangla. In *Proceedings of the 30th edition of the Conference of the Student Organisation of Linguistics in Europe (ConSOLE XXX)*, 125–144.
- Biezma, María & Kyle Rawlins. 2017. Rhetorical questions: Severing asking from questioning. In *the proceedings of the 27th Semantics and Linguistic Theory (SALT 27)*, 302–322.
- Caponigro, Ivano & Jon Sprouse. 2007. Rhetorical questions as questions. In *the 11th proceedings of Sinn und Bedeutung (SuB 11)*, 121–133.
- Dayal, Veeneta. 2012. Bangla classifiers: Mediating between kinds and objects. *Rivista di Linguistica* 24(2). 195–226. [https://www.italian-journal-linguistics.com/app/uploads/2021/05/2\\_dayal.pdf](https://www.italian-journal-linguistics.com/app/uploads/2021/05/2_dayal.pdf).
- Dayal, Veeneta. 2014. Bangla plural classifiers. *Language and Linguistics* 15(1). 47–87. <https://doi.org/10.1177/1606822X13506151>.
- Guha, Ambalika. 2020. Wh-movement and the Distribution of *Sei* and Wh in the Bangla DP. In *Proceedings of the 8th Formal Approaches to South Asian Linguistics (FASAL 8)*, 1–7.
- Rett, Jessica. 2008. *Degree Modification in Natural Language*: Graduate School-New Brunswick Rutgers, The State University of New Jersey dissertation.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar: Handbook in generative syntax*, 281–337. Dordrecht: Springer Netherlands.
- Simpson, Andrew & Tanmoy Bhattacharya. 2003. Obligatory overt wh -movement in a wh -in-situ language. *Linguistic Inquiry* 34. 127–142. <https://doi.org/10.1162/002438903763255940>.

## Kannada through the lens of the NP/DP parameter

JOSHUA DEES, *University of Illinois at Urbana-Champaign*

### ABSTRACT

Bošković's (2008, 2012) NP/DP parameter suggests that languages with definite articles are DP languages and languages without definite articles are NP languages. However, more recent literature on the topic demonstrates that some article-less languages may be DP languages (see, e.g., Syed and Simpson 2017 on Bangla; Dees 2020 on Dholuo). This paper explores Kannada, an article-less Dravidian language, based on a number of Bošković's NP/DP-divide generalizations. The results demonstrate that Kannada patterns like the NP languages from Bošković (2008, 2012). It is then illustrated that Kannada may lack certain movements within the nominal domain that have been associated with other article-less languages which have been proposed as DP languages. These results provide necessary details for better understanding what the NP/DP 'divide' looks like cross-linguistically.

### 1 Introduction

Since Abney (1987) and Fukui & Speas (1986), it has been widely assumed that DP, at the very least, exists in languages with determiners. This assumption has led to three major camps of thought regarding DP level of structure: DP is universal (i.e. Bowers 1991; Longobardi 1994; Li 1998, 1999; Progovac 1998; Bašić 2004; Simpson 2005; Watanabe 2006; Park 2008; a.o.); DP is not present in article-less languages (i.e. Fukui 1988; Cover 1992; Chierchia 1998; Cheng & Sybesma 1999; Willim 2000; Baker 2003, 2005; Bošković 2008, 2012; Despić 2011; Talić 2015, a.o.); and there is no DP (i.e. Bruening 2020). In this paper, I focus primarily on the camp in which it is assumed that DP is not present in article-less languages.

Specifically, I explore Kannada, an article-less Dravidian language, from the perspective of Bošković's (2008, 2012) parameter. Kannada does not have a definite article, as is illustrated by the ambiguity between definite and indefinite interpretations of the bare noun, *ma:wannu* 'mango.ACC' in (1).

- (1) ra:da ma:w-annu ko|eda ſeniwa:ra tinda|u  
Rada mango-ACC last Saturday eat.3SG.F.PAST  
'Rada ate a/the mango last Saturday'

Bošković's (2008, 2012) NP/DP parameter proposes that languages with (definite) articles are DP languages, whereas languages without (definite) articles are NP languages. However, in more recent literature on the topic, it has been argued that Bošković's NP/DP divide is not as simple as a two-way divide. Some article-less languages appear to pattern like DP languages and present evidence for a DP projection in the nominal domain (see, e.g., Syed & Simpson 2017 on Bangla; Dees 2020 on Dholuo). For these languages, it has been argued that phrasal and/or head movement can also trigger the need for a DP projection in a language. Therefore, I seek to find out whether Kannada patterns like the article-less languages Bošković (2008, 2012) discusses or whether it is more similar to article-less Bangla and/or Dholuo.



The data for this study comes from one native-speaker consultant who identifies Kannada as their dominant language. The results of this study, indicate that Kannada patterns like the article-less languages in Bošković (2008, 2012). I suggest that the reason Kannada doesn't pattern like Bangla or Dholuo is due to the fact that phrasal movement and/or nominal head movement doesn't occur in Kannada. I demonstrate that the Kannada nominal modifier order is compatible with such an analysis.

This paper is structured as follows: Section 2 provides an overview of the relevant literature on Bošković's (2008, 2012) NP/DP parameter. Section 3 tests Kannada data against Bošković's (2008, 2012) NP/DP-divide generalizations, demonstrating that Kannada patterns like the article-less languages discussed in this work. Section 4 investigates the nominal modifier order of Kannada, leading to the proposal that Kannada is not a DP language. Section 5 provides concluding remarks and directions for future research.

## **2 The NP/DP parameter**

As previously mentioned, Bošković (2008, 2012) proposes an NP/DP parameter in which it is argued that some languages project DP in the nominal domain (i.e. English) and others do not (i.e. Bosnian-Croatian-Serbian). Accordingly, Bošković argues that the presence of overt definite articles is what triggers this parameter setting. Thus, languages with overt definite articles are DP languages and languages without overt definite articles are NP languages. Additionally, Bošković (2008, 2012) posits a set of generalizations associated with the presence/absence of definite articles, which he suggests corresponds to the NP/DP divide.

While Bošković's (2008, 2012) NP/DP typology remains borne out in a significant number of languages, recent literature on the topic of the NP/DP parameter suggests that it is too restrictive. More specifically, some authors have argued there is a split amongst DP languages. Talić (2017), for example, argues that there is a three-way distinction of languages: Article-less languages, affixal article languages, and non-affixal article languages. Oda (2022) further argues that a three-way distinction is also too restrictive, and offers that the NP/DP divide is more of a "fine-grained scale". On the other end of the parameter, authors have argued that some article-less languages are DP languages. For example, Syed & Simpson (2017) argues that Bangla (Indo-Aryan), which lacks a definite article is a DP language. They propose that phrasal movement, and possibly head movement can act as a trigger for the NP/DP parameter (an overview of this analysis is provided in section 2.2). Dees (2020) similarly argues that Dholuo (Nilo-Saharan), which is an article-less language, projects DP. This study proposes that head movement can act as a trigger for the NP/DP parameter (an overview of this analysis is provided in section 2.3).

In what follows, I provide an overview of a set of Bošković's (2008, 2012) generalizations that are relevant to the study (section 2.1), an overview of Syed & Simpson's (2017) analysis of Bangla (section 2.2), and an overview of Dees' (2020) analysis of Dholuo (section 2.3).

### **2.1 Bošković's (2008, 2012) generalizations**

As previously mentioned, Bošković (2008, 2012) posits a set of generalizations associated with the presence/absence of a (definite) article in a given language. Crucially, Bošković

takes this as support for the NP/DP parameter. In this section, I provide an overview of a subset of these generalizations that are associated with Syed & Simpson (2017), Dees (2020), and the current analysis of Kannada.

### 2.1.1 Negative raising generalization

Bošković (2008, 2012) posits a negative-raising generalization, in which languages without (definite) articles disallow negative raising (i.e. strict negative polarity item licensing under negative raising) and languages with (definite) articles allow it. Negative raising is understood as negation being interpreted in the matrix or embedded clause of sentences like *John does not believe Mark is smart*, as is illustrated in (2).

- (2) a. [John does **not** believe [Mark is smart]]  
 b. [John believe [Mark is **not** smart]]

As Bošković (2008, 2012) notes, we cannot rely solely on interpretation to diagnose whether negative raising is present in a language. Instead, drawing from Lakhoff (1969), Horn (1978), and Gajewski (2007), we can confirm the embedded clause option in (2b) using strict clause-mater negative polarity items (NPIs).

Strict NPIs require negation, as is demonstrated in the grammaticality of sentences like (3a-b) and the ungrammaticality of sentences like (3c-d).

- (3) a. John didn't leave **until yesterday**  
 b. John hasn't visited her **in at least two years**  
 c. \*John left **until yesterday**  
 d. \*John has visited her **in at least two years** (Bošković 2008: 106)

In (3a-b), the strict NPI is licensed by negation, whereas in (3c-d) there is no negation to license the strict NPI resulting in ungrammatical sentences. Furthermore, long distance licensing of strict NPIs is not allowed. This is demonstrated in (4) with a non-negative-raising verb *claim* in the matrix clause.

- (4) a. \*[John didn't claim [that Mary would leave [NPI **until tomorrow**]]]  
 b. \*[John doesn't claim [that Mary visited him [NPI **in at least two years**]]]  
 (Bošković 2008: 106)

In (4), since *claim* is a non-negative raising verb, negation must originate in the matrix clause (as opposed to raising from the embedded clause). Thus, due to the ungrammaticality of (4a-b), it is concluded that long-distance licensing of NPIs is disallowed. However, with attitude predicates like *believe*, it appears NPIs can occur in the embedded clause despite negation being in the matrix clause (5).

- (5) a. [John didn't believe [that Mary would leave [NPI **until tomorrow**]]]  
 b. [John doesn't believe [that Mary has visited him [NPI **in at least two days**]]]  
 (Bošković 2008: 106)

Since long-distance licensing is disallowed, as is demonstrated in (4), Bošković & Gajewski (2011) argues that (5a-b) are not instances of long-distance licensing. Instead, the licensing of the NPIs in (5a-b) occurs in the embedded clause and negation raises out of the embedded clause into the matrix clause.

Bošković (2008, 2012) highlights that negative raising is disallowed in article-less languages such as Bosnian-Croatian-Serbian, Slovenian, Polish, Russian, Turkish, Korean, Japanese, and Chinese. However, it is allowed in article languages such as English, German, French, Portuguese, Romanian, Bulgarian, and Spanish.

The rationale for connecting negative raising to the presence/absence of (definite) articles comes from Bošković & Gajewski (2011), in which it is argued that there is a similarity in interpretation of definite plurals and negative-raising predicates (NRPs). This similarity is the Excluded Middle (see Bartsch 1973; Horn 1989; Gajewski 2007). Based on this similarity Bošković & Gajewski (2011) propose that while attitude predicates are typically analyzed as quantifiers over worlds (6a), some (such as NRPs like *believe*) may also take part in distributive plural predication and denote sums of worlds (6b).

- (6) a.  $\text{all}(\text{BELa}) = \lambda p. \text{BELa} \subseteq p$   
 b.  $\text{the}(\text{BELa}) = \text{the sum of a's belief worlds}$  (Bošković & Gajewski 2011)

Distributive plural predication is triggered when an attitude predicate is constructed with a definite article, as in (6b). Bošković & Gajewski (2011) suggest that because of the Excluded Middle, these types of attitude predicates create statements that are true if the modal base is a subset of the embedded proposition, but false if the modal base is separate from the embedded proposition. Because of this, they argue that when this type of attitude predicate is negated the negation is interpreted as if it is in the embedded clause. And from this, Bošković & Gajewski (2011) propose that attitude verbs which select the distributive definite plural semantics are NRPs, while those that select universal quantification are not NRPs. They further argue that, in languages like English, which allow negative raising, the NRP *believe* involves the definite determiner. Therefore, the presence of D is responsible for the presence of negative raising in a language.

### 2.1.2 Sequence of Tense generalization

Bošković (2012) posits a sequence of Tense generalization, in which sequence of Tense (SOT) is only found in languages with (definite) articles. SOT is observed in languages like English where sentences in which a past-tensed clause is embedded under a past-tensed attitude predicate have two possible readings: The non-past/simultaneous reading and the anteriority reading, as is illustrated in (7).

- (7) *John thought that Mark was ill.*  
Non-past/simultaneous reading: John thought, “Mark is ill”  
Anteriority reading: John thought, “Mark was ill”

Bošković (2012) argues that languages without (definite) articles do not have SOT, this is exemplified with Bosnian-Croatian-Serbian in (8) and (9).

- (8) Jovan je vjerovao da je Marija bolesna

- Jovan is believe that is Marija ill  
 ‘John believed that Mary is ill’ (non-past/simultaneous) (Bošković 2012: 214)
- (9) Jovan je vjerovao da je Marija bila bolesna  
 Jovan is believe that is Marija been ill  
 ‘John believed that Mary was ill’ (anteriority) (Bošković 2012: 214)

The only way to obtain a non-past/simultaneous reading in Bosnian-Croatian-Serbian is with a present under past construction, as in (8). A past under past construction only has an anteriority reading in Bosnian-Croatian-Serbian, as in (9). It is also important to note that some languages with definite articles (e.g., Hebrew) do not have SOT, hence the one-way generalization.

The rationale for connecting SOT and the presence/absence stems from the concept of parallelism across domains (i.e. Abney 1987; Chomsky 2005; Bošković 2008, 2012; Todorović 2016; Talić 2015, 2017). In this case, the generalization relies on parallelism across the nominal and clausal domains. Specifically, if a language projects DP (a definite article) it should project TP (Tense morphology). SOT fits into this story when considering Stowell (1993, 1995a,b) and Kusumoto (2005). Under these approaches to SOT, the past tense morpheme receives its value from a higher anteriority operator PAST located in T. For the anteriority reading of an English sentence like (7), there are PAST operators in both the matrix and the embedded T, so the past tense morphemes receive their anteriority values from two different operators, locating each event at distinct points in the past, hence the anteriority reading. This is represented in (10).

- (10) [**PAST** John believe-**past** [that **PAST** Mark be-**past** sick]]

For the simultaneous reading, the PAST operator is only present in the matrix T. Both past tense morphemes, in this case, receive their value from the same operator, locating each event at the same point in the past, hence the non-past/simultaneous reading. This is represented in (11).

- (11) [**PAST** John believe-**past** [that Mark be-**past** sick]]

Assuming parallelism across the nominal and clausal domains, an NP (article-less) language would lack TP. As Bošković (2012) describes, due to the lack of TP in these languages, the operator PAST is not available. Therefore, NP languages cannot have a past tense morpheme like that in (10) and (11), which introduces variables and is licensed by PAST. Instead of a past tense morpheme, these languages have elements that carry a lexically specified meaning and add to temporal interpretations by saturating the time argument slot of the predicate. Since these elements are not variables, when they are embedded in the complement of an attitude predicate that is anchored with the past tense, it is not possible for them to be quantified-in by an intentional verb. Hence, for NP languages like Bosnian-Croatian-Serbian, the simultaneous reading is unavailable for past-under-past, as in (9).

### 2.1.3 Adnominal genitives generalization

Bošković (2008, 2012) posits an adnominal genitives generalization in which languages without (definite) articles don't allow transitive nominals with two genitives. Willim (2000) demonstrates that article languages such as English, Arabic, Dutch, German (12a), and Catalan (12b) allow two nominal genitive arguments (both external and internal arguments can be genitive, where the genitive is realized via a clitic/suffix or a dummy P).

- (12) a. Hannibals Eroberung Roms (German)  
 Hannibal-GEN conquest Rome-GEN  
 'Hannibal's conquest of Rome'
- b. l'avaluació de la comissió dels resultats (Catalan)  
 The evaluations of the committee of the results  
 'The committee's evaluation of the results'
- (Bošković 2012: 186)

Bošković (2012) notes that the same holds for Portuguese, Basque, French, Greek, Hebrew, Icelandic, Macedonian, Bulgarian, Spanish, Welsh, Maltese, Maori, Samoan, Swedish; all of which are article languages. Willim (2000) demonstrates, on the other hand, that languages without (definite) articles such as Polish (13a), Czech (13b), Russian, and Latin disallow lexical genitives. Bošković (2008, 2012) also notes that Ukrainian, Chinese, Quechua, and Turkish (which lack a definite article) also don't allow adnominal genitives.

- (13) a. \*odkrycie Ameryki Kolumba (Polish)  
 discovery America-GEN Columbus-GEN  
 'Columbus' discovery of America'
- b. \*zničení Říma barbarů (Czech)  
 destruction Rome-GEN barbarians-GEN  
 'The barbarian's destruction of Rome'

Instead, in languages without (definite) articles the external argument is often realized via a PP headed by an adposition analogous to English *by* or inherent oblique Case.

- (14) a. odkrycie Ameryki przez Kolumba (Polish)  
 discovery America-GEN by Columbus  
 'The discovery of America by Columbus'
- b. zničení Říma barbary (Czech)  
 destruction Rome-GEN barbarians-INSTR  
 'The destruction of Rome by the barbarians'

#### 2.1.4 Superlative majority reading

Bošković (2008, 2012) posits a majority superlative reading generalization in which only languages with a definite article allow the superlative majority reading. Živanović (2008), for example, notes that English, German, Hungarian, Romanian, Macedonian, and Bulgarian [languages with (definite) articles] have a superlative majority reading; and Slovenian, Czech, Polish, Bosnian-Croatian-Serbian, Chinese, Turkish, and Punjabi [languages without (definite) articles] do not allow the superlative majority reading.

For instance, an English sentence like *most people drink beer* has a plurality reading in which ‘more than half the people drink beer’ in addition to a superlative majority reading in which ‘more people drink beer than any other drink’. While English has both of these readings, Slovenian (article-less language) only has the plurality reading.

- (15) Največ ljudi pije pivo (Slovenian)  
 Most people drink beer  
 ‘More people drink beer than any other beverage’ (Plurality reading, PR)  
 ‘\*More than half the people drink beer’ (Majority reading, MR)  
 (Bošković 2008: 106)

### 2.1.5 Inverse scope generalization

Bošković (2012) posits an inverse scope generalization, in which inverse scope readings are only available in languages with (definite) articles. Consider the English sentence *someone love everyone* in (16).

- (16) Someone loves everyone  
Narrow scope: One person that loves all the people  
Wide (inverse) scope: Everyone is loved by someone

As is illustrated in (16), both the narrow scope reading and the inverse scope reading are available in English. Bošković (2012) presents that the inverse scope reading is available in English, Spanish, Brazilian Portuguese, Macedonian, and Hebrew. However, it is not available in languages such as German, Basque, Dutch, Icelandic, Bulgarian, Welsh, Romanian, Japanese, Korean, Turkish, Persian, Hindi, Bangla, Chinese, Russian, Polish, Slovenian, Ukrainian, and Bosnian-Croatian-Serbian.

### 2.1.6 Exhaustivity presupposition

Bošković (2012) posits that possessors may induce exhaustivity presupposition only in DP languages. Partee (2006) observes that the English possessor in the phrase *Zhangsan’s three sweaters* presupposes that Zhangsan has exactly three sweaters. Whereas, the possessor in Chinese (17) does not exhibit any such exhaustivity presupposition.

- (17) Zhangsan de [san jian maoxianyi]  
 Zhangsan DE<sub>POSS</sub> three CL sweater  
 ‘Zhangsan’s three sweater’ (Bošković 2012: 191)

Bošković (2012) notes that Russian, Bosnia-Croatian-Serbian, Turkish, Japanese, Korean, Hindi, Bangla, Malayalam, and Maghi all pattern like Chinese (and Partee notes the same for Russian), whereas Spanish, Brazilian Portuguese, Italian, Basque, Hebrew, Dutch, and Arabic pattern like English.

## 2.2 Bangla and the NP/DP parameter

Bangla, an Indo-Aryan language, does not have a definite article. However, as Syed & Simpson (2017) notes, Bangla demonstrates word-order alternations that are associated



Non-negative raising predicates, like *dekhi* ‘see’, however, block the licensing of strict NPIs in lower clauses.

- (22) \*ami **dekhi-ni** je ram kal parti-te **kono**  
 I see-NEG that Ram yesterday party-at **any**  
**khabar** kheyechhe  
**food** eat.PRES.PERF (Syed & Simpson 2017: 6)

Therefore, Syed & Simpson (2017) argues that (21) is an instance of negative raising, rather than long-distance licensing. In this way, Bangla patterns like languages with (definite) articles following Bošković (2008, 2012).

### 2.2.2 Superlative majority reading in Bangla

According to Syed & Simpson (2017) Bangla also has the superlative majority reading. For example (23) has both the relative reading and the majority reading.

- (23) beši-r-bhag lok kal parti-te beer khelo  
 most people yesterday party-at beer drink-PAST  
PR: ‘more people drank beer than any other beverage at the party yesterday’  
MR: ‘more than half the people drank beer at the party’ (Syed & Simpson 2017: 7)

In this way, Bangla patterns like languages with (definite) articles following Bošković (2008, 2012).

### 2.2.3 Adnominal Genitives in Bangla

Finally, Syed & Simpson also demonstrate that Bangla permits transitive nominals with two genitives. This is illustrated in (24) and (25) in which there are two genitive arguments in each example.

- (24) ram-er kukur-er bheeti  
 Ram-GEN dog-GEN fear  
 ‘Ram’s fear of dogs’ (Syed & Simpson 2017: 8)

- (25) feluda-r badšahi angti-r rohosyo somadhan  
 Feluda-GEN royal ring-GEN mystery solution  
 ‘The solution of the mystery of the royal ring by Feluda’ (Syed & Simpson 2017: 8)

In this way, Bangla patterns like languages with (definite) articles following Bošković (2008, 2012).



### 2.3 Dholuo and the NP/DP parameter

As Dees (2020) demonstrates, Dholuo does not have articles. This is illustrated with the use of bare nouns for unique entities in (26) and the use of bare noun *mpira* ‘ball’ for both the indefinite and definite in (27).

- (26) a. a-neno                    **tjiej**  
           1SG-see                **sun**  
           ‘I see the sun’  
       b. n-a-neno                **ker**  
           PST-1SG-see         **president**  
           ‘I saw the president’

- (27) auma n-o-gwejo    n-a    **mpira** to    n-a-gwejo    **mpira**  
       Auma PST-PFV-kick to-1SG **ball**    and    PST-1SG-kick **ball**  
       ‘Auma kicked a ball to me and I kicked the ball back’

Dees (2020) argues that the order of nominal modifiers in Dholuo is indicative of N-to-D movement (cf. Carstens 1991, 1993, 2008). All modifiers are post-nominal, and the order of elements is N > Num > Dem. Adjectives are set aside in this analysis, as Dees (2020) argues they are relative clauses.

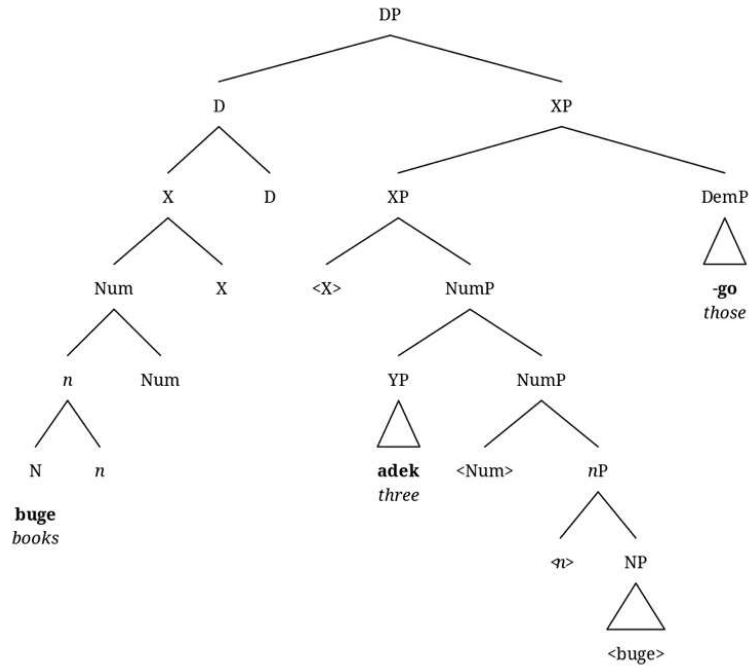
- (28) **bug-e**            **adek-go**        gin    ei        sanduk                    (N < Num < Dem)  
       **book-PL**        **three-those**    are    in        box  
       ‘Those three books are in the box’

Following Carstens (2008), Dees proposes the following structure for Dholuo nominals<sup>1</sup>:

---

<sup>1</sup> [DP  
    [D  
    [X  
    [Num  
    [\*n\*  
    [N  
    \*\*buge\*\*\  
    \*books\*  
    ]  
    [\*n\*]  
    ]  
    [Num]  
    ]  
    [X]  
    ]  
    [D]  
   ]  
 [XP  
 [XP  
 [ <X> ]  
 [NumP  
 [YP  
 ^\*\*adek\*\*\  
 ]

(29)



In (28), the demonstrative is adjoined to right edge of XP, accounting for the surface-level order. Num (YP) is an adjunct to NumP, and seems to always be adjoined to the left in Dholuo. Finally, Head movement of the noun occurs cyclically, moving to the left edge of the DP. For this reason, as it is hinted at in Syed & Simpson (2017), Dees (2020) argues that N-to-D movement can also trigger the NP/DP parameter.

In addition to this N-to-D movement analysis, Dees (2020) demonstrates that, like Bangla, Dholuo patterns like languages with (definite) articles following Bošković's (2008, 2012) generalizations. Specifically, Dees (2020) highlights the negative raising and sequence of Tense generalizations. A more recent investigation of Dholuo demonstrates that it also patterns like languages with (definite) articles following several other generalizations from Bošković (2008, 2012), as is demonstrated in Table 1.

---

```
*three*
]
[NumP
  [<Num>]
  [*n*P
    [<*n*>]
    [NP ^\<buge>]
  ]
]
]
]
]
[DemP ^\-*go**\
  *those*]
]
]
```

	Neg Raising	Sequence of Tense	Superlative Majority	Inverse Scope	Exhaustivity Presupposition
<b>Dholuo</b>	Y	Y	Y	Y	Y
<b>English (DP)</b>	Y	Y	Y	Y	Y
<b>BCS (NP)</b>	N	N	N	N	N

Table 1. A comparison of Dholuo, English (DP-language), Bosnian-Croatian-Serbian (NP-language) using six of Bošković's (2008, 2012) generalizations.

Following the data exemplifying N-to-D movement in Dholuo and the evidence that Dholuo patterns like languages with (definite) articles following Bošković (2008, 2012), Dees (2020) argues that Dholuo is a DP language.

### 3 Kannada and the NP/DP generalizations

As previously mentioned, Kannada does not have a (definite) article. This is demonstrated in (30) where the indefinite use of *ma:wu* 'mango' and *papajawannu* 'papaya.ACC' occurs with the numeral *undu* 'one' and the definite use is article-less (only accusative marked).

- (30) ra:da **undu** **ma:wu** mattu **papajawannu** karidisida|u.  
Rada **one** **mango** and **papaya.ACC** buy.3SG.F.PAST  
awa|a maga|u **ma:wannu** tinda|u maga **papajawannu**  
3SG.F.GEN daughter **mango.ACC** eat.3SG.F.PAST son **papaya.ACC**  
tinda  
eat.3SG.M.PAST  
'Rada bought a mango and a papaya. Her daughter ate the mango and her son ate the papaya'

Bare nouns are also used for unique readings. In (31a), the sentence 'the moon is beautiful' uses the bare noun *tʃendra* 'moon'. Likewise, in (31b), the sentence 'the sun rises in the morning' uses the bare noun *surija* 'sun'.

- (31) a. tʃendra sundarawa:gide  
moon beautiful  
'The moon is beautiful'  
b. surija be|aginadzawa udajawa:gutade  
sun morning.in rises  
'The sun rises in the morning'

Following Bošković's (2008, 2012) generalizations, Kannada should pattern like other languages without (definite) articles. However, as was demonstrated with Bangla and Dholuo, this is not always the case. Thus, in what follows, I test Kannada against a set of Bošković's (2008, 2012) NP/DP-divide generalizations. The results demonstrate that Kannada patterns like the languages without (definite) articles in Bošković (2008, 2012) and, thus, unlike Bangla and Dholuo.

### 3.1 Negative raising in Kannada

Like the article-less languages in Bošković (2008, 2012), Kannada does not have negative raising. This is demonstrated using the NPI *nanejawregu* ‘until yesterday’, which requires negation (i.e. strict NPI).

- (32) a. ra:da            **nanejawregu**            horada[-**ella**  
          Rada            **until.yesterday**            leave.3SG.F.PAST-NEG  
          ‘Rada didn’t leave until yesterday’  
       b. \*ra:da            **nanejawregu**            horata[u  
          Rada            **until.yesterday**            leave.3SG.F.PAST

The grammaticality of (32a), and the ungrammaticality of (32b), demonstrates that *nanejawregu* ‘until yesterday’ is a strict NPI. Furthermore, licensing of a strict NPI in and embedded clause when negation is in the matrix clause is disallowed for all attitude predicates. In (33), when the phrase *he:liko[[al-ella* ‘didn’t claim’ is in the matrix clause and the strict NPI, *naljaveregu* ‘until tomorrow’, is in the embedded clause, the sentence is ungrammatical. The same is true if *namba[-illa* ‘didn’t believe’ is in the matrix clause (34).

- (33) \*ra:ma **naljaveregu**            horaduta:nendu            ra:da  
       Rama **until.tomorrow**            leave.3SG.M.FUT            Rada  
       **he:liko[[al-ella**  
       **claim.3SG.F.PAST-NEG**  
       (34) \*ra:ma **naljaveregu**            horaduta:nendu            ra:da  
       Rama **until.tomorrow**            leave.3SG.M.FUT            Rada  
       **namba[-illa**  
       **believe.3SG.F.PAST-NEG**

Additionally, a sentence like (35) only has a reading in which negation is interpreted in the matrix clause. The reading in which the matrix clause is in the embedded clause is completely out.

- (35) ra:ma horaduta:nendu            ra:da namba[-illa  
       Rama leave.3SG.M.FUT            Rada believe.3SG.F.PAST-NEG  
       Reading A: ‘Rada didn’t hold the belief that that Rama left’  
       Reading B: ‘\*Rada held the belief that Rama didn’t leave’

Thus, like the article-less languages in Bošković (2008, 2012), and unlike article-less Bangla (Syed & Simpson 2017) and Dholuo (Dees 2020), Kannada does not have negative raising.

### 3.2 Adnominal genitives in Kannada

Kannada also does not have adnominal genitives, patterning like the article-less languages in Bošković (2012). In (36), the use of transitive nominals with two genitives is disallowed in Kannada.

- (36) \*nagara-**da**    na:ʃa            mahile-**da**  
           city-GEN    destruction    woman-GEN  
           ‘The woman’s destruction of the city’

Instead, just as Bošković (2012) highlights for other article-less languages, the external argument is realized via a PP headed by a post-position *inda* ‘from’, as is demonstrated in (37).

- (37) mahil-**inda**    nagara-da    na:ʃa  
           woman-**from**    city-GEN    destruction  
           ‘The destruction of the city from the woman’

Again, Kannada patterns like the article-less languages in Bošković (2012), and unlike article-less Bangla (Syed & Simpson 2017) and Dholuo (Dees 2020).

### 3.3 Superlative majority reading in Kannada

Kannada does not have a superlative majority reading, again, patterning like the article-less languages in Bošković (2008, 2012). To determine this, the context in (38) was provided to my consultant.

- (38) Context: Five students attend a gathering. Student 1 drank lemonade, student 2  
           Drank lemonade and wine, student 3 drank beer and wine, student 4 drank beer  
           and wine, and student 5 drank beer and wine.

When presented with a statement that would induce the plurality reading (39), my consultant claimed the statement was true within the context of (38). However, when my consultant was presented with a statement that would induce the majority reading (40), they claimed this was false within the context of (38).

- (39) bahalʃtu            vidjarti-gaʃu    wain    koɟjuttare  
           most            student-PL    wine    drink.3PL  
           ‘Most people drink wine’
- (40) \*bahalʃtu            vidjarti-gaʃu    bir      koɟjuttare  
           most            student-PL    wine    drink.3PL  
           ‘Most people drink beer’

Based on these results, it is concluded that Kannada patterns like the article-less languages in Bošković (2008, 2012), and unlike article-less Bangla (Syed & Simpson 2017) and Dholuo (Dees 2020).

### 3.4 Inverse scope reading in Kannada

Inverse scope readings are also unavailable in Kannada. A sentence like (41), which is in the unmarked SOV order, only has the narrow scope reading.

- (41) jaro            jalar-annu    pridisutarre  
         someone    everyone-ACC loves  
Narrow scope: ‘Someone love everyone’  
Wide (inverse) scope: ‘\*Everyone is loved by someone’

In Kannada the only way to get the inverse scope reading is to change the word order. Thus, it is concluded that Kannada patterns like the article-less languages in Bošković (2012), and unlike Bangla (Syed & Simpson 2017) and Dholuo (Dees 2020).

### 3.5 Exhaustivity presupposition in Kannada

Possessor’s do not induce an exhaustivity presupposition in Kannada either. A sentence like (42) cannot possibly mean Rada has exactly three sweaters.

- (42) ra:da-|a        moru    sweter-ga|u  
         Rada-GEN    three    sweater-PL  
         ‘Rada’s three sweaters’

This final test, is consistent with the rest of the tests, further illustrating that Kannada patterns like the article-less languages in Bošković (2012).

## 4 Kannada nominal modifiers

In section 3, I have established that, unlike Bangla (i.e. Syed & Simpson 2017) and Dholuo (i.e. Dees 2020), Kannada patterns like an article-less language following Bošković’s (2008, 2012) generalizations. In Syed & Simpson (2017), Bangla is argued to pattern like languages with (definite) articles due to phrasal movement higher in the nominal domain causing a DP projection. Similarly, Dees (2020) argues that Dholuo patterns like languages with (definite) articles due to N raising to D, thus DP being projected in the nominal domain. Both of these arguments are based on the idea that certain positional patterns in the nominal domain can trigger the NP/DP parameter. For this reason, I explore patterns in the Kannada nominal domain, specifically nominal modifier ordering, to see if there is reason to believe DP is projected in the language.

Kannada is consistently head-final. The Kannada nominal domain presents a strict prenominal ordering of modifiers (Dem > Num > Adj > N). This order is illustrated in (43). In (44), moving the noun changes the interpretation of the sentence. For example, in (44a), the nominal modifier (Adj) *kempu* ‘red’ can only modify the noun it precedes, thus the interpretation of the sentence changes from (43) ‘those two big red books are in the box’ to ‘those two big books are in the red box’.

- (43) a:        eraḍu    doḍḍa    kempu            pustaka-ga|u            *pettige-jallive*  
         those two    big    red            book-PL            *box-in*  
         ‘Those two big red books are in the box’

- (44) a. **a:**            **eraḍu** **doḍḍa**            **pustaka-gaḷu** *kempu*            *pettige-jallive*  
**those**    **two**    **big**            **book-PL**    *red*            *box-in*  
‘Those two big books are in the red box’
- b. **a:**            **eraḍu**            **pustaka-gaḷu** *doḍḍa* *kempu*            *pettige-jallive*  
**those**    **two**            **book-PL**    *big*    *red*            *box-in*  
‘Those two book are in the big red box’
- c. **a:**            **pustaka-gaḷu** *eraḍu* *doḍḍa* *kempu* *pettige-jallive*  
**those**    **book-PL**    *two*    *big*    *red*    *box-in*  
‘those book are in two big red boxes’/‘\*those two red books are in the box’
- d. **pustaka-gaḷu**    *a:*    *eraḍu* *doḍḍa* *kempu* *pettige-jallive*  
**book-PL**            *those* *two*    *big*    *red*    *box-in*  
‘the books are in those two big red boxes’/‘\*those two big red books are in the box’

Assuming a universal merge order of Dem > Num > Adj > N (i.e. Cinque 2000, 2005; Carstens 2008), the data in (42) and (43) is not incompatible with the idea that nouns remain low in Kannada.

Kannada nouns remaining low could provide an explanation for the fact that Kannada doesn’t pattern like Bangla (i.e. Syed & Simpson) or Dholuo (i.e. Dees 2020), and instead patterns like the article-less languages from Bošković (2008, 2012). In this case, Kannada both lacks a (definite) article and/or phrasal/head movement to trigger a DP projection. Thus, I conclude that Kannada is an NP language, whereas languages like Bangla and Dholuo are DP languages.

## 5 Conclusions

In this paper, I have demonstrated that Kannada patterns like the article-less (NP) languages from Bošković (2008, 2012). Following more recent work on article-less languages that demonstrates some project DP, it is crucial to investigate why some article-less languages would project DP whereas others may not. I propose that NP languages like Kannada differ from other article-less languages like Bangla (i.e. Syed & Simpson 2017) and Dholuo (i.e. Dees 2020) in that Kannada does not have phrasal or head movement in the nominal domain that would trigger a DP projection. This argument is still preliminary, however. While the nominal modifier order in Kannada is not inconsistent with the idea that there is no phrasal or head movement in the nominal domain, it is not certain that this is the case. Therefore, a deeper dive into the nominal domain of Kannada is necessary for future research on this topic.

## Acknowledgments

This paper wouldn’t have been possible without the help of my Kannada consultant, Veeresh Chitradurga Gangadhara. I would like to express my gratitude to Veeresh. Thank you so much for spending many hours with me teaching me about your language. Many thanks to my advisor, Dr. Aida Talić, who not only started my connection with Veeresh, but also gave advice and feedback at different stages of this project. I also want to extend my gratitude to the organizers and attendees of (f)ASAL 13.

## References

- Abney, Steven. 1987. *The English noun phrase in its sentential aspect*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
- Baker, Mark. 2003. *Lexical categories*. Cambridge University Press.  
<https://doi.org/10.1017/CBO9780511615047>
- Bartsch, Renate. 1973. 'Negative transportation' gibt es nicht. *Linguistische Berichte* 27. 1-7.
- Bašić, Monika. 2004. *Nominal subtractions and the structure of NPs in Serbian and English*. Tromsø: University of Tromsø MA thesis.  
<https://hdl.handle.net/10037/238>
- Bhattacharya, Tanmoy. 1999. *The structure of the Bangla DP*. London: University of London dissertation.
- Bošković, Željko. 2005. On the locality of left branch extraction and the structure of NP. *Studia Linguistica* 59(1), 1-45. <https://doi.org/10.1111/j.1467-9582.2005.00118.x>
- Bošković, Željko. 2008. What will you have, DP or NP? *Proceedings of the Northeast Linguistic Society* 37(1), 101-114. Amherst, MA: GLSA Publications.
- Bošković, Željko. 2012. On NPs and clauses. In G. Grewendorf & T.E. Zimmerman, eds., *Discourse and grammar: From sentence types to lexical categories*, 179-242. Boston, MA: DeGruyter Mouton. <https://doi.org/10.1515/9781614511601.179>
- Bošković, Željko. & Jon Gajewski. 2011. Semantic correlates of the DP/NP parameter. *Proceedings of the Northeast Linguistic Society* 39(1), 121-134. Amherst, Ma: GLSA Publications.
- Bowers, John. 1991. The syntax and semantics of nominals. In *Cornell University Working Papers in Linguistics* 10: 1-30. (Proceedings of the First Semantics and Linguistic Theory Conference.) Department of Modern Languages and Linguistics, Cornell University, Ithaca, NY.
- Bruening, B., (2020) "The head of the nominal is N, not D: N-to-D Movement, Hybrid Agreement, and conventionalized expressions", *Glossa: a journal of general linguistics* 5(1): 15. <https://doi.org/10.5334/gjgl.1031>
- Carstens, Vicki. 1991. *The morphology and syntax of determiner phrases in Kiswahili*. Los Angeles, CA: University of California dissertation.
- Carstens, Vicki. 1993. On nominal morphology and DP structure. In Sam Mchombo (ed.), *Theoretical aspects of Bantu grammar*. 151-182. Stanford, CA: Stanford University Press.
- Carstens, Vicki. 2008. DP in Bantu and Romance. In C. De Cat & K. Demuth, eds., *The Bantu- Romance connection: A comparative investigation of verbal agreement, DPs, and information structure*. 131-166. Amsterdam: John Benjamins.  
<https://doi.org/10.1075/la.131.10car>
- Chacón, Dustin. 2012. Head movement in the Bangla DP. *Journal of South Asian Linguistics* 4, 3-24.
- Cheng, Lisa Lai-Shen. & Rita Sybesma. 1999. Bare and not-so-bare nouns and the structure of np. *Linguistic Inquiry* 30, 509-542.  
<https://doi.org/10.1162/002438999554192>
- Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6, 339-405. <https://doi.org/10.1023/A:1008324218506>



- Chomsky, Noam. 2005. Three factors in the language design. *Linguistic Inquiry* 36, 1-22. <https://doi.org/10.1162/0024389052993655>
- Cinque, Guglielmo. 2005. On Deriving Greenberg's Universal 20. *Linguistic Inquiry* 36(3), 315-332. <https://doi.org/10.1162/0024389054396917>
- Dasgupta, Probal. 1983. The Bangla classifier /Ta/, its penumbra, and definiteness. *Indian Linguistics* 44. 11–26.
- Dayal, Veneeta. 2012. Bangla classifiers: Mediating between kinds and objects. *Rivista di Linguistica* 24(2). 195–226.
- Dees, Joshua. 2020. Dholuo: Evidence of negative raising and sequence of Tense in an article-less language. Presentation at the Annual Meeting of the Linguistic Society.
- Despić, Miloje. 2011. *Syntax in the absence of determiner phrase*. Storrs, Connecticut: University of Connecticut dissertation. <https://www.proquest.com/docview/905841026>
- Fukui, Naoki. 1988. Deriving the differences between English and Japanese. *English Linguistics* 5, 249–270. <https://doi.org/10.9793/elsj1984.5.249>
- Fukui, Naoki & Margaret Speas. 1986. Specifiers and projection. *MIT Working Papers in Linguistics* 8, 128-172.
- Gajewski, Jon. 2007. Neg-raising and polarity. *Linguistics and Philosophy* 30, 289-328. <https://doi.org/10.1007/s10988-007-9020-z>
- Horn, L. 1978. Remarks on neg-raising. *Syntax and Semantics* 9, 129-220.
- Kusumoto, K. 2005. On the Quantification over Times in Natural Language. *Natural Language Semantics* 13, 317–357. <https://doi.org/10.1007/s11050-005-4537-6>
- Lakoff, R. 1969. A syntactic argument for negative transportation. *Proceedings of Chicago Linguistics Society* 5, 140-147.
- Li, A. 1998. Argument determiner and number phrases. *Linguistic Inquiry* 29, 693–702.
- Li, A. 1999. Plurality in a classifier language. *Journal of East Asian Linguistics* 8, 75–99. <https://doi.org/10.1162/ling.1998.29.4.693>
- Longobardi, G. 1994. Reference and proper names. *Linguistic Inquiry* 25, 609-665.
- Park, S. 2008. *Functional categories: The syntax of dp and DegP*. Los Angeles, California: University of Southern California dissertation.
- Progovac, L. 1998. Determiner Phrase in a language without determiners. *Journal of Linguistics* 34, 165-179. <https://doi.org/10.1017/S0022226797006865>
- Simpson, Andrew. 2005. Classifiers and dp structure in southeast Asian languages. In R. Kayne & G. Cinque, eds., *Handbook of comparative syntax*. 806–838. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195136517.013.0018>
- Stowell, Tim. 1993. *Syntax of tense*. Los Angeles, California: University of California MS thesis.
- Stowell, Tim. 1995a. The Phrase Structure of Tense. In J. Rooryck & L. Zaring, eds., *Phrase Structure and the Lexicon*. 277-291. Dordrecht: Kluwer. [https://doi.org/10.1007/978-94-015-8617-7\\_10](https://doi.org/10.1007/978-94-015-8617-7_10)
- Stowell, Tim. 1995b. What is the meaning of the present and past tenses? In P.M. Bertinetto, V. Bianchi, J. Higginbotham, & M. Squartini, eds., *Temporal Reference: Aspect and Actionality Vol. 1: Semantic and Syntactic Perspectives*. 381-396. Rosenberg and Sellier.

- 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), 1- 24. <https://doi.org/10.5334/gjgl.186>
- Talić, Aida. 2015. Adverb extraction, specificity, and structural parallelism. *Canadian Journal of Linguistics* 60, 417-445. <https://www.muse.jhu.edu/article/618397>
- Talić, Aida. 2017. Long-form and short-form prenominal adjectives are not reduced relative clauses in Bosnian/Croatian/Serbian. *Linguistic Inquiry* 48(1), 194-212. [https://doi.org/10.1162/LING\\_a\\_00241](https://doi.org/10.1162/LING_a_00241)
- Todorović, Neda. 2016. *On the presence/absence of TP: Syntactic Properties and Temporal Interpretation*. Storrs, Connecticut: University of Connecticut dissertation.
- Watanabe, Akira. 2006. Functional projections of nominals in Japanese: Syntax of classifiers. *Natural Language and Linguistic Theory* 24, 241–306. <https://doi.org/10.1007/s11049-005-3042-4>
- Willim, Ewa. 2000. On the grammar of Polish nominals. In R. Martin, D. Michaels & J. Uriagereka, eds., *Step by step*. 319-346. MIT Press.

## ‘-Te/-e’ marker as a plural in Bangla

CHANDNI DUTTA, *IIT Delhi*

SAYANTANI BANERJEE, *IIT Delhi*

### ABSTRACT

The current work deals with the morphological marker *-Te/-e* in Bangla. It analyzes the form, function, and structural position of the marker and proposes that, among various other roles, this marker yields the interpretation of a plural marker in restricted contexts. This study further claims that along with *-raa*, *-Te* is another associative plural in the language found with conjoined DPs with added ‘collective’ or ‘together’ semantics. This claim becomes interesting in the backdrop of Bangla being a classifier language. This novel proposal deconstructs the dominant thesis regarding the absence of plural in a numeral classifier language. It also suggests that the definition of the plural is much more varied than what has been discussed in the literature.

## 1 Introduction

Bangla, along with other Eastern Indo-Aryan languages, is a nominative-accusative language with person agreement and no number-gender agreement. Bangla nominals do not exhibit a canonical number marker, as seen in the examples below.

- |                       |                 |                |
|-----------------------|-----------------|----------------|
| (1) (a) <i>chhele</i> | (b) <i>goru</i> | (c) <i>boi</i> |
| boy                   | cow             | book           |
| ‘boy/ boys’           | ‘cow/ cows’     | ‘book/books’   |

In the absence of any canonical number marker, Bangla uses numeral classifiers to denote the distinction between singular and plural (Dasgupta 1983, Bhattacharya 1999, Dayal 2012 a.o.). However, recent studies claim the existence of non-canonical plural in Bangla- *-raa* and *gulo* (Chácon, 2011; Biswas, 2013; Dayal, 2014). These two markers are used to express plurality in contexts without numeral. The paper provides novel empirical evidence that shows that the polyfunctional marker *-Te/-e* acts as a plural in Bangla. It gives a non-singular, collective interpretation in restricted contexts. *-Te/-e* comes with conjoined DPs to show plurality and makes a distinction between canonical co-ordinate phrases in the language. This proposal adds to the list of literature that supports the existence of plural in classifier languages. It also suggests that the definition of the plural is much more varied than what has been discussed in the literature.

Section 2 introduces the polyfunctional nature of *-Te*. Section 3 establishes that *-Te* expresses plural interpretation in restrictive contexts. However, it also establishes that *-Te* does not fit anywhere in Wiltschko’s (2008, 2021) number cline. Section 4 shows the overlap between *-raa* and *-Te*. Section 5 gives empirical evidence for the semantics of *-Te*. Section 6 proposes the structural position of the *-Te* marker. Section 7 concludes the study.

## 2 *-Te* as a multifunctional marker

This section shows how *-Te* is a multifunctional morpheme in Bangla, which has been the focus of discussion in the literature for a long. The marker *-Te* fulfills different functions in Bangla, such as locative (2), instrumental (3), or subject marker (4). In (2), it attaches to *bombe*, and in (3), it attaches to *chhuri*.

(2) *ami bombe-Te thak-i*  
 I bombay-LOC stay.PRS.HAB-1P  
 ‘I stay in Bombay’

(3) *ama-r angul chhuri-Te kete geche*  
 I-GEN finger knife-INS cut be.PERF-3P  
 ‘My finger got cut by knife’

This marker can not only appear with 3rd person DPs, but also with 1st and 2nd person pronouns (see below). Presently, they are primarily seen in poetic uses, but they are infrequent.

(4) *ami nei ama-Te*  
 I be.NEG I-LOC  
 ‘I am not in me’

(5) *toma-Te ami mugdho*  
 you-LOC I impress  
 ‘I am impressed with you’

(6) *dube achi Toma-Te*  
 drown be.1P you-LOC  
 ‘I am drowned in you’

Further, this *-Te* marker can also be seen on subjects as a subject marker irrespective of the verb types. For example, in (7-8), *-Te* comes with the transitive verb ‘cross’ and unaccusative verb ‘fall’ respectively. However, note that this marker is optional in such contexts. It means that the structures are not ungrammatical in *-Te*’s absence. Additionally, the agreement system remains obligatorily similar both in the presence and absence of the marker.

(7) *sita-(te) lokkhonrekha par koreche*  
 Sita-s lakshman-border cross do-PERF-3P  
 ‘Sita crossed the Lakshmanrekha’

(8) *sita-(te) aj khad-e pore g-ech-e*  
 Sita-s today cliff-LOC fall-NF go-PERF-3P  
 ‘Sita fell on the cliff’

Now, in the following sections, we concentrate on *-Te* acting as a marker imparting plural sense in restricted contexts.

### 3 *-Te* As a Plural Marker

Bangla presents a typical example of a numeral classifier language with no existence of canonical inflectional plural (Dasgupta 1983, Bhattacharya 1999, 2000, 2001, Dayal 2012 inter alia). The plurality of a nominal is expressed in terms of a numeral and a classifier (9).

(9) (a) <i>paanch-ta chhele</i>	(b) <i>paanch-ta goru</i>	(c) <i>paanch-ta boi</i>
Five-CL boy	five-CL cow	five-CL book
‘Three boys’	‘Three cows’	‘Three books’

Recent studies, however, contradict the above claim and present evidence for the presence of non-canonical plurals in Bangla- *raa* and *gulo* (Chácon 2011, Biswas 2013, Dayal 2014, Dutta, Kumari & Chandra 2021). Without a numeral, these two markers are used to express plurality (10-11). While *-raa* is acknowledged as an associative plural (Chácon 2011, Biswas 2013), *gulo* is considered a plural marker (Biswas 2013, Dutta, Kumari & Chandra 2021). See bellow for illustration.

(10) (a) <i>meye</i>	(b) <i>meye-ra</i>
girl	girl-RAA

‘(A) girl’

‘Girls’

- (11) (a) *chhagol*                      (b) *chhagol-gulo*  
goat                                      goat-GULO  
‘(A) goat’                              ‘Goats’

Along with these non-canonical plurals, we claim that Bangla has an optional marker *-Te/-e*, that yields non-singular readings in certain environments.

### 3.1 Properties of *-Te/e* as a plural

In specific contexts, the marker shows non-singular meaning with extra semantics. It is optional in nature but quite robustly used in the language. Some of the significant properties are discussed below:

The marker *-Te* yields a **together and intense** meaning involving two active participants/groups.

- (12) *montri-(Te)*                      *montri-(Te)* *kolakuli kor-ch-e*  
minister-TE                      minister-TE hug do-PROG-3P  
‘Two (groups of) ministers are hugging each other’

- (13) [*baba-ma*]-(*Te*)                      *khub jhogra kor-e*  
father-mother-TE                      very quarrel do-3P.PRS.HAB  
‘Father and mother argue a lot (among each other)’

The marker is limited to **two nominals**. The marker disappears if the number of arguments is more than two, as seen in (14). It again does not impart the plural sense when attached to a single DP as in (15).

- (14) \**ma baba chhele-Te khub jhogra kor-e*  
mother father boy-TE very argue do-3P.PRS.HAB  
‘Mother, father, and son argue a lot (among each other)’

- (15) *baba-Te khub jhogra kor-e*  
father-TE very argue do-3P.PRS.HAB  
‘Father argues a lot’

Another characteristic of *-Te* is that it strictly comes with DPs and NPs, as seen in the examples above, and not with AdjP (see 16), VP (see 17), IP (see 18) or CP (see 19).

(16) \* *lal-Te sobuj-Te boigulo*  
 red-TE green-TE book-GULO  
 ‘Red and green balls’

(17) \* *chheleti khachhe-Te ghumachhe-Te*  
 boy-D eat-TE sleep-TE  
 ‘Boys have eaten and slept’

(18) \* *ami jani na raja ash-Te ravi ja-Te kina*  
 I know NEG raja come-TE ravi go-TE whether  
 ‘I do not know whether Raja will come or Ravi will go’

(19) \* *tumi jabe-Te ami ashbo-Te*  
 you go-TE I come-TE  
 ‘You will go, and I will come’

In examples (16-19) above, we see that *-Te* is seen in very restricted contexts and participates in c-selection.

The next characteristic is that it is not found in all types of plural arguments. It is restricted only to the external argument in the case of transitives and unergatives and appears only in internal arguments in unaccusatives. Hence, in a purely nominative-accusative language such as Bangla, it appears only on nominative subjects. For example,

(20) *gari-Te lori-Te dhakka lag-l-o*  
 car-TE lorry-TE clash hurt-PST-3P  
 ‘The car and the lorry clashed’

(21) \* *crane-ti gari-Te lori-Te dhakka lag-l-o/ lag-a-l-o*  
 crane-CL car-TE lorry-TE clash hurt-PST-3P/hurt-PST-CAUS-3P  
 ‘The crane crashed the car and lorry’

In (20), the *-Te* marker is only seen on the nominative subject *gari lori*. However, it is not seen on the accusative DP in (21).

Interestingly, the variant of this marker *-e* is optionally found in restricted domains in a related language, Agartala Bangla (22). Agartala Bangla is another Eastern Indo-Aryan language in contact with the Bodo-Garo language family.

(22) *sasuri-(e) bou-(e) khub kaijja kor-s-e*

Mother-in-law-e          daughter-in-law-e          very          quarrel do-PRS.PERF-3P  
 ‘The mother-in-law and the daughter-in-law quarrelled with each other’

In (22), the *-e* marker exhibits an intense involvement of both the subjects, collectively, in the event. The collective interpretation is unavailable in the absence of the marker *-e*. Such *-e* marked plural subjects are also found in Bangla when the subject ends with a consonantal sound without changing meaning. For example, similar to (22), we see the *-e* in (23).

(23) *saikel-e bus-e dhakka lag-l-o*  
 cycle-e bus-e clash hurt-PST-3P  
 ‘The cycle and the bus clashed’

The distribution of *-Te* is clear from this data set. Let us now turn to the next section, where we assess the nature of the plural *-Te* theoretically.

### 3.2 What type of Plural *-Te* is?

The *-Te* marker plays a role in exhibiting non-singular collective meanings in Bangla. Therefore, it falls under the definition of plurals. Let us now verify what type of plural it is. In other words, we ask whether it is a syntactic ‘head plural’ similar to the English canonical plural or a ‘modifying plural’ (in terms of Wiltschko 2008, 2021). Below, we apply a few of the major diagnostics from Wiltschko (ibid.) to find out the status of *-Te* as a plural.

I. Obligatory plural marking: A head plural is obligatory in nature. The head plural needs to be present obligatorily in the plural context. In the current context, the marker is optional, unlike a head plural.

(24) (a) *dui montri-Te alochona kor-l-o*  
 two minister-TE discuss do-PST-3P  
 ‘Two ministers discussed (this)’

(b) *dui montri alochona kor-l-o*  
 two minister discuss do-PST-3P  
 ‘Two ministers discussed (this)’

II. Obligatory plural agreement: Bangla does not show number agreement.

III. Plural inside compounds: *-Te* can occur inside compounds, unlike inflection (with subtle semantic change).



(25) Compound	-Te marking
(a) <i>raat-din</i> day-night 'Entire day'	<i>raate-Te dine-Te</i> day-TE-night-TE 'At day and at night'
(b) <i>ma-baba</i> mother-father 'Parents'	<i>ma-Te baba-Te</i> mother-TE father-TE 'Mother and father'

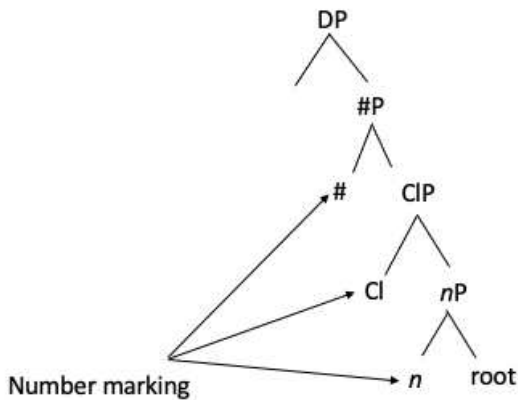
IV. Inside a derivational morphology: Similar to inflectional marker, it does not occur inside derivation

(26) Root> Derivation	-Te marking
(a) <i>din</i> > <i>doinik</i> day> daily	* <i>doinike-Te</i> daily-TE
(b) <i>pita</i> > <i>poitrik</i> father> paternal	* <i>poitrik-Te</i> paternal-TE

Most of these diagnostics thus suggest that -Te is structurally a 'non-inflectional' plural marker.

It is to be noted that Wiltschko (2008, 2021) has suggested that inflectionality and headship may not have one-to-one correspondence. Although inflectional markers are mostly located on syntactic heads, there are no grammar-internal restrictions for non-inflectional plurals not to be on syntactic heads. Further, no restriction exists on the type of syntactic head that plural can merge at- #, D, *n*. According to Wiltschko, the non-canonical plural can merge at any head other than the canonical Num head as seen in (27).

(27)



Let us now analyze if *-Te* is merged at any of these heads.

### 3.2.1 *-Te* as a plural head or a modifying plural?

This section concentrates on analysing whether the *-Te* is merged at any other heads, as suggested in (27), or whether it merges as a modifying plural adjunct to any of these heads. Below, we mention some diagnostics for the same.

#### I. *-Te* not a D-head

There is no selectional restriction on the definiteness based on the appearance of this marker. Its presence does not necessarily render definiteness. In the following example, the DP (‘dogs and cats’) renders a generic meaning and not a definite one.

(28) *...kukur-berale-Te ento kheyeye ne-b-e*  
 dog-cat-TE leftover eat take-FUT-3P  
 ‘Dogs and cats will eat the leftovers’

This suggests that *-Te* is not a D head.

#### II. *-Te* not a Num head

The *-Te* marker is not a number head. It is optional and does not render any obligatory plural agreement.

(29) *montri-(Te) montri-(Te) kolakuli kor-ch-e*  
 minister-TE minister-TE hug do-PROG-3P  
 ‘Two (groups) ministers are hugging each other’

Further, it can co-occur with a productive plural *gulo*, defying the headship quality.

- (30) *kukur-gulo-(Te) khub bhab*  
 dog-GULO-TE very bond  
 ‘(The) dogs have good bonding’

III. -Te not a CI head

Obeying ‘split Num-hypothesis,’ let us check if the *-Te* marker is a classifier (CI) head. The empirical data suggest it is not a regular classifier (31 a vs. b). It does not appear as a single constituent with the numeral. It does not select nominals based on some semantic attributes.

- (31) a. *paanch jon chhele*                      b. *paanch -Te chhele*  
 five CL boy                                      five TE boy  
 ‘Five boys’                                      ‘Five boys’

IV. -Te not a n head

Unlike n-plurals, it is neither lexically attached to any idiosyncratic *n*-roots nor acts as a nominalizer (Wiltschko 2008)

- (32) Root > nominalized form                      *Te marker*
- |                             |                                     |
|-----------------------------|-------------------------------------|
| <i>nach</i> > <i>nacha</i>  | <i>*nach-Te</i> (not a nominalizer) |
| dance (v.) > Dancing (Ger.) |                                     |

The tests show that the *-Te* marker has no head-like qualities. Therefore, we claim it is a modifying plural that (most probably) merges at root *n* as an adjunct with a ‘unifying’ or a ‘together’ meaning (see 33, extracted from Wiltschko 2021:191).

(33)



We consider it as a modifier to *n* -root because for two other reasons. It does not change

the category of the root into a plural when attached. It only attaches as adjuncts. It can co-occur with regular *gulo* plural (34), suggesting it is not a head plural.

- (34) *kukur-gulo-Te*  
 dog-PL-TE  
 ‘Dogs’

We can, therefore, safely conclude that it is a modifying plural merging at *n*. However, the story gets complicated when we look deeper into (i) its interaction with the other plural marker *-raa*, and (ii) the restricted DP structures in which it appears.

#### 4 Associative plural *-raa*

The *-raa* marker is an established plural in Bangla, better known as an associative plural (Chacón 2011, Biswas 2013). This plural marker differs from the other plural *gulo* in its distribution and semantics. The *gulo* can be added to any nominals, but *raa* can only appear with +human and animate nouns (in some contexts). When added to a nominal, while *gulo* yields a regular additive plural (35), the *-raa* yields an associative plural meaning (36). The associative semantics is clearer when attached to a proper noun (37).

- |      |                                      |  |
|------|--------------------------------------|--|
| (35) | a. <i>chhele</i><br>boy<br>‘(A) boy’ | b. <i>Chhele-gulo</i><br>boy-GULO<br>‘Boys’                |
| (36) | a. <i>chhele</i><br>boy<br>‘(A) boy’ | b. <i>Chhele-raa</i><br>boy-RAA<br>‘(Association of )boys’ |

- (37) *nita-raa*  
 nita-RAA  
 ‘Nitas’

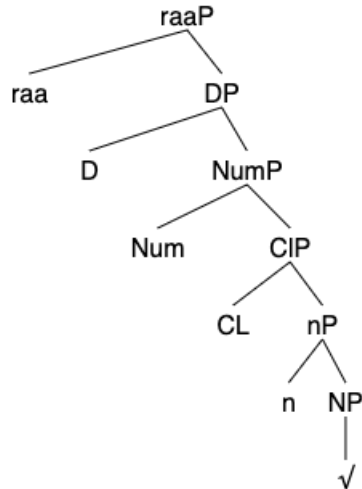
- (38) i. Nita and her associates/friends/family  
 ii. \* Multiple girls named Nita

(37) can have only one interpretation among (38 i- ii). When *-raa* appears with proper nouns, it gives out a collective meaning, as seen in (38i), and not the additive meaning (38ii).

Now, as the data set shows, *-raa* behaves differently from *gulo*, indicating that their structural positions are also different. The literature (Biswas 2013) claims that *-raa* has

some unique properties that suggest that this plural marker is not situated in any of the DP internal positions stated earlier. Instead, it is situated above DP (39).

(39)



Let us look at some of these characteristics with examples.

This associative marker can co-exist with a numeral classifier when the NP is fronted, as seen in (40).

(40) *Chhele-raa*      *du-jon ehete,*  
 boy-RAA      two-CL  
 ‘Two boys’

The *-raa* marker can also be associated with the nominal in the presence of an inverted quantifier. This suggests the NP has moved up to *-raa* across the DP quantifier. In (40), we see that NP has moved out of the universal quantifier, which takes DP complement. The agentive marker on the quantifier makes it more vivid that the *-raa* has to be above DP, where the NP moves crossing the stand-alone quantifier.

(41) *chhele-raa*      *sob-ai*  
 boy-RAA      all-AGT  
 ‘All of the boys’

Further, the proper names generated at D can also attach to *-raa*, strengthening the claim

that *-raa* is above DP.

The position of *-raa* is relevant in the present study because it indicates that the plural can be located in positions beyond what has been discussed in the current analysis, thereby broadening our understanding of the structure of plurals in Bangla. Now, this *-raa* marker has an interesting interaction with *-Te*. The two cannot co-exist. At the end of the previous section, it is shown that the *-Te* can co-occur with the regular plural *gulo*. But this is not true with *-raa*.

- |                               |                           |
|-------------------------------|---------------------------|
| (42) (a) * <i>meye-raa-Te</i> | (b) * <i>kukur-raa-Te</i> |
| girl-RAA-TE                   | dog-RAA-TE                |
| Intended: Girls               | Intended: dogs            |

This complementarity indicates that the *-raa* and *-Te* might have the same structural positions. The argument favoring the *-Te* being above DP gets further justification from the semantics of the *Te*, as detailed below.

## 5 Evidence for *-Te* as a Plural marker with special ‘together’ Semantics

The data discussed till now suggests that the marker *-Te* expresses plural meaning but with a special semantics- non-singular ‘together’ semantics. This is similar to what has been suggested in the seminal work of Heycock & Zamparelli's (2005) about the ‘plurality-forming *and*.’ They show that the conjunction in English has the power not only to exhibit “joint reading” (43 a) but also “split reading” /plurality reading (43 b, c)

- (43) (a) My uncle is short **and** stout  
 (b) My uncle **and** auntie are in love  
 (c) My uncle **and** auntie danced

We propose a similar ‘plurality reading’ arises when *-Te* is attached to the nominals. Let us go through some more examples from Bangla to justify our claim.

- |   |                |             |                 |
|---|----------------|-------------|-----------------|
| (44) (a) <i>Rony-Te</i>                 | <i>Ritu-Te</i> | <i>khub</i> | <i>prem</i>     |
| roni-TE                                 | ritu-TE        | very        | love            |
| ‘Rony ( <b>and</b> ) Ritu are in love’  |                |             |                 |
| (b) <i>Rony-Te</i>                      | <i>Ritu-Te</i> | <i>khub</i> | <i>nach-l-o</i> |
| rony-TE                                 | ritu-TE        | very        | dance-PST-3P    |
| ‘Rony ( <b>and</b> ) Ritu danced a lot’ |                |             |                 |

The ‘together’ meaning comes out well when we use the adverb overtly, and the meaning

remains the same (44b vs. 45)

- (45) *Rony aar Ritu mile khub nach-l-o*  
Rony and ritu together very dance-PST-3P  
'Rony and Ritu dance a lot'

Additionally, the rest of this section checks whether the *-Te* marker imparts collective or distributive interpretation. We check the distribution of the marker with distributive (46-47) and collective predicates (48-49). The choice of verbs is from Syrett and Musolino (2023).

- (46) \*/? *bachha-Te buro-Te ghum-achh-e*  
baby-TE old-men-TE sleep-PROG-3P  
'Babies and old men are sleeping'

- (47) \* *Chhele-Te chhele-Te lomba ho-ech-e*  
boy-TE boy-TE tall be-PERF-3P  
'Boys have become tall'

- (48) *bachha-Te bachha-Te dhakkadhakki kor-ch-e*  
boy-TE boy-TE push do-PROG-3P  
'Boys are pushing each other'

- (49) *Chhele-Te chhele-Te jaega-ti ghir-e fel-ech-e*  
boy-TE boy-TE place-D surround-NF finish-PERF-3P  
'Boys have surrounded the place'

- (50) *Ama-Te toma-Te dekha ho-ech-il-o*  
I-TE you-TE meet be-PERF-PST-3P  
'You and I met'

The data suggest that the distributive predicates (46-47) do not allow *-Te* marker, whereas collective predicates (48-50) do. More examples of the presence of collective reading and absence of distributive reading in relation to *-Te* can be seen below:

- (51) *raja ar rani bera-te g-ech-e*  
Raja and rani visit-NF go-PERF-3P  
'Raja and Rani went for a visit'

(51) can have both interpretations (52-53).

(52) *raja purulia-e bera-te g-ech-e ar rani*  
 raja purulia-LOC visit-NF go-PERF-3P and rani  
*malda-e bera-te g-ech-e*  
 malda-LOC visit-NF go-PERF-3P  
 ‘Raja went to Purulia and Rani went to Malda’

(53) *raja ar rani eksathe bera-te g-ech-e*  
 raja and rani together visit-NF go-PERF-3P  
 ‘Raja and Rani went (somewhere) together for a visit’

(54) *raja-Te rani-Te bera-te gece*  
 Raja-TE rani-TE visit-NF go-PERF-3P  
 ‘Raja and Rani went for a visit’

However, (54) can have only one interpretation of (53). Therefore, apart from imparting plurality, it gives a meaning of collectivity. This collective semantics is missing when canonical Bangla conjunctions are used.

Now, we test whether any quantificational restrictions are found with *-Te*. From the evidence below, we see that the ‘each’ meaning is disallowed in the case of ‘*-Te*’ (55) unless we attach an overt quantifier *protyek*, as seen in (56).

(55) *chhele-Te meye-Te kaj-ta kor-ech-e*  
 boy-TE girl-TE work-D do-PERF-3P  
 ‘Boys and girls have done the work’

(56) *protyek chhele-Te meye-Te kaj-ta kor-ech-e*  
 each boy-TE girl-TE work-D do-PERF-3P  
 ‘Each boy and each girl has done the work’

*-Te* also differs from the genitive counterpart in relation to the choice of verb/ light verb. For example, in (57), the genitive DP cannot appear with a verb like *koreche*, as seen in (55-56). Additionally, the sense of direct involvement of the DPs in the work is also missing.



- (57) *chhele-r*      *meye-r*      *kaj-ta*      *ho-ech-e*  
 boy-GEN      girl-GEN      work-D      do-PERF-3P  
 ‘Boys and girls have done the work’

Moreover, the marker *-Te* is also mutually exclusive with *-raa* marker in every context as already mentioned in section 4. Let us take another example:

- (58) \* *chhele-ra-Te*    *chhele-ra-Te*  
 Boy-RAA-TE    boy-RAA-TE

We see an overlap between the *-raa* and *-Te* marker in Bangla where both are not allowed together. This is a crucial diagnostic of the study. The following sub-section elaborates on this phenomenon.

## 6 Structure of the *-Te*

The above sections clearly show that *-Te* exclusively comes when there are two nominals in the structure. In other words, the plural sense of *-Te* is seen when there are strictly two DP. Therefore, it becomes crucial to check structures with two DPs. Co-ordinate structures are one such construction where two DPs come together.

Let us discuss the structure of the co-ordinate DPs as suggested in the literature. According to some accounts (Munn 1987, Larson 1900), the co-ordinate phrase (&P) is headed by the co-ordinate phrase &P.

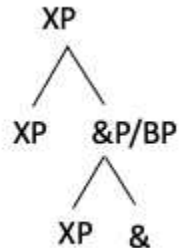
- (59)



On the contrary, some other accounts, such as Munn (1992, 1993), Boskovic and Franks (2000) among others, propose that the whole co-ordinate complex is not headed by the

co-ordinate phrase as seen in (60).

(60)

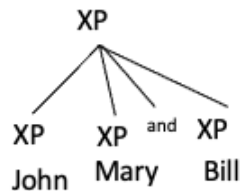


However, Bangla has dedicated co-ordinate markers- *ebong*, *o*, and *aar*. The *-Te* marker is not one of them. This is quite evident from the examples where *-Te* is attached to both the DPs individually in the phrase. Additionally *-Te* is inflectional (contra to derivational marker). Therefore, none of the structures (59-60) fit the bill.

We claim that although *-Te* involves two nominals, it has a different structure from what has been suggested for the co-ordinate DPs. *-Te* is a part of a flatter structure as commonly seen in the case of multiple adjuncts ( cf. Ross 1967 , Jackendoff 1977) a.o. See (61-62) for illustration.

(61)

- a. John, Mary and Bill
- b.



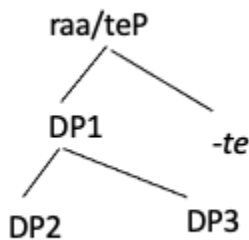
A piece of evidence for our claim comes from the structure itself. In a construction like (62), the first DP (*baba*) cannot c-command the anaphor (*tar*), suggesting the absence of hierarchy needed for the c-command relation. However, such relation is fulfilled in a co-ordinate DP (63).

(62) \**Baba<sub>i</sub>-Te tar<sub>i</sub> meye-Te lorai kor-ch-e*  
 father-TE his daughter-TE fight do-PROG-3P  
 ‘Father and his daughter are fighting’

(63) *Baba<sub>i</sub> ar tar<sub>i</sub> meye lorai kor-ch-e*  
 father and his daughter fight do-PROG-3P  
 ‘Father and his daughter are fighting’

Now, let us see how this flatter structure of *-Te* fits in the Bangla DP structure. We have already suggested in sections 4 and 5 that *raa* and *-Te* are in complementary distribution with a similar kind of associative/collective plural meaning. Therefore, we claim that *-Te* and *-raa* occupy the same position above DP. Structurally, see 64.

(64)



The analysis and the structure show that *-Te* behaves like Bangla associative plural of *-raa*. Both of them give associative meaning. However, it is to be noted that *-Te* can only come exclusively with two DPs unlike *-raa*. The associative meaning of *-Te* is connected to collective participation of both the DPs in contrast to *-raa* which is more about group.

## 7 Conclusion

The present study suggests that the classifier language Bangla has a plural marker *-Te*, along with the established markers *-raa* and *gulo*. The support for the claim also comes from the diachronic literature (Chatterji, 1926, p.822) that suggests that the form *-Te* is derived in Bangla from the Old Indo-Aryan instrumental plural *-tehi*.

It further suggests that this marker is similar to the associative plural marker *-raa*, yielding the meaning of an association. It differs from *-raa* in that it is restricted to two DPs. This marker also gives out the nuanced semantics of a ‘togetherness’ implying that the two arguments are actively and collectively participating in the action/event referred to by the verb. The study provides a novel claim since this marker has not yet been

studied as a plural marker in the literature. It thereby adds to the list of current claims in the literature that propose the presence of a variety of plurals in classifier languages.

## References

- Bhattacharya, Tanmoy. 1999. The structure of the Bangla DP. Unpublished PhD dissertation, UCL, London, UK.
- Bhattacharya, Tanmoy. 2000. In search of the vague 'One.' In Proceedings of ConSOLE 7. 33-48
- Bhattacharya, Tanmoy. 2001. Numeral/Quantifier–Classifier as a Complex Head. In Henk van Riemsdijk and Norbert Corver (eds.), *Semi-lexical Heads*. Mouton de Gruyter.
- Biswas, Priyanka. 2013. Plurality in a classifier language: Two types of plurals in Bangla. Proceedings of Generative Linguists of the Old World in Asia (GLOW in Asia), 1-14.
- Bošković, Željko. and Steven. Franks. 2000. 'Across-the-Board Movement and LF', *Syntax* 3: 107-128.
- Chacón, Dustin. A. 2011. Bangla and company: the distribution of associative plurals in Bangla, Japanese, and Mandarin Chinese. *Handout in Formal Approaches to South Asian Languages*.
- Dasgupta, Probal. 1983. On the Bangla classifier. *Ta/, its penumbra and definiteness Indian Linguistics*44, 11, 26.
- 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.
- Dutta, Chandni., Kumari, Preeti., Chandra, Pritha. 2023. Plural Inflection in Classifier Language. IATL 36 Conference, *MIT Working Papers*.
- Heycock, Caroline., & Zamparelli, Roberto. 2005. Friends and colleagues: Plurality, coordination, and the structure of DP. *Natural Language Semantics*, 13, 201-270.
- Larson, Richard. K. (1990). Double objects revisited: Reply to Jackendoff. *Linguistic Inquiry* 21(4), 589-632.
- Munn, Alan. 1987. Coordination and X-bar theory. In McGill working papers in Linguistics 4(1):121-140. Department of Linguistics, McGill University, Montreal.
- Munn, Alan. 1992. A null operator analysis of ATB gaps. *The Linguistic Review* 9:1-26.
- Munn, Alan. 1993. Topics in the Syntax and Semantics of coordinate structures. Doctoral dissertation, University of Maryland, College Park.

- Syrett, Kristen., & Musolino, Julien. 2013. Collectivity, distributivity, and the interpretation of plural numerical expressions in child and adult language. *Language Acquisition*, 20(4), 259-291.
- Wiltschko, Martina. 2008. The syntax of non-inflectional plural marking. *Natural Language & Linguistic Theory*, 26(3), 639-694.
- Wiltschko, Martina. 2021. *The grammar of interactional language*. Cambridge University Press.

# The internal syntax of *iva*- clauses in Vedic Sanskrit

MARK HALE, *Concordia University, Montréal*

MADELYN KISSOCK, *Concordia University, Montréal*

## ABSTRACT

In this paper we argue that a proper characterization of the synchronic internal syntax of clauses of comparison in Vedic Sanskrit can clarify issues of interpretation. We also briefly explore aspects of the diachrony of these structures.

## 1 Introduction

In Hale & Kissock (2021), we discussed the relationship between ‘comparison’ clauses, typically marked by *iva* or *ná* in the language of the mantras of the Vedas, and the matrix clauses within which they were embedded. That paper demonstrates that when the matrix clause is negated, the scope of the negation vis-à-vis the embedded ‘like’-clause, introduced by *iva* or *ná*, is a function of the structural position of the comparison clause.

This discovery has serious implications for two widely-held, but in our view incorrect, views on the syntax of the language of the Vedic mantras:

- (1a) that the word order is ‘free,’ and/or
- (1b) that the word order is metrically conditioned.

The claim in (1a) cannot be true if, as Hale & Kissock (2021) demonstrated, the positioning of elements directly impacts standard syntactic concepts such as ‘scope of negation’. This is inconsistent with ‘flat’ phrase structure which one would need to posit to justify labelling the word order as ‘free’. Likewise, the claim in (1b) cannot be true given our demonstration of the scope facts, which show that scope interpretation is based on structural position. Such conditioning would be inconsistent with metrically-determined positioning of the relevant elements, since the meter does not have access to information about ‘scope of negation,’ and thus cannot appropriately place elements which need to have the relevant scope relations.

We conclude, then, that the language of the Vedic mantras displays the properties of a natural human language, with hierarchical syntax of the familiar type.<sup>1</sup>

## 2 *iva* Placement

In this brief paper, we would like to turn to one of the many issues which arise concerning the *internal* structure of the clause of comparison (we’ll just call them *iva*-clauses from

---

<sup>1</sup>Of course, the texts were composed over a period of time, and thus actually represent texts produced by mildly divergent grammars — i.e., different ‘dialects,’ if you will. This fact appears irrelevant to the present investigation.

now on, though comparison clauses marked by *ná* should be considered as included in the discussion). As a leaping-off point, consider the following recently edited passage from the Paippalāda recension of the Atharva Veda:

(1) AVP 6.6.6ab

*sindhuprajāno*      *madhugho*      // *aśva*      *iva nīyate*  
 Sindhu-born-NOMSG Madhugha-NOMSG // horse-NOMSG **like** is lead  
*janān*      *anu*  
 men-ACCPL among

‘Madhugha is led like a Sindhu-born horse among men’. (transl. of Griffiths 2009:84)

As you can see from the translation, Griffiths takes the adjective *sindhuprajāna-* ‘Sindhu-born, originating in Sindhu’ as modifying ‘horse’ (part of the *iva*-clause), rather than as modifying the plant-name *madhugha-*. He notes (84): ‘I take this adjective [*sindhuprajāno*—mh/mk] with *aśva-* in the next pāda, in view of BĀU 6.1.13 *mahāsuḥayaḥ saindhavaḥ*, ŚāṅkhĀ 9.7 *saindhavaḥ suḥayaḥ* ‘a (great) prize-stallion from the Indus region’...’

Of course, we must point out that the fact that there are prize stallions which come from Sindhu does not entail that there are not also plants which come from that same region. For this reason, we do not consider Griffiths’ argument for his interpretation compelling, though, up to now, nothing precludes that interpretation.

Since the *mad(h)ugha-* is a plant<sup>2</sup> and *aśva-* is a horse, and both plants and horses can be *saindhava-*, can we tell, from our study of the structure of *iva*-clauses, what *sindhuprajāna-* is modifying?

The way we will approach answering this question here is by asking a structural question: is *iva* properly positioned (a) under the interpretation which takes *sindhuprajāna-* as a modifier of *madhugha-* or (b) under the interpretation which takes *sindhuprajāna-* as a modifier of *aśva-*, or (c) under both interpretations? Relatedly, we will ask whether the kind of discontinuity seen within the *iva* clause if we take *sindhuprajāna-* as a modifier of *aśva-* is in fact permissible, or are there some relevant constraints on such things?

### 3 *iva* Placement

MacDonell (1916: §180, s.v. *iva*) describes the most frequent usage of *iva* in the Rigveda (and the only one that will interest us) as:<sup>3</sup>

<sup>2</sup>Perhaps < *madhu-dugha-* or even *madhu-dogha-*, ‘giving honey as milk’ *vel sim.*, both RV+, via haplology, as already suggested by Brugman (1897:860). The explanatory possibilities are quite broad for each of the attested forms (with and without aspiration on the first non-nasal stop), including *madhudugha-* > \**madhugha-* while Grassmann’s Law was still active and *madhudugha-* > *madugha-* directly, via anticipatory haplology, for the unaspirated form, and, for the aspirated form, either of those developments plus folk-etymological restoration of the aspiration, or *madhudugha-* > *madhugha-* directly *after* Grassmann ceased to be active.

<sup>3</sup>This is virtually an English translation of Grassmann’s entry, s.v. *iva*.

**iva** . . . means *as if, as, like* in abbreviated similes in apposition, never introducing a clause like **yáthā**. It follows the word with which comparison is made; if the comparison consists of several words, the particle generally follows the first, less commonly the second.

The vast, vast majority of instances of *iva* involve simple ‘second position’ placement of the particle. In the interest of time, we will not present supporting evidence here. There is a wealth of data both familiar and readily accessible to those interested.

Nevertheless, the ‘vast, vast majority’ is not *all*, and, as the MacDonell (1916) quote indicates, we do in fact find occasional instances of *iva* in what we might term a ‘delayed’ position in the mantra texts. This can be seen from examples such as the following:

(2) RV 2.5.3cd

*pári víśvāni káviyā // nemís cakrám iva-*  
 around-PV all-ACCPL wisdoms-ACCPL // **rim-NOMSG wheel-ACCSG like**  
*-abhavat*  
 he comes to be

‘He surrounds (=comes to be around) all wisdoms, like the rim (surrounds) the wheel’.

(3) RV 9.50.1ab

*út te súśmāsa īrate // síndhor ūrmér*  
 upwards-PV your-CL strengths-NOMPL move // **river-ABL SG waves-GEN SG**  
***iva svanáḥ***  
**like rush-NOMSG**

‘up move your strengths, like the rush of waves (move up) from the river’

MacDonell pretty clearly implies, with the statement ‘less commonly the second,’ that *iva* may not occur *later* in the string making up the comparison than after the second word. As can be seen from these examples, that generalization does not hold:

(4) RV 1.116.15a

***parṇá mṛgásya patáror iva- -ārábha // úd***  
**wings-ACCPL wild-animal-GENSG flying-GENSG like to seize // up-PV**  
*aśvinā ūhathuḥ śrómatāya kám*  
 Aśvin-VOC DU you two conveyed for obedience-DATSG PostP

‘You have conveyed (him) upwards for obedience like (you convey upwards) the feathers of the flying wild animal for seizing’

Unfortunately, since the comparison in (1) would “consist of several words” if *sind-**huprajāna-* modifies *aśva-*, the MacDonell observation would allow *iva* to follow *aśva-* in such a case. And, of course, if it does not, then *iva* would be in ‘second position’ after *aśva-*—the most common pattern. So it isn’t clear that the distribution of *iva* can be helpful in this case. What about considerations of *discontinuity*?



#### 4 Discontinuity in *iva* Clauses

If we are to take the elements of the comparison in (1) to be *sindhuprajāno... aśva iva* ... ‘like a Sindhu-born horse,’ we must recognize that the elements which make up the terms of comparison can be *discontiuous* — in this case broken up by the intervention of the subject *madhugha*. In general, as can be seen from examining the data we have already cited, the elements of the term of comparison form an uninterrupted constituent. So we can ask the question: do the elements of a comparison *have* to be continuous, or can they be interrupted (as required for Griffiths’ interpretation of (1))?

In fact, discontinuity amongst the elements of the comparison clause is attested in mantra texts. We will cite these examples somewhat more fully, since we are going to need them in the discussion which follows, and since, while attested, they are somewhat rare and can thus be hard to find.

(5) RV 1.116.15a

*carítram hí vér iva- -āchedi parṇám*  
leg-NOMSG because **bird-GENSG like** is broken **wing-NOMSG**

‘For the leg is broken, like the wing of a bird’.

The comparison clause is *vér iva... parṇám* ‘like the wing of a bird’. The verb form *āchedi* ‘interrupts’ the elements of the comparison clause.

(6) RV 5.60.1c

*ráthair iva prá bhare vājayádbhiḥ*  
**chariot-INSTRPL like** forth-PV I bear myself **prize-seeking-INSTRPL**

‘I hasten forth like (one hastens forth) with prize-seeking chariots’.

Here the elements in the comparison are *ráthair iva... vājayádbhiḥ* ‘like with prize-winning chariots,’ and they are ‘interrupted’ by the verb form *prá bhare* ‘I bear myself’ = ‘I hasten forth’.

(7) RV 6.75.4ab

*té ācárantī sámanā- -iva*  
these-NOMDU wandering-hither-NOMDU gathering-ACCPL like  
*yóṣā // mātā- -iva putrám bibhṛtām*  
maiden-NOMSG // **mother-NOMSG like son-ACCSG** bear-3DUIMPV  
*upásthe*  
**lap-LOC SG**

‘let these two wandering ones bear (it), like a maiden (bears it) to gatherings, like a mother (bears) a son in her lap’

In this example, the comparison of interest (the second one) consists of the elements *mātā-iva putrām... upāsthe* ‘like a mother a son in her lap’. It is interrupted by the third person dual imperative *bibhṛtām*.

(8) RV 7.103.5ab

*yád eṣām anyó anyásya vācam // śāktásya-*  
 when of-them one-NOMSG another-GENSG word-ACCSG // **teacher-GENSG**  
**-iva vádati śíkṣamāṇaḥ**  
**like speaks learning-one-NOMSG**

‘when one of them speaks the word of the other, like the learning one (speaks the word) of the teacher’

Here we once again see that the elements in the comparison, *śāktásya-iva śíkṣamāṇaḥ* ‘like the learning one (the word) of the teacher,’ are interrupted by the finite verb form *vádati* ‘speaks’.

(9) AVP 5.25.2c

*sénā- -iva- eṣi tvíṣīmatī*  
**army-NOMSG like you go impetuous-NOMSG**

‘you go like an impetuous army (goes)’ (after Lubotsky 2002)

Finally, in this example we have a comparison clause *sénā-iva... tvíṣīmatī* ‘like an impetuous army,’ which is interrupted by the second person singular finite verb *eṣi*.

These examples provide a clear characterization of the context within which ‘discontinuity’ is permitted in the elements which make up the comparison: the only open-class lexical item<sup>4</sup> which can ‘interrupt’ an *iva* clause is the verb (which may be accompanied by its preverb, unsurprisingly).

We can also see something else from these examples: when the *iva* clause displays ‘discontinuity,’ the *iva* seems to invariably accompany the first element. Thus although in general, given a multi-element *iva* clause, ‘postponement’ of *iva* is possible, it seems that this is *not* possible in DISCONTINUOUS *iva* clauses.

The interpretation of (1) in which *sindhuprajāna-* is interpreted as a modifier of *aśva-* would violate both of these generalizations. We would have a discontinuous *iva*-clause whose discontinuity arose via the interposition of an open-class element *madhughas* (pre-sandhi) which is not a verb and, in spite of the discontinuity, we would have a ‘postponed’ *iva*. For these syntactic reasons, therefore, *sindhuprajāna-* should be interpreted as a modifier of *madhugha-* and, when seeking to identify this plant, we should look for one for which ‘produced in the river (area)’ or ‘in the area of the Sindhu’ is a sensible attribute.

<sup>4</sup>Occasionally, a second position clitic intervenes, which is obviously to be attributed to a quite distinct set of factors.

## 5 A Possible Diachrony of *iva* Clauses

We now turn briefly to the question of how *iva* clauses may have come to have their somewhat peculiar properties. Vine (1978) has presented a plausible story about the similar-seeming *ná* clauses, but it seems quite implausible that that account would work for *iva*, which shows no traces of the negative semantics which is required for his explanation to work here. Instead, *iva* seems like the clitic version of *yáthā* (like *ca* is the clitic version of *utá*), but differs significantly from *yáthā* in not allowing a verb to occur in its ‘clause’. Could these two peculiar properties of *iva* clauses, — namely, that it permits ‘interruption’ only by verb forms and that it cannot itself contain a verb form, — be plausibly related?

In examples with an ‘interrupting’ verb it is on some occasions clear that that verb must be from the main clause (see exx. 6, 7, 9 above). On other occasions, however, the verb works equally well as either an *iva*-clause predicate or as the main-clause predicate (see exx. 4 and 8 above). The latter is most common when one is dealing with third person arguments in both clauses, as is quite frequently the case. Never in Vedic mantra texts does the verb agree only with the subject of the *iva*-clause. Thus, the verb can *always* be taken as representing the matrix finite verb.

Imagine that, originally, like *yáthā*, clauses introduced by *iva* could have their own verb, and that, since that verb was often identical to the main clause predicate, it was frequently gapped. When not gapped, and appropriate to either clause, it would be ambiguous as to whether it was the main clause predicate, or the *iva* clause predicate. If a speaker wrongly concluded that it was the main clause verb, (s)he would then need to allow main clause verbs to intervene amongst the elements of the *iva* clause.

This speaker would then, of course, begin to insert unambiguously main clause verbs into *iva* clauses (but only verbs). When such examples were combined with the many ambiguous ones, plus the ones where the verb of the subordinate *iva* clause was gapped, one can see, perhaps, how the current situation came into being.

## 6 Diachrony and Synchrony

We know that diachrony can give rise to idiosyncratic *morphological* properties: *keep* : *kept*, *go* : *went*, *foot* : *feet*, etc. And diachrony can give rise to ‘unnatural’ phonological rules, weird gaps in segment inventories, etc. In the case of morphological oddities, we have made allowance in our formal models for ‘listed’ (i.e., stored, not generated) forms, as seems required. In the case of phonology, we have allowed a computational component which licenses a relatively unconstrained and idiosyncratic ‘rule’ system.

But, if something like our diachronic story about *iva*-clauses containing matrix verbs is correct, how are we to account for this *synchronically*? Does the syntax license such oddities? Is there, as Chomsky has often said (but never elaborated on much) both a ‘core’ and a ‘periphery’ in the syntactic computational system? What do the two systems, if they both exist, look like and how do they interact? How do we constrain them? We leave these questions for future research, or as an exercise to the interested reader.

## 7 Conclusions

In this brief contribution we have presented arguments from the internal syntax of *iva*-clauses for the following points:

- The only open-class lexical item that can interrupt an *iva*-clause is the matrix verb.
- The plant *mad(h)ugha-* is from a river basin (or, less plausibly, is a river itself).
- Understanding the syntax of *iva* clauses can be important for discovering the correct semantics for lexical items.
- There is still much we don't know about *iva*-clauses, but which we hope to get clear about relatively soon.

More importantly, we can see how critical a proper characterization of the *synchronic* syntactic details of a structure is to developing a plausible account for the diachrony of that structure. There is nothing surprising in this — it would be strange indeed if we could develop a compelling story of how a structure came to be what it synchronically is without first knowing, in fact, precisely *what* it is, synchronically — however, it is our feeling that this fundamental fact, known since Saussure as the ‘primacy of synchronic linguistics,’ is overlooked by many of those working on diachronic syntax.

## References

- Bergaigne, Abel. 1887. La syntaxe des comparaisons védiques. In *Mélanges Renier*, 75–101. Paris: Vieweg.
- Bhattacharya, Dipak. 1997. *The Paippalāda-Saṃhitā of the Atharvaveda: Volume one, consisting of the first fifteen Kāṇḍas*. Kolkata: The Asiatic Society.
- Bhattacharya, Dipak. 2008. *The Paippalāda-Saṃhitā of the Atharvaveda: Volume two, consisting of the sixteenth Kāṇḍa*. Kolkata: The Asiatic Society.
- Brugmann, Karl. 1897. *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen, Volume 1, 2nd edition*. Strassburg: Trübner.
- Griffiths, Arlo. 2009. *The Paippalādasamhitā of the Atharvaveda, Kāṇḍas 6 and 7: a new edition with translation and commentary*. Groningen Oriental Studies XXII. Groningen, NL: Egbert Forsten.
- Hale, Mark & Madelyn Kissock. 2021. On the syntax of comparative clauses in vedic sanskrit: ...like someone eating the foam off the water. In G. Sharma & J. Lowe (eds.), *Trends in South Asian Linguistics*, 1–21. Berlin: de Gruyter-Mouton.
- Jamison, Stephanie. 1982. Case disharmony in Ṛgvedic similes. *Indo-Iranian Journal* 24. 251–271.
- Lubotsky, Alexander. 2002. *Atharvaveda-Paippalāda, Kāṇḍa 5: text, translation, commentary*. Columbia, Missouri: South Asia Books.

- MacDonell, Arthur A. 1916. *A Vedic Grammar for Students*. Oxford, UK: Oxford University Press.
- Pinault, Georges-Jean. 2004. On the usages of the particle *iva* in the Ṛgvedic hymns. In Arlo Griffiths & Jan Houben (eds.), *The Vedas: texts, language and ritual*, Groningen: Forsten.
- Schrapel, Dieter. 1970. *Untersuchung der Partikel iva und anderer lexikalisch-syntaktischer Probleme der vedischen Prosa nebst zahlreichen Textemendationen und der kritischen Übersetzung von Jaiminīya-Brāhmaṇa 2,371-373 (Gavāmayana i)*. Marburg: University of Marburg dissertation.
- Vine, Brent. 1978. On the metrics and origin of rig-vedic *ná* 'like, as'. *Indo-Iranian Journal* 20. 171–193.

# Relative clause constructions in New Indo-Aryan languages: Hierarchies of macro roles

SAKURA ISHIKAWA, *Tokyo University of Foreign Studies*  
SHIGEKI YOSHIDA, *University of Tokyo*

## ABSTRACT

Since the seminal work by Keenan & Comrie (1977), typological studies have shown that languages vary with respect to the range of arguments that can be relativized on. In this study, we systematically examine what can be relativized in five New Indo-Aryan (NIA) languages: Hindi-Urdu, Nepali, Early Nepali, Sinhala, and Bengali. Inspired by typological studies on relative clauses, we conducted our examination using a novel systematic methodology. First, we examined both headless and headed relative clauses. Second, we examined relativization on arguments for each of the macro roles S, A, P, T, and R. Lastly, we examined every participial strategy for relative clause constructions when a language had different participles for tense or aspect. Our investigation showed that there are both similarities and differences in the relativizability of NPs in relative clause constructions in the five NIA languages examined. On the one hand, in each language examined, arguments of the same range of macro roles can be relativized on in both headed and headless relative clauses. On the other hand, the five languages differ as to which macro roles can be relativized on. Based on this difference of the relativizability of NPs and our novel methodology, we propose hierarchies of relativizability for these NIA languages. The hierarchies are the onset-oriented hierarchy  $\{S\} > \{A\} > \{P, T, R\}$  for relative clause constructions by imperfective/nonpast participles and the termination-oriented hierarchy  $\{S, P, T\} > \{A\} > \{R\}$  for those by perfective/past participles. We explained these hierarchies in terms of viewpoint, localist metaphor, and a metonymy relationship.

## 1 Introduction

Relative clauses have been a major area of interest in linguistic typology, due in part to the fact that languages vary with respect to the range of arguments that can be relativized on. In the literature on the typology of relative clauses, Keenan & Comrie (1977) proposed the NP Accessibility Hierarchy to capture the universality and diversity of relative clauses in languages. They claim the following implicational hierarchy for the relativizability of NPs.

- (1) The NP Accessibility Hierarchy (Keenan & Comrie 1977)  
subject > direct object > indirect object > oblique > genitive > object of comparison

The hierarchy in (1) shows that the subject can always be relativized, and that if a strategy in a language is available for one grammatical relation, it is also available for grammatical relations higher up on the hierarchy.

Relative clauses in New Indo-Aryan (henceforth NIA) languages seemingly exhibit counterexamples to the NP Accessibility Hierarchy. It has been reported that some relative clauses in these languages do not follow the hierarchy in (1) (Subbārāo 2012). For example, in Bengali, NPs of direct object and oblique (e.g., locative) can be relativized, but indirect object cannot

(Faquire 2014; Subbārāo 2012: 331), as shown in (2).

(2) Bengali

- a. [*amar dek<sup>h</sup>-a*] *lok=ti*  
 1SG.GEN see-PTCP person=CLF  
 ‘The person whom I saw’ (Faquire 2014: 26)<sup>1</sup>
- b. \*[*amar tʃi<sup>h</sup>i de-wa*] *lok=ti*  
 1SG.GEN letter give-PTCP person=CLF  
 ‘The person to whom I send a letter’ (Faquire 2014: 26)
- c. [*alta pɔɾ-a*] *pa*  
 alta wear-PTCP foot  
 ‘The foot on which alta dye is worn’ (Subbārāo 2012: 332)

(2a) shows the relativization of the direct object *lok* ‘person’, and (2c) shows the relativization of the oblique *pa* ‘foot’. As shown in (2b), the indirect object *lok* ‘person’ cannot be relativized. The examples in (2) deviate from the predicted pattern outlined in the NP Accessibility Hierarchy. Since indirect objects fall between direct objects and obliques in the hierarchy, if an oblique can be relativized in a language, it is predicted that an indirect object can also be relativized. The Bengali data in (2) do not follow this prediction.

Situations like the above that go against the predictions of the NP Accessibility Hierarchy in NIA languages are found only in participial strategies for relative clause constructions. Most NIA languages have two strategies for relative clauses: participial and relative-correlative strategies. Relative-correlative strategies have little restriction on relativizability (Subbārāo 2012: 271). In this paper, we focus on participial strategies for relative clause constructions.

This study aims to provide a systematic survey of participial strategies in five NIA languages. We investigated both headless and headed relative clauses created by participles for each of the macro roles S, A, P, T, and R in Hindi-Urdu, Nepali, Early Nepali, Sinhala, and Bengali.

This study is systematic in three respects. First, it examines relative clauses both with and without a head NP. A relative clause with a head NP is illustrated in (3).

(3) Nepali

- [*pok<sup>h</sup>ara dza-ne*] *bʌs k<sup>h</sup>ʌhã pa-i-nt<sup>h</sup>a?*  
 Pokhara go-IMPF.PTCP bus where get-PASS-3SG  
 ‘Where can I get a bus going to Pokhara?’ (Matthews 1998: 160)

In the example above, the relative clause *pok<sup>h</sup>ara dza-ne* ‘going to Pokhara’ modifies the head NP *bʌs* ‘bus’. This type of relative clauses is called a headed relative clause. Some languages have relative clauses that do not modify nouns or pronouns (Dryer 2007: 197). For example, in Nepali, a participle can occur without modifying a head noun.

(4) Nepali

- [*b<sup>h</sup>an-eko na-man-ne*]=*lai* *sallah di-erʌ ke kam?*  
 say-PFV.PTCP NEG-listen-IPFV.PTCP=DAT advice give-CVB what work  
 ‘What’s the use of giving advice to someone who does not listen to what you say?’  
 (Matthews 1998: 171)

<sup>1</sup>We altered the glossing of examples from other studies if necessary throughout this paper.

In (4), the relative clause formed by the participial phrase *bʰan-eko na-man-ne* ‘one who does not listen to what you say’ functions as a noun phrase without modifying a noun. This type of relative clauses is called a headless relative clause, as opposed to a headed relative clause (Dryer 2007: 197). In recent typological studies, both headed and headless relative clauses have been considered equally important. Shibatani Masayoshi (Shibatani 2019 among others) argues that relative clauses should be reanalyzed as nominalizations, and that so-called headed and headless relative clauses are the two uses of nominalizations. Except for Nepali (Wallace 1985; Paudyal 2010), the relativizability of the gapped argument in a headless relative clause, or nominalization, has not often been described. In the literature, Nepali data seem to show that the relativizability of an NP can differ between headed and headless relative clauses. Wallace (1985) shows that only the subject can be relativized in headless relative clauses (‘nominalizations’ in his terminology), while Paudyal (2010) provides data for headed relative clauses whose head NP is something other than the subject. This study examines both headed and headless relative clauses when a language has both.

Second, this study is systematic because it examines relativization for each of the macro roles S, A, P, T, and R. Here we deviate from Keenan & Comrie (1977). Their discussion is based on grammatical relations like subject and object. Describing relative clauses based on macro roles enables us to accomplish more accurate generalizations, as some grammatical relations cover more than one macro role. For example, subject is the syntactic generalization over S and A, and direct object is the syntactic generalization over P and T. However, the macro roles covered by a grammatical relation do not necessarily behave similarly especially in a language with ergativity. For example, in the ergative language Central Alaskan Yup’ik, S and P can be relativized, while A cannot (Shibatani 2021). In such a situation, we cannot syntactically generalize S and A as subjects in relativization because they behave differently syntactically. Similarly, a number of NIA languages, including Hindi-Urdu and Nepali, show ergativity to varying degrees (Verbeke 2013). In describing these languages, it is especially necessary to focus on macro roles rather than on grammatical relations like subject and direct object. In previous studies on relative clauses in NIA languages, however, the difference in relativizability based on macro roles has not often been described. More focus has been put on grammatical relations like subject and object. For example, it is repeatedly mentioned that the Hindi-Urdu imperfective participial strategy is available for subjects (see, for example, Kachru (2006)), but it is not clearly shown whether this strategy is available for both S and A. In order to describe relativizability in NIA languages, macro roles must be investigated separately.

Third, this study systematically examines every participial strategy for relative clause constructions when a language has different participles depending on tense or aspect. Among the five languages investigated, Hindi-Urdu, Nepali, Early Nepali, and Sinhala have two participles: perfective or past participle and imperfective or nonpast participle. A large number of NIA languages show split ergativity in their marking of argument or agreement depending on tense and aspect (Abbi 2001: 29). For example, in Nepali, the A argument is marked by the ergative marker *=le* in the simple past tense (Matthews 1998: 94). Relativizability can also be different depending on tense or aspect. Thus, we investigate both forms of participles when a language has two participial strategies.

Through the systematic investigation described in this study, we are able to offer generalizations about relative clause constructions in the NIA languages examined. Our investigation shows that there are both similarities and differences between these languages. On



the one hand, none of the five languages examined shows any difference of relativizability between headed and headless relative clauses. On the other hand, the five languages differ as to which macro roles can be relativized. We propose hierarchies of relativizability for NIA languages based on our results, namely the onset-oriented and termination-oriented hierarchies. We then present an explanation for these hierarchies in terms of viewpoint, localist metaphor, and a metonymy relationship.

This paper is organized as follows. In Section 2, we introduce the languages examined and the methodology we used for testing grammaticality. In Section 3, we discuss the geographical locations of the languages and summarize what is known about these languages from previous studies. In Section 4, we provide the results of our investigation. In Section 5, we discuss the similarities and differences between the four NIA languages and propose hierarchies of relativizability and an explanation for these hierarchies. In Section 6, we conclude the paper.

## 2 Methodology

To investigate the behavior of the participle strategies of relative clause constructions in NIA languages, we selected five NIA languages: Hindi-Urdu, Nepali, Early Nepali, Sinhala, and Bengali. By Early Nepali, we mean Nepali of the 18<sup>th</sup> to 19<sup>th</sup> centuries. The data pertaining to Early Nepali was sourced from Wallace (1985). As for the remaining four languages, we utilized the data from the literature and from the stories, and we also collected data through direct elicitation from our informants. To elicit data, we conducted grammatical judgment tests with a single informant for each language. Table 1 shows the basic information on our informants.

	Gender	Year of birth	Origin	The first language	Other languages	Elicitation methods
Hindi-Urdu	Male	1972	Karachi, Pakistan	Urdu	Punjabi, English	virtual meeting, checking written examples
Nepali	Male	1989	Jhapa, Nepal	Nepali	English, Japanese	in-person session, virtual meeting
Sinhala	Female	1998	Colombo, Sri Lanka	Sinhala	English	telephone
Bengali	Male	1975	Kolkata, India	Bengali	English, Hindi	in-person sessions

Table 1. The informants for the present study

The Hindi-Urdu informant is a male individual born in the year 1972. He hails from Karachi, Pakistan. Urdu is his first language, and he is also proficient in Punjabi and English. We elicited data from him through both virtual meetings and the checking of written examples. The Nepali informant is a male born in the year 1989. He originates from Jhapa, Nepal. Nepali is his first language, and he also speaks English and Japanese. We elicited data from him through both virtual meetings and in-person sessions. The Sinhala informant is a female born in the year 1998. She is from Colombo, Sri Lanka. Sinhala is her first language, and she also speaks English. We elicited data from her via telephone conversations. The Bengali informant is a male born in the year 1975. He hails from Kolkata, India. Bengali is his first language, and he also speaks English and Hindi.

we elicited data from him through face-to-face sessions.

In this study we focus on three elements to carry out a systematic study of the relative clause constructions in NIA languages: (i) headed and headless relative clauses, (ii) macro roles, and (iii) participial strategies based on tense or aspect.

During our elicitation sessions, we presented informants with headed and headless relative clause constructions contrastively with information on the context. See the English example below.

- (5) You should marry a man [whom you love] and you should not marry [whom you do not love].

The first half of the example in (5) contains a headed relative clause construction, and the second half contains a headless relative clause construction. The contrastive illustration of headed and headless relative clause constructions enables an informant to interpret a headless relative clause easily. This is due to the fact that the interpretation of headless relative clauses relies on the context in many cases since a head noun phrase is absent in a headless relative clause construction.

As mentioned earlier, previous studies have focused more on the grammatical relation of an extracted argument in relativization. However, we investigated relative clause constructions with a focus on the macro roles of an extracted argument, that is S, A, P, T, and R. Each macro role corresponds to the single argument of an intransitive construction, the agent of transitive construction, the patient of transitive construction, the theme of a ditransitive construction, and the recipient of a ditransitive construction, respectively. English examples of each macro role are given in (6).

- (6) Macro roles
- S macro role: **A train** is coming from Delhi.
  - A macro role: **A boy** is reading a book.
  - P macro role: A boy is reading **a book**.
  - T macro role: I will give **a gift** to my friend.
  - R macro role: I will give a gift **to my friend**.

Finally, in our study, we focused on the participial strategies of relative clause constructions. Many NIA languages have multiple participial strategies for relative clause constructions based on aspect or tense. For example, Hindi-Urdu has two distinct participial strategies based on aspect: imperfective and perfective participles. See the examples below.

- (7) The imperfective participle strategy in Hindi

[ <i>ro-t-a</i>	<i>ɦʊ-a</i> ]	<i>bəttʃa</i>	<i>mā=ko</i>	<i>dekʰ-kəʀ</i>
cry-IPFV.PTCP-M.SG	be-PFV.PTCP	child.M.SG	mother=DAT	see-CP
<i>tʃʊp</i>	<i>ɦo</i>	<i>gə-ja</i>		
quiet	be	go-PFV.PTCP.M.SG		

‘The child who was crying became quiet when he saw his mother.’ (Kachru 2006: 137)

- (8) The perfective participle strategy in Hindi

[ <i>kʰət =pəʀ</i>	<i>bətʰ-a</i>	<i>ɦʊ-a</i> ]	<i>admi:</i>
cot=on	sit-PFV.PTCP.M.SG	be-PFV.PTCP	man
<i>koi:</i>	<i>upənjəs</i>	<i>pətʰ</i>	<i>rəɦ-a</i>
		<i>tʰ-a</i>	

some novel read PROG-M.SG be.PST-M.SG  
 ‘The man sitting on the cot was reading some novel.’

(Kachru 2006: 137)

In the example in (7), the imperfective participle of the verb *ro* ‘cry’ is used for relativization. It corresponds to the progressive event of crying. In the example in (8), the perfective participle of the verb *beɽʰ* ‘sit’ is used for relativization. It corresponds to the stative interpretation of the event of sitting.

When a language has two participial strategies based on the differences of aspect or tense, we included both strategies in our study. The imperfective or nonpast participle strategies and the perfective or past participle strategies can be observed in Hindi-Urdu, Nepali, Early Nepali, and Sinhala. On the other hand, Bengali has a sole participial strategy for the relative clause construction, which can be used in both perfective and imperfective aspects depending on the context.

In Hindi-Urdu, another strategy, namely *vala* construction or “agentive participle” is included in the participial strategies of relative clauses in some studies (Kachru 1980; Kachru 2006; Hook 1979). This construction is composed of “inflected infinitive form of the verb followed by the item *vala*” (Kachru 2006: 136). This *vala* construction is not included in our study, as it does not code a specific tense or aspect and behaves differently from other participles (e.g., it can also follow elements other than verbs).

We focused on these three elements mentioned at the beginning of this section in our investigation: (i) headed and headless relative clauses, (ii) macro roles, and (iii) participial strategies based on tense or aspect. The elements we focused on in this study are summarized in Table 2 below. When a language has two participial strategies based on tense or aspect, it is necessary to investigate the possibility of relative clause formation in 20 patterns.

Strategy	Head	Macro roles				
		S	A	P	T	R
Imperfective/nonpast participle strategy	headed	✓	✓	✓	✓	✓
	headless	✓	✓	✓	✓	✓
Perfective/past participle strategy	headed	✓	✓	✓	✓	✓
	headless	✓	✓	✓	✓	✓

Table 2. The summary of the parameters for the survey

### 3 The investigated languages

We investigated five NIA languages, Hindi-Urdu, Nepali, Early Nepali, Sinhala, and Bengali in the present study. The four currently-spoken languages are distributed across South Asia, as shown in the map in Figure 1.



Figure 1. The geographical location of the languages under examination

As noted earlier, among the five languages investigated, Hindi-Urdu, Nepali, Early Nepali, and Sinhala have two participial strategies for relative clause constructions based on aspect or tense. On the other hand, Bengali has a sole participial strategy for the relative clause construction, which can be used in both perfective and imperfective aspects depending on the context.

We decided to investigate the five languages listed above for two reasons. First, we wanted to investigate both ergative languages like Hindi-Urdu and Nepali and accusative languages like Sinhala and Bengali. We included both Early Nepali and Modern Nepali in the present study because Wallace (1985) notes that a change is observed between the two stages of Nepali regarding ergativity in the headless relative clauses with perfective participle strategy (it is called *-eko* nominalization by Wallace). Thus, it is worth investigating Early Nepali and Modern Nepali to observe the development of relative clause constructions. Second, each language genetically belongs to a distinct subgroup of the NIA linguistic group. As per the subcategorization of NIA languages by Chatterji (1923), Hindi-Urdu belongs to the Midland group, Nepali belongs to the North group, Sinhala belongs to the Southwest group, and Bengali belongs to the Eastern group of NIA languages, respectively. Investigating these languages enabled us to observe possible variations within the NIA languages.

Several researchers have investigated the behavior of participial strategies of relative clause constructions in these languages (see Hook & Koul 2014; Kachru 1980; Subbārão 2012; Nishioka & Kumar 2021; Ahmed 2010 for Hindi-Urdu, Wallace 1985; Paudyal 2010 for Nepali, Subbārão 2012; Chandralal 2010 for Sinhala, Dasgupta 1980; Faquire 2014; Subbārão 2012 for Bengali). Among them, the study by Subbārão (2012) is noteworthy because it focuses on macro roles to investigate relative clause constructions in South Asian languages including NIA languages. However, previous studies have not conducted a systematic investigation focusing on the three elements altogether, namely (i) headed and headless relative clauses, (ii) macro roles, and (iii) participial strategies based on tense or aspect. Thus, previous descriptions are incomplete since they do not fully address the patterns and characteristics of participial strategies employed in relative clause constructions across these languages.

#### 4 Data

In this section, we present the data from our study. The results are summarized in Tables 3 and 4. Both tables represent the results of the respective participial strategies, namely imperfective/nonpast and perfective/past participle strategies. When “OK” appears in a cell, it indicates that a specific macro role was observed to be relativized in a certain type of event. It does not necessarily mean that macro roles in all types of events can be relativized when “OK” is shown.

Language	S	A	P	T	R
Hindi-Urdu	OK	NO	NO	NO	NO
Early Nepali	OK	OK	NO	NO	NO
Modern Nepali	OK	OK	OK	OK	OK
Sinhala	OK	OK	OK	OK	OK
Bengali	OK	OK	OK	OK	OK

Table 3. The summary of the results: imperfective/nonpast participle strategy

Language	P	T	S	A	R
Early Nepali	OK	OK	OK	NO	NO
Hindi-Urdu	OK	OK	OK	OK	NO
Modern Nepali	OK	OK	OK	OK	OK
Sinhala	OK	OK	OK	OK	OK
Bengali	OK	OK	OK	OK	OK

Table 4. The summary of the results: perfective/past participle strategy

We discuss the results presented in Tables 3 and 4 for each language in the following section.

## 4.1 Hindi-Urdu

Hindi-Urdu has two participial strategies, namely imperfective and perfective participle strategy. We demonstrate the data related to the imperfective participle strategy and the perfective participle strategy in Sections 4.1.1 and 4.1.2, respectively.

### 4.1.1 Imperfective participle strategy

In Hindi-Urdu, only S is relativized with the imperfective participle strategy. Both headed and headless relative clauses are accepted. Other macro roles are not relativized with this strategy.

#### (9) S relativization

[*tʃəl-t-i:* (*ɦʊ-i:*)]      *gɑ:ri:=se*    *kʰu:d*    *pəɽ-na*    *beɳəqu:fi:*    *ɦɛ*  
 move-IPFV.PTCP-F    (be-PFV.PTCP.F)    train=from    jump    fall-INF    foolish    be.3.PRS  
 ‘To jump from a moving train is stupidity.’ (McGregor 1986: 156)

#### (10) S relativization

[*məɾ-t-a*]      *kja*    *nə*    *kəɾ-t-a?*  
 die-IPFV.PTCP-M.SG    what    NEG    do-PTCP-M.SG

‘What wouldn’t a dying man do?’ (McGregor 1986: 158)

(11) A relativization

\*[*kitab* *pəɽʰ-t-a* *ɦʊ-a*] *ləɽka*  
 book read-IPFV.PTCP-M.SG be-PFV.PTCP.M.SG boy  
*mera* *tʰoɽa* *bʰai:* *ɦe*  
 1.GEN.M.SG small.M.SG brother be.PRS.3SG  
*ɔ:r* [*əkʰbar* *pəɽʰ-t-i:* *ɦʊ-i:*] *meri:* *baɽi:* *bəɦən* *ɦe*  
 and newspaper read-IPFV.PTCP-F be-PFV.PTCP-F 1.GEN.F big.F sister be.PRS.3SG  
 ‘The boy who is reading a book is my younger brother and the one who is reading the newspaper is my elder sister.’

(12) P relativization

\*[*mere* *bʰai:=ki:* *pəɽʰ-t-i:* *ɦʊ-i:*] *kitab*  
 1.SG.GEN.OBL brother=GEN.F read-IPFV.PTCP-F be-PFV.PTCP.F book.F  
*mẽ* *bʰi:* *bəɦpən=mẽ* *pəɽʰ-t-i:* *tʰ-i:*  
 1.SG.NOM also childhood=in read-IPFV.PTCP-F be.PAST-F.SG  
*ɔ:r* [*meri:* *bəɦən=ka* *pəɽʰ-t-a* *ɦʊ-a*]  
 and 1.SG.GEN.F sister=GEN.M.SG read-IPFV.PTCP-M.SG be-PFV.PTCP.M.SG  
*mẽ* *bʰi:* *roz* *pəɽʰ-t-i:* *ɦũ*  
 1.SG.NOM also everyday read-IPFV.PTCP-F be.PRS.1SG  
 ‘I used to read the book which my brother is reading, and I also read the one which my sister is reading every day.’

(13) T relativization

\*[*mera* *apne* *dost=ko* *de-t-a*] *tofa is*  
 1.SG.GEN.M.SG self.M.OBL friend=DAT give-IPFV.PTCP-M.SG gift this.OBL  
*kəmre=mẽ* *ɦe* *ɔ:r* [*tumhara* *apni:* *dost=ko* *de-t-a*]  
 room=in be.PRS.3SG and 2.SG.GEN.M.SG self.SG.F friend=DAT give-IMP.F.PTCP-M.SG  
*us* *kəmre=mẽ* *ɦe*  
 that.OBL room.OBL=in be.PRS.3SG  
 ‘The gift which I will be giving to my friend is in this room and the one which you will be giving to your friend is in that room.’

(14) R relativization

\*[*mera* *adʒ* *tofa* *de-t-a*] *admi:*  
 1.SG.GEN.M.SG today gift give-IPFV.PTCP-M.SG man  
*mera* *dost* *ɦe* *ɔ:r* [*meri:* *bi:vi:=ki:* *tofa* *de-t-i:*]  
 1.SG.GEN.M.SG friend be.PRS.3.SG and 1.SG.GEN.F wife=GEN.F gift give-IPFV.PTCP-F  
*us=ki:* *səɦeli:* *ɦe*  
 that=GEN.F female.friend be.PRS.3.SG  
 ‘The person to whom I will be giving a gift today is my friend, and the one to whom my wife will be giving a gift is her friend.’

To summarize, in Hindi-Urdu, only S is relativized with the imperfective participle strategy in headed and headless relative clause constructions. The literature mentions that the subject as a grammatical relation can be relativized by the imperfective participle strategy in Hindi-Urdu (Kachru 1980: 35). However, our data demonstrated that only S is possible.

#### 4.1.1 Perfective participle strategy

Hindi-Urdu allows P, T, S, and A macro roles to be relativized with the perfective participle strategy. Both headed and headless relative clauses are accepted for these macro roles.

(15) P relativization

[*səlma=ki: pɪtʃ<sup>h</sup>le sal lɪk<sup>h</sup>-i: ɦɪv-i:] kɪtab*  
 Salma=GEN.F last.M.OBL year write-PTCP.F be-PFV.PTCP.F book  
*əttʃ<sup>h</sup>i: t<sup>h</sup>-i: ɔr [səlma=ki: ɪs sal*  
 good.F be.PST-F.SG and Salma=GEN.F this.OBL year  
*lɪk<sup>h</sup>-i: ɦɪv-i:] b<sup>h</sup>i: tʃi:k tʃ-i:*  
 write-PFV.PTCP.F be-PFV.PTCP.F also fine be.PST-F.SG  
 ‘The book which Salma wrote last year was good, and the one which Salma wrote this year was also fine.’

(16) T relativization

*g<sup>h</sup>ər=mẽ [i:fwər=ka dɪ-ja ɦɪv-a]*  
 house=in god=GEN.M.SG give-PFV.PTCP.M.SG be-PFV.PTCP.M.SG  
*səb kʊtʃ<sup>h</sup> ɦɛ*  
 all anything be.PRS.3SG  
 ‘Everything that God/the god gave us is in the house.’ (Premchand, *Nirmala*)

(17) T relativization

[*ɔn=ka dɪ-ja ɦɪv-a] ɦəm kəb<sup>h</sup>i:*  
 3PL.OBL-GEN.M.SG give-PFV.PTCP.M.SG be-PFV.PTCP.M.SG 1PL.NOM never  
*nəɦĩ tʃʊka sək-t-e*  
 NEG complete be.able-IPFV.PTCP-M.PL  
 ‘You can never repay what they gave.’ (Nishioka & Kumar 2021: 91)

(18) S relativization

*am=ke bag=mẽ gað=ke ləɽke ləɽkiyã*  
 mango=GEN.M.OBL garden=in village=GEN.M.PL boy.PL girl.PL  
 [*ɦəwa=se ɡɪr-e ɦɪv-e] am*  
 wind.F=from fall-PFV.PTCP.M.PL be-PFV.PTCP.M.PL mango  
*tʃʊn rəɦ-e t<sup>h</sup>-e*  
 select PROG-M.PL be.PST-M.PL  
 ‘The boys and girls from the village were picking up mangos which fell through the air into the mango garden.’ (Premchand, *Algyojha*)

(19) S relativization

[*ɡɪr-ð]=ko ʊt<sup>h</sup>a-o*  
 fall-PFV.PTCP.M.PL.OBL=DAT raise-IMP  
 ‘Raise up the fallen.’ (McGregor 1986: 158)

(20) A relativization

[*pɪ-ja ɦɪv-a] admi: tʃəl rəɦ-a ɦɛ ɔr*  
 drink-PFV.PTCP.M.SG be-PFV.PTCP.M.SG man move PROG-M.SG be.PRS.3.SG and  
*ʊd<sup>h</sup>ər [pɪ-ja ɦɪv-a] natʃ rəɦ-a ɦɛ*  
 there drink-PFV.PTCP.M.SG be-PFV.PTCP.M.SG dance PROG-M.SG be.PRS.3.SG  
 ‘The drunken man is walking and another drunken man is dancing over there.’

(21) R relativization

\*[*ɔ:rət=ka k<sup>h</sup>ilɔna dɪ-ja ɦɪv-a] bəttʃa dərəsəl*  
 lady=GEN.M.SG toy give-PFV.PTCP.M.SG be-PFV.PTCP.M.SG child actually

*mera b<sup>h</sup>ai: hε pər [ɔ:rət=ke mɪt<sup>h</sup>ai: dɪ-e*  
 1SG.GEN.M.SG brother COP.PRS.3SGbut lady=GEN.M.OBL sweet give-PFV.PTCP.M.OBL  
*hɪv-e]=ko mē nəhī: dʒan-t-a*  
 be-PFV.PTCP.M.SG.OBL=DAT 1SG.NOM NEG know-IPFV.PTCP-M.SG  
 ‘The child to whom the lady gave a toy is actually my brother, but I do not know the one  
 to whom she gave a sweet.’

In summary, Hindi-Urdu allows P, T, S, and A macro roles to be relativized with the perfective participle strategy. Headed and headless relative clauses show the same behavior with respect to the macro roles to be relativized on.

## 4.2 Early Nepali

Early Nepali has two participial strategies, namely imperfective and perfective participle strategies, which we demonstrate in Sections 4.2.1 and 4.2.2, respectively.

### 4.2.1 Imperfective participle strategy

According to Wallace (1985), in Early Nepali, S and A are relativized with the imperfective participle strategy.

(22) S relativization

*g<sup>h</sup>a vΛ-nja dek<sup>h</sup>i [kirat=baʎΛ b<sup>h</sup>agi-dza-nja]=kana pΛkΛr-erΛ*  
 union make-IPFV.PTCP after Kirat=from flee-go-IPFV.PTCP=ACC capture-CVBS  
*hami=lai saũpi di-nja t<sup>h</sup>Λ*  
 1PL=DAT ally give-IPFV.PTCP COP.PRS.3  
 ‘After the alliance is made, our ally will give us those who fled from Kirat whom he  
 captured.’ (Wallace 1985: 108)

(23) A relativization

*[tʃita-jako kamΛna purjau-nja] adʒΛ dlibΛ tʃ<sup>h</sup>Λ ʌrko tʃ<sup>h</sup>Λina*  
 think-PFV.PTCP desire fulfill-IPFV.PTCP today fate COP.PRS other COP.NEG  
 ‘That which fulfills our desires is fate and nothing else.’ (Wallace 1985: 108)

### 4.2.2 Perfective participle strategy

According to Wallace (1985), in Early Nepali, P, T, and S are relativized with the perfective participle strategy.

(24) P relativization (18<sup>th</sup> century)

*[b<sup>h</sup>Λn-jako] sunjΛũ*  
 say-PFV.PTCP hear.PST.1PL  
 ‘We heard what was said.’ (Wallace 1985: 109)

(25) T relativization (19<sup>th</sup> century)

*tΛsΛrt<sup>h</sup>Λ taha [mΛ=kΛne pɾΛkafΛ gΛr-jako]*  
 therefore then 1SG=DAT clear do-PFV.PTCP  
*timi=le nΛ-dzan-jako ho*  
 2SG=ERG NEG-know-PFV.PTCP be.PRS.3SG



‘Therefore, you do not understand that which has been made clear to me.’  
(Wallace 1985: 109)

(26) S relativization (19<sup>th</sup> century)

[*bãtf-jak-i*]                      *mer-i*                      *huntʰʌ*  
survive-PFV.PTCP-F      1SG.GEN-F      be.PRS.3SG

‘The one who survived is my wife.’ (Wallace 1985: 109)

### 4.3 Nepali

Nepali (Modern Nepali) has two participial strategies, namely imperfective and perfective participle strategies, which we demonstrate in sections 4.3.1 and 4.3.2, respectively.

#### 4.3.1 Imperfective participle strategy

In Nepali, the relativization of all macro roles with the imperfective participle strategy was accepted by our informant. Also, both headed and headless relative clauses are accepted in each macro role.

(27) S relativization

[*biratnagar=baʃʌ au-ne*]                      *bʌs ʌhile=sʌmma*                      *pug-eko*                      *tʃʰʌina*  
Biratnagar=from      come-IPFV.PTCP      bus      now=till                      arrive-PFV.PTCP                      COP.NEG

*tʌʀʌ [kaʃʰmãdʒʌũ=baʃʌ au-ne]*                      *ek gʰʌntʌ ʌgaʒi nʌi*  
but Kathmandu=from      come-IPFV.PTCP      one hour                      before EMPH

*pug-i*                      *sʌk-j-o*  
arrive-CP                      finish-PST-3

‘The bus which comes from Biratnagar has not arrived yet, but the one which comes from Kathmandu arrived one hour ago.’

(28) A relativization

[*fʉʃbol dʰerʌi dzit-ne*]                      *def*                      *brazil ʃio*  
football      much      win-IPFV.PTCP      country      Brazil      be.PRS.3

*ʌni [kriket dzit-ne]*                      *ʌstrelija ʃio*  
and cricket      win-IPFV.PTCP      Australia      be.PRS.3

‘The national team which wins soccer games is Brazil, but the one which wins cricket games is Australia.’

(29) P relativization

[*brazil=le dʰerʌi dzit-ne*]                      *kʰel fʉʃbol ʃio*  
Brazil=ERG      much      win-IPFV.PTCP      game      football      be.PRS.3

*tʌʀʌ [ʌstrelija=le dzit-ne]*                      *kriket ʃio*  
but      Australia=ERG      win-IPFV.PTCP      cricket      be.PRS.3

‘The game which Brazil wins is football, but the one which Australia wins is cricket.’

(30) T relativization

[*mʌi=le us=lai di-ne*]                      *kura=ʃʌru tjo koʃʰa=ma tʰie*  
1SG=ERG      3SG.OBL=DAT      give-IPFV.PTCP      thing=PL      that room=in      be.PST.3

*rʌ [mʌi=le tʌpʌĩ=lai di-ne]=ʃʌru*                      *jo koʃʰa=ma tʰie*  
and      1SG=ERG      2SG=DAT      give-IPFV.PTCP=PL      this room=in      be.PST.3

‘The things which I gave to him/her were in that room, and the ones which I gave to you were in this room.’

(31) R relativization

[*mai=le adʒa gift di-ne*]                      *mantʃe mero satʃi ho*  
 1SG=ERG today gift give-IPFV.PTCP person 1SG.GEN.M friend be.PRS.3  
*ani [meri srimati=le gift di-ne]*                      *unki satʃi hun*  
 and 1SG.GEN.F wife=ERG gift give-IPFV.PTCP 3SG.HON.GEN.F friend be.PRS.3.HON  
 ‘The person to whom I will give a gift today is my friend, and the one to whom my wife will give a gift is her friend.’

In summary, S, A, P, T, and R are relativized with the imperfective participle strategy in headed and headless relative clauses in Modern Nepali.

### 4.3.2 Perfective participle strategy

In Nepali, the relativizations of all macro roles, namely P, T, S, A, and R with the perfective participle strategy were accepted by our informant. Also, both headed and headless relative clauses are accepted in each macro role.

#### (32) P relativization

[*brazil=le olimpik=ma dzit-eko*]                      *sport fuṭbol tʃijo*  
 Brazil=ERG Olympic=in win-PFV.PTCP sport football be.PST.3  
*ani [aʃtrelija=le dzit-eko]*                      *hʌkki tʃijo*  
 and Australia=ERG win-PFV.PTCP hockey be.PST.3  
 ‘The sport which Brazil won in the Olympics was football, and the one which Australia won was hockey.’

#### (33) T relativization

[*mai=le us=lai di-eko*]                      *kura=hʌru tjo koṭʰa=ma tʃie*  
 1SG=ERG 3SG=DAT give-PFV.PTCP thing=PL that room=in be.PST.3  
*ra [mai=le tapaĩ=lai di-eko]=hʌru jo koṭʰa=ma tʃie*  
 and 1SG=ERG 2SG=DAT give-PFV.PTCP=PL this room=in be.PST.3  
 ‘The things which I gave to him/her were in that room, and the ones which I gave to you were in this room.’

#### (34) S relativization

[*biratnagar=baṭʌ a-eko*]                      *bas tʃik ʃaim=ma a-i pug-y-o*  
 Biratnagar=from come-PFV.PTCP bus fine time=in come-CP arrive-PST-3  
*ani [kaṭʰmāḍʌũ=baṭʌ a-eko]*                      *ek gʰʌnṭa aḡʌi nai*  
 and Kathmandu=from come-PFV.PTCP one hour before EMPH  
*pug-i sak-j-o*  
 arrive-CP finish-PST-3

‘The bus which came from Biratnagar has already arrived on time, and the one which came from Kathmandu arrived one hour ago.’

#### (35) A relativization

[*2022 sal=ma fuṭbol warlḍkap dzit-eko*]                      *deʃ ʌrzentina ho*  
 2022 year=in football worldcup win-PFV.PTCP country Argentina be.PRS.3  
*ani [tjoʃi varʃa krikeṭ warlḍkap dzit-eko]*                      *ingland=le ho*  
 and that year cricket worldcup win-PFV.PTCP England=ERG be.PRS.3  
 ‘The national team which won the Soccer World Cup in 2022 was Argentina, and the one which won the Cricket World Cup in that year was England.’

#### (36) R relativization

[*mʌhila=le kʰilɔna di-eko*]                      *tfora*                      *bastʌv=ma*  
 lady=ERG                      toy                      give-PFV.PTCP                      child                      actuality=in  
*mero*                      *bʰai*                      *ɦio*  
 1SG.GEN.M                      brother                      be.PRS.3SG  
*tʌʌ*                      [*mʌhila=le mitʰai di-eko*]=*la*                      *mʌlai*                      *tʰaɦa*                      *tʰai-na*  
 but                      lady=ERG                      sweet                      give-PFV.PTCP=DAT                      1SG.DAT                      known                      be.1SG-NEG  
 ‘The child to whom the lady gave a toy is actually my brother, but I do not know the one  
 to whom the lady gave a sweet.’

In summary, In Nepali, S, A, P, T, and R are relativized with the perfective participle strategy in headed and headless relative clause constructions.

#### 4.4 Sinhala

##### 4.4.1 Nonpast participle strategy

In Sinhala, the relativizations of all macro roles with the nonpast participle strategy are accepted. Examples of these are given in (37)–(41).

(37) S relativization

[*mehe: innə*]                      *lamajə*  
 here                      exist.NPST.PTCP                      child  
 ‘the child who exists here’

(38) A relativization (Chandralal 2010: 131)

[*darua-wə hojənə*]                      *amma*  
 child-ACC                      search.NPST.PTCP                      mother  
 ‘the mother who searches for her child’ or ‘the mother, who searches for her child’

(39) P relativization (Chandralal 2010: 131)

[*amma hojənə*]                      *darua*  
 mother                      search.NPST.PTCP                      child  
 ‘the child whom the mother searches for’

(40) R relativization

[*randzit potə denə*]                      *lamea*  
 Ranjit                      book                      give.NPST.PTCP                      child  
 ‘the child to whom Ranjit gives the book’

(41) T relativization

[*randzit lamea-tə denə*]                      *potə*  
 Ranjit                      child-DAT                      give.NPST.PTCP                      book  
 ‘the book which Ranjit gives to the child’

In Sinhala, the nonpast participle does not function as a noun phrase without modifying a noun or pronoun.

##### 4.4.2 Past participle strategy

The relativizations of all macro roles with the past participle strategy were accepted by our Sinhala informant, as shown in (42)–(46).

(42) S relativization

[*mehe: hitijə*] *lamajə*  
here exist.PST.PTCP child  
'the child who existed here'

(43) A relativization

[*darua-wə hojəpu*] *amma*  
child-ACC search.PST.PTCP mother  
'the mother who searched for her child' or 'the mother, who searched for her child.'

(44) P relativization

[*amma hojəpu*] *darua*  
mother search.PST.PTCP child  
'the child whom the mother searched for.'

(45) R relativization (Chandralal 2010: 131)

[*randzit potə dunnə*] *lamea*  
Ranjit book give.PST.PTCP child  
'the child to whom Ranjit gave the book'

(46) T relativization

[*randzit lamea-tə dunnə*] *potə*  
Ranjit child-DAT give.PST.PTCP book  
'the book which Ranjit gave to the child'

In Sinhala, the past participle does not function as a noun phrase without modifying either a noun or pronoun.

#### 4.5 Bengali

Bengali does not have multiple participial strategies for relative clause constructions based on the differences of aspect or tense. There is only one participial strategy. Our Bengali informant accepted examples of headed and headless relative clauses with all macro roles.

(47) S relativization

[*tfennai t'eke af-a*] *tren=guli ekʰɔn-o pountʰo-e-ni*  
Chennai from come-PTCP train=CLF now-also arrive-PRS.3-NEG  
*təbe [dilli t'eke af-a]=guli æk gʰɔntə age pountʰ-etʰ-e*  
but Delhi from come-PTCP=CLF one hour before arrive-PRF-3  
'The trains which come from Chennai have not arrived yet but the ones which come from Delhi arrived one hour ago.'

(48) A relativization

[*futbol biʃʃokap dzit-e ne-wa*] *def=guli ho-ttʰ-e bradzilar ardʒentina*  
football worldcup win-CP take-PTCP country=CLF be.PROG.3 Brazil and Argentina  
*ar [kriket biʃʃokap dzit-e ne-wa]=guli ho-l-o ɔʃtrelija ar bʰarɔt*  
and cricket worldcup win-CP take-PTCP=CLF be-PST-3 Australia and India  
'The countries which win the Soccer World Cup are Brazil and Argentina, and the ones which win the Cricket World Cup are Australia and India.'

(49) P relativization

[*amar adzke badzar-e ken-a*] *dzinif=guli amar baʀi-te atʰ-e*  
1.SG.GEN today market-LOC buy-PTCP thing=CLF 1.SG.GEN house-LOC be-3

*ar* [tomar kalke badzar-e ken-a]=guli ek<sup>h</sup>an-e ro-et<sup>h</sup>-e  
 but 2.SG.GEN yesterday market-LOC buy=CLF here-LOC stay-PRF-3  
 ‘The things which I bought in the market today are in my house and the ones which you bought in the market yesterday are here.’

(50) T relativization

[amar take de-wa] dzinif=guli fei g<sup>h</sup>or-e t<sup>h</sup>il-o  
 1SG.GEN 3SG.DAT give-PTCP thing=CLF that room-LOC be.PST-3  
*ar* [amar apnake de-wa]=guli ei g<sup>h</sup>or-e t<sup>h</sup>il-o  
 and 1SG.GEN 2SG.HON.DAT give-PTCP=CLF this room-LOC be.PST-3  
 ‘The things which I gave to him were in that room, and the ones which I gave to you were in this room.’

(51) R relativization

[mohila-r k<sup>h</sup>elna de-wa] battfa=guli aʃɔl-e amar b<sup>h</sup>ai,  
 lady-GEN toy give-PTCP child=CLF actual-LOC 1SG.GEN brother  
*tɔbe* [mohila-r miʃti de-wa]=guli=ke ami tʃin-i na  
 but lady-GEN sweet give-PTCP=CLF=DAT 1SG.NOM know-PRS.1 NEG  
 ‘The children to whom the lady gave a toy are actually my brothers, but I do not know the ones to whom the lady gave a sweet.’

In summary, in Bengali, all macro roles, namely S, A, P, T, and R are relativized with the participial strategy in headed and headless relative clauses.

#### 4.6 Summary

In this section, we presented data on participial strategies for relative clause constructions in Hindi-Urdu, Early Nepali, Modern Nepali, Sinhala, and Bengali. The findings of the investigation are summarized in Tables 3 and 4 presented above.

### 5 Discussion

The NIA languages examined in this study show both similarities and differences with regard to the relativizability of relative clause constructions. On the one hand, both headed and headless relative clauses are found in the same range of macro roles if a language has both. Among the languages examined, Hindi-Urdu, Early Nepali, Nepali, and Bengali have both headed and headless relative clauses. What can be relativized is the same regardless of the presence or absence of the head NP. In previous studies, headed and headless relative clauses have not been examined together except in the case of Nepali (Wallace 1985). As for Nepali, it has been shown that only the subject is relativized in headless relative clauses via imperfective participles, while there is no such restriction for grammatical relations in headed relative clauses by imperfective participles. This study systematically examined the relativization of S, A, P, T, and R both with and without the head NP for the five languages. We did not find relative clauses that always lack the head NP or that cannot lack the head NP in any of the languages examined.

On the other hand, the five languages differ as to which macro roles can be relativized. Even inside a language, different ranges of macro roles can be relativized by different participles. In previous studies, grammatical relations, such as subject and direct object, are often the parameters of the examination, and consideration is not given to possible differences among macro roles. For example, Hook & Koul (2014) show that relativization by imperfective

participle is only available for subject in Hindi-Urdu. A subject can be interpreted to be composed of S and A. It is not clearly mentioned whether both S and A behave in the same way. In contrast, this study showed that the macro roles treated under one grammatical relation in a given language can show different syntactic behaviors with respect to relativization. We showed that S can be relativized with the imperfective participle in Hindi-Urdu, but A cannot.

Based on the results of our investigation, we propose aspect-based implicational hierarchies of relativizability for NIA languages. For relative clauses with imperfective/nonpast participles, we propose the implicational hierarchy in (52).

(52) Hierarchy of macro roles in imperfective/nonpast (Onset-oriented Hierarchy):

$$\{S\} > \{A\} > \{P, T, R\}$$

We consider the macro roles between parentheses to have equal status in the hierarchy. For example, in (52), P, T, and R are written together between parentheses, and we do not consider there to be any hierarchical order among them. The order of the macro roles in a parenthesis is irrelevant. The data for Hindi-Urdu and Early Nepali create the breakpoints. Hindi-Urdu allows relativization for S, but not for A, P, T, and R. Early Nepali allows relativization for S and A, but not for P, T, and R. The other languages in this study allow relativization for all macro roles.

For relative clauses with perfective/past participles, we propose the implicational hierarchy in (53).

(53) Hierarchy of macro roles in perfective/past (Termination-oriented Hierarchy):

$$\{S, P, T\} > \{A\} > \{R\}$$

Again, the Hindi-Urdu and Early Nepali data create the breakpoints. Early Nepali allows relativization for S, P, and T, but not for A and R. Hindi-Urdu allows relativization for S, P, T, and A, but not for R. The other languages in this study allow relativization for all the macro roles.

We explain the hierarchies in (52) and (53) uniformly based on the concept of viewpoint (DeLancey 1981; DeLancey 1982). DeLancey (1981) proposes that the domains of space, time, and transitivity have a vector from the onset to the termination as in (54) and that an event can be construed with the viewpoint on either the onset or the termination of one of these domains.

(54) The vectors in the domains of space, time, and transitivity:

Space:	Source → Goal
Time:	Onset → Termination
Transitivity:	Agent → Patient

In the domain of space, the onset is Source and the termination is Goal. In the domain of time, the onset is Onset and the termination is Termination. In the domain of transitivity, the onset is Agent and the termination is Patient. For example, an event described by the verb *go* is seen from the source of the domain of space. These three domains are not independent from one another. The three vectors in (54) are related metaphorically based on localist metaphor (Croft 2001). DeLancey (1982) argues that an imperfective event has the viewpoint on the temporal onset (Onset) and a perfective event has the viewpoint on the temporal termination (Termination). In this paper, we propose that an event with the viewpoint on the temporal onset

(Onset) metaphorically evoke an event with the viewpoint on the transitivity onset (Agent) based on localist metaphor. The event with the viewpoint on the transitivity onset (Agent) further evokes the agentive participant of the event based on a the event for THE EVENT FOR THE PROTAGONISTS metonymy relationship. The hierarchies in (52) and (53) are explained from these processes. We argue that the reason why S and A come before the other macro roles in the onset-oriented hierarchy in (52) is because an imperfective event tend to evoke the agentive participant of the event. Similarly, a perfective event tend to evoke the patientive participant of the event, namely, P and patientive S. S and P come before the other macro roles in the termination-oriented hierarchy in (53) because of these processes.

We believe that the relativizability of NPs in the five languages examined in this study is better captured by the onset-oriented and termination-oriented hierarchies we present in (52) and (53) than by the NP Accessibility Hierarchy (Keenan & Comrie 1977). In the NP Accessibility Hierarchy, generalizations are made with reference to grammatical relations like subject and direct object. Subject is the syntactic generalization over S and A, and direct object is the generalization over P and T. P and T are in the same position both in (52) and (53). This pattern can be generalized by the NP Accessibility Hierarchy. However, S and A behave differently both in (52) and (53). The differences between S and A cannot be appreciated when these macro roles are grouped in a single category subject, as in the NP Accessibility Hierarchy. Moreover, by presenting two different hierarchies, we can see the difference of relativizability of NPs depending on tense and aspect. These hierarchies are explained in terms of viewpoint, localist metaphor, and THE EVENT FOR THE PROTAGONISTS metonymy relationship.

## 6 Conclusion

In this study, we examined the relativizability of NPs in the five NIA languages: Hindi-Urdu, Nepali, Early Nepali, Sinhala, and Bengali. First, we investigated both headless and headed relative clauses of participial strategies. Second, we examined relativization on arguments for each of the macro roles S, A, P, T, and R. Third, we examined every participial strategy for relative clause constructions when a language has different participles depending on tense or aspect. Our investigation showed that there are both similarities and differences in relative clause constructions in the five NIA languages examined. On the one hand, none of the languages examined shows any difference of relativizability between headed and headless relative clauses. On the other hand, the five languages differ as to which macro roles can be relativized. Based on these findings, we proposed two novel hierarchies of relativizability for the five NIA languages. We proposed the onset-oriented hierarchy  $\{S\} > \{A\} > \{P, T, R\}$  for relative clauses with imperfective/nonpast participles and the termination-oriented hierarchy  $\{S, P, T\} > \{A\} > \{R\}$  for those with perfective/past participles. We argued that the generalizations discussed in this study can only be made by examining imperfective/nonpast participles and perfective/past participles separately and by using macro roles rather than grammatical relations. We explained these hierarchies in terms of viewpoint, localist metaphor, and a metonymy relationship.

## Acknowledgements

Author contributions: Both authors contributed to the analysis. The data for Hindi-Urdu, Nepali, and Bengali were collected by the first author, and the data for Sinhala were collected by the second author. Sections 2–4 were written by the first author, and Sections 1 and 5–6 were written by the second author.

Acknowledgments: We are grateful to Mai Hayashi, Naonori Nagaya, Yuta Sakon, Yui Suzuki, Mizuki Tanigawa, and Duhai Zhou for their valuable comments on the paper. Previous versions of this paper were presented at “Empirical Study of the Typology of Nominalization” Project First Research Meeting of fiscal year 2022 at the University of Tokyo in September 2022, at the 165th meeting of the Linguistic Society of Japan online in November 2022. We are grateful to the audiences of these events. This work was supported by JSPS KAKENHI Grant Number JP23KJ0748 (PI: Shigeki Yoshida), JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2110, and the NINJAL collaborative research project ‘Empirical Study of the Typology of Nominalization—from Theoretical, Fieldwork, Historical and Dialectal Perspective’.

## Abbreviations

1	first person	F	feminine	PL	plural
3	third person	GEN	genitive	PRF	perfect
ACC	accusative	HON	honorific	PFV	perfective
CLF	classifier	IPFV	imperfective	PROG	progressive
COP	copula	INF	infinitive	PRS	present
CP	conjunctive participle	LOC	locative	PST	past
CVB	converb	M	masculine	PTCP	participle
DAT	dative	NEG	negative	Q	question marker
ERG	ergative	NPST	nonpast	SG	singular
EMPH	emphasis	OBL	oblique		

## References

- Abbi, Anvita. 2001. *A manual of linguistic field work and structures of Indian languages*. Muenchen: Lincom Europa.
- Ahmed, Tafseer. 2010. The unaccusativity/unergativity distinction in Urdu. *Journal of South Asian Linguistics* 3(1). 3–22.
- Chandralal, Dileep. 2010. *Sinhala*. Amsterdam: John Benjamins.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Dasgupta, Probal. 1980. *Questions and relative and complement clauses in a Bangla grammar*, Ph.D. dissertation, New York: New York University.
- DeLancey, Scott. 1981. An interpretation of split ergativity and related patterns. *Language* 57(3). 626–657.
- DeLancey, Scott. 1982. Aspect, transitivity and viewpoint. In Paul J. Hopper (ed.), *Tense-aspect: Between semantics and pragmatics*, 167–183. Amsterdam/Philadelphia: John



- Benjamins.
- Dryer, Matthew S. 2007. Noun Phrase Structure. In Timothy Shopen (ed.), *Language Typology and Syntactic Description*, vol. II: Complex Constructions, 151–205. Cambridge: Cambridge University Press.
- Faquire, Razaul Karim. 2014. Revisiting relative clauses in Japanese, with reference to Bangla. *NINJAL Research Papers* 8. 15–31.
- Hook, Peter & Okmar N. Koul. 2014. The noun phrase accessibility hierarchy and participial noun-modifying constructions in Hindi-Urdu and Kashmiri. 11th International Conference of South Asian Languages and Literatures (ICOSAL-11).
- Kachru, Yamuna. 1980. *Aspects of Hindi grammar*. New Delhi: Manohar.
- Kachru, Yamuna. 2006. *Hindi*. Amsterdam: John Benjamins.
- Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8(1). 63–99.
- Matthews, David. 1998. *A course in Nepali*. New York, NY: Routledge.
- McGregor, Ronald Stuart. 1986. *Outline of Hindi grammar: With exercises*. Reprinted. New York: Oxford University Press.
- Nishioka, Miki & Rajesh Kumar. 2021. Hindigo no meishiku shushoku kozo to kino ni tsuite [Noun modifying structures and their function in Hindi]. In Chung Sung-Yeo & Masayoshi Shibatani (eds.), *taigenka riron to gengo bunseki* [Nominalization theory and linguistic analysis], 59–151. Osaka: Osaka University Press.
- Paudyal, Netra. 2010. The syntax of three-argument verbs in Nepali. A qualifying paper submitted to Max Planck Institute Leipzig.
- Shibatani, Masayoshi. 2019. What is nominalization? Towards the theoretical foundations of nominalization. In Roberto Zariquiey, Masayoshi Shibatani & David W. Fleck (eds.), *Nominalization in Languages of the Americas*, 15–167. Amsterdam/ Philadelphia: John Benjamins.
- Shibatani, Masayoshi. 2021. Syntactic typology. In *Oxford Research Encyclopedia of Linguistics*.
- Subbārāo, Kārumūri V. 2012. *South Asian languages: A syntactic typology*. Cambridge: Cambridge University Press.
- Verbeke, Saartje. 2013. *Alignment and ergativity in new Indo-Aryan languages*. Berlin/Boston: Walter de Gruyter.
- Wallace, William David. 1985. *Subjects and subjecthood in Nepali: An analysis of Nepali clause structure and its challenges to relational grammar and government & binding*. Urbana, IL: University of Illinois at Urbana-Champaign dissertation.

# Language contact and sound change: Reasons for mutual unintelligibility between formal and colloquial registers of Tamil

ANUSHRI KARTIK-NARAYAN, *University of Colorado Boulder*

## ABSTRACT

Tamil has since its origination been diglossic, separating the formal high register from the colloquial low register. These two registers are currently mutually unintelligible (Shanmugam Pillai 1965). This analysis explores the reasons why they became unintelligible, which are proposed to be two-fold: historic language contact between Tamil and Sanskrit; and sound changes demonstrated using the Comparative Method. It has been suggested that the decline in mutual intelligibility is due to the removal of Sanskrit loanwords from the formal high register during the Tamil Purist Movement of the 20th century (Kailasapathy 1979). The earliest evidence of Tamil and Sanskrit reciprocal borrowing dates to the first Tamil literary works (Krishnamurti 2003). Where and when this language contact occurred is unclear, but it may have occurred during overlapping occupation of the Indus River Valley region by Sanskrit and Proto-Dravidian (Steever 2009). During the 20th century, the formal register replaced these loanwords with Tamil equivalents wherever possible (Kailasapathy 1979). Currently, low register Tamil is composed of 50% loanwords whereas high register Tamil is composed of only 20% loanwords (Krishnamurti 2003). It has been attested, however, that some diglossia was present before contact between Tamil and Sanskrit. Early diglossia can thus instead be explained by sound changes, which also account for current differences between the registers not attributed to loanwords. Sound changes identified in this analysis include: syncope, apocope, paragoge, stop to fricative lenition, and others. This analysis finds that language contact and sound changes contributed to the decline in intelligibility between formal and colloquial Tamil, however the nature of the language contact is still under investigation.

## 1 Introduction

Tamil is a Dravidian language spoken in the southern third of the Indian peninsula as well as parts of Sri Lanka, Malaysia, and Singapore (Steever 2009). It is the most direct descendent of reconstructed Proto-Dravidian, which has been dated to circa 8,000 BCE. Ancient Tamil arose from Proto-Dravidian in 300 BCE, which then evolved into Medieval Tamil in 700 CE, and Modern Tamil in 1600 CE (Steever 2009).

Within Modern Tamil, multiple mutually intelligible dialects are divided by regional variety, caste, and even religious sect (*Vaishnavite*: those who worship Vishnu; *Shaivite*: those who worship Shiva) (Steever 2009). However, a divide between formal register Modern Tamil (hereafter referred to as “formal”) and colloquial register Modern Tamil (hereafter referred to as “colloquial”) varieties has created an unintelligibility that is only overcome by formal education of the formal variety and native acquisition of the colloquial variety. Native speakers of colloquial Tamil who do are not educated in formal Tamil cannot understand it, and people who are only educated in formal Tamil cannot understand colloquial Tamil (Shanmugam Pillai 1965). This phenomenon of formal and colloquial registers with different methods of acquisition, known as diglossia, is common to diverse languages globally, but is known to be mutually intelligible in

such cases (Ferguson 1959). The question this paper explores is thus, how did formal Tamil become unintelligible to colloquial Tamil?

## **2 Data analysis**

In order to determine how colloquial Tamil became different from formal Tamil, the following data were analyzed following the Comparative Method of Sound Change. This data was sourced from iLearnTamil (2020), transcribed from Tamil script into the International Phonetic Alphabet (IPA) using Rajan (2014), and is presented here in tables with columns delineated English Gloss – Formal Tamil (Romanized) – Formal Tamil (IPA) – Colloquial Tamil (Romanized) – Colloquial Tamil (IPA). It is important to note that it is assumed that colloquial Tamil derives from formal Tamil, rather than both deriving from Proto-Dravidian simultaneously, and thus the proposed sound changes are in the direction of formal-to-colloquial. Additionally, the following sound changes are tendencies rather than rules because the changes do not always apply uniformly to a given context.

### **2.1 Syncope**

Syncope is deletion of a phoneme or entire syllable word-internally. Syncope in Tamil can affect syllables such as [ri], [ru], and [gɪ]. This process is depicted by the data in Table 1.

### **2.2 Apocope**

Apocope is deletion of a phoneme or entire syllable word-finally. The colloquial register of Tamil prefers words to end in vowels, with exceptions made for nasal consonants. Thus, non-nasal word-final consonants are deleted, such as [l] and [k]. This process is depicted by the data in Table 2.

### **2.3 Paragoge**

Paragoge is the insertion of a phoneme word-finally. The colloquial register of Tamil inserts a vowel such as [ə] or [u] word-finally due to the aforementioned preference for words to end in vowels. This process is depicted by the data in Table 3.

### **2.4 High front unrounded vowel becomes high not-front unrounded vowel**

The vowel [ɪ] becomes either [u] or [ʊ] as a backing process while maintaining height and rounding. This process is depicted by the data in Table 4.

### **2.5 Alveolar tap becomes velar nasal**

The alveolar tap [ɾ] becomes the velar nasal [ŋ] preceding a velar stop as a partial assimilation process via the place-of-articulation feature. This process is depicted by the data in Table 5.

### **2.6 Vowel reduction / monophthongization**

Word-final diphthongs either monophthongize to the first vowel in the segment or reduce to [ə]. This further contributes to the low register's preference for word-final monophthong vowels. This process is depicted by the data in Table 6.

## 2.7 Voiced coronal stop becomes voiceless fricative

Voiced coronal stops such as [ḍ] and [ḍ] become voiceless coronal fricatives such as [s] as a lenition process where manner of articulation is weakened, and voicing is lost. Place-of-articulation is not always retained in the change. This process is depicted by the data in Table 7.

## 2.8 Exceptions to sound change processes

Not all differences between the formal and colloquial registers of Tamil can be explained with the aforementioned sound change tendencies, as depicted by the data in Table 8. The differences between these terms cannot be explained by sound changes and thus must be the result of borrowing through language contact. The most likely contact language candidate is Sanskrit, which will be discussed in the next section.

## 3 Contact with Sanskrit

Sanskrit is an Indo-European language that no longer has native speakers and whose current relevance is limited to historic texts and Hindu functions. It is derived from the Indo-Aryan branch of Proto-Indo-European. Vedic Sanskrit arose circa 1500 BCE, then evolved into Classical Sanskrit circa 700 BCE, and fell out of spoken use by 1350 CE (Steever 2009). Sanskrit's regional distribution, as attested by the controversial but still presently acknowledged Aryan Invasion Hypothesis, began in the Indus River Valley and entered the Indian peninsula circa 2000 BCE (Steever 2009). Language contact may have occurred within the Indian peninsula during the origination of Tamil from Proto-Dravidian, however alternate theories suggest contact in the Indus River Valley during trade. The true temporal and spatial nature of contact between Tamil and Sanskrit is still under investigation.

Sanskrit literary scholars are known to have interacted and collaborated with Tamil *sangams* (literary scholars) from the beginning of Tamil's literary tradition, leading to reciprocal borrowing between the two languages. (Krishnamurti 2003). The oldest known Tamil text, the *tolkappiyam* grammar of Tamil, contains borrowed terms from Sanskrit (Krishnamurti 2003). Following a trace of the number of Sanskrit words in Tamil literature, the highest increase in proportion occurred during Medieval Tamil from 300-600 CE (Krishnamurti 2003). Borrowing of Sanskrit words was common in all of the Dravidian languages of the area, and it continued well into modern forms of these languages (Krishnamurti 2003). However, at the turn of the 20<sup>th</sup> century a surge in Tamil nationalism led to a drive for Tamil revivalism (Kailasapathy 1979). Scholars actively attempted to recreate a "pure" Tamil free of borrowed terms from Sanskrit. This movement was somewhat successful and led to a reduction in Sanskrit loanwords from 50% of formal Tamil to 20% in present day, leaving only particularly abstract concepts and proper names of religious figures (Krishnamurti 2003). The high proportion of loanwords was retained in colloquial Tamil, creating a strong separation between the two registers.

## 4 Conclusion

Tamil's formal and colloquial registers both utilized loanwords from Sanskrit to an equal extent until the purist movement of the 20<sup>th</sup> century. However, diglossia has been attested as early as the literary tradition began during the period of Ancient Tamil. It is therefore possible that until the

20<sup>th</sup> century, the two registers were mutually intelligible, and the removal of Sanskrit loanwords from formal Tamil reduced mutual intelligibility. This would better fit the understood state of global diglossia (Ferguson 1959), but may force a reckoning of whether formal and colloquial Tamil may still be considered registers or should instead be considered a reconstructed language and a mixed language.

### **Acknowledgements**

Thank you to Dr. Hannah Haynie of the University of Colorado Boulder for guidance and feedback on this research.

### **References**

- Ferguson, Charles A. 1959. Diglossia. *Word* 15(2). 325-340. DOI: 10.1080/00437956.1959.11659702
- ilearnTamil. 2020. Spoken tamil lessons. <https://ilearntamil.com/spoken-tamil-lessons-colloquial-tamil-greetings-and-introduction-lesson1/> (11 December, 2022).
- Kailasapathy, K. 1979. The tamil purist movement: A re-evaluation. *Social Scientist* 7(10). 23-51. DOI: <https://doi.org/10.2307/3516775>
- Krishnamurti, B. 2003. *The Dravidian Languages*. 1st edn. Cambridge University Press.
- Shanmugam Pillai, M. 1965. Merger of literary and colloquial tamil. *Anthropological Linguistics* 7(4.II). 98-103. DOI: <https://www.jstor.org/stable/30022544>
- Steever, Sanford B. 2009. Tamil and dravidian languages. In Comrie, Bernard (ed.), *The World's Major Languages*. 2nd edn. 635-652. New York, NY & Canada: Routledge.
- Rajan, V. 2014. Tamil transcription tool. <https://anunaadam.appspot.com/> (12 December, 2022)

# The unique functionality of Urdu light verb *jaa* and Voice head variation

SANA KIDWAI, *University of Cambridge*

FRANCES SOBOLAK, *Cornell University*

## ABSTRACT

Variation in the properties and structural position of Hindi-Urdu light verbs is well-established. Similar accounts across the literature agree on three positions within the verbal spine: a lower *v/V* position, an intermediate position, and a high external-argument-introducing head (see Butt & Ramchand, 2005; Suliman, 2015; Sobolak, 2023). In this paper, we add light verb *jaa* to this discussion. Specifically, we show that *jaa* occupies an external-argument-introducing Voice head, using evidence from instrumental causers in *jaa*-constructions. We also show that, within the Voice head typology, Voice-*jaa* is distinct from the canonical active and passive Voice heads, and is, in fact, akin to Voice in marked anticausatives.

## 1 Introduction

Variation in the properties and structural properties of Hindi-Urdu light verbs (LVs) is well-established. Similar accounts across the literature agree on three positions within the verbal spine: a lower *v/V* position, an intermediate position, and a high external-argument-introducing head (see Butt & Ramchand, 2005; Suliman, 2015; Sobolak, 2023). In this paper, we add LV *jaa* to this discussion. Specifically, we show that *jaa* occupies an external-argument-introducing Voice head, using evidence from instrumental causers in *jaa*-constructions. We also show that, within the Voice head typology, Voice-*jaa* is distinct from the canonical active and passive Voice heads, and is, in fact, akin to Voice in marked anticausatives.

The organization of this paper is as follows. In Section 1.1, we briefly review the literature around LVs and specifically LVs in Hindi-Urdu. In Section 2, we present the structural properties of LV *jaa*: its distribution (§2.1), its semantic effect on the interpretation of the event (§2.2), and its effect (or lack thereof) on case (§2.3). In Section 3, we zero in on the structural properties of *jaa*; namely, its ability to introduce an instrumental-marked causer. We show that the structural properties of this causer pattern exactly like those of other external arguments, specifically oblique causers found in marked anticausative constructions. We propose an analysis for this *jaa* Voice head, situating it within the canonical Voice typology. Throughout this paper, we discuss the properties of *jaa* specifically in Urdu, and highlight relevant points of variation between Hindi and Urdu. We then step back and compare the properties of *jaa* to other light verbs in Urdu, in Section 4: namely, *de* ‘give’ (§4.1), *le* ‘take’ (§4.2), and *paR* ‘fall(en)’ (§4.3). We conclude in Section 5.

## 1.1 Relevant properties of light verbs

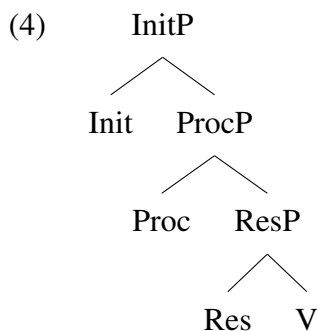
The first observation of LVs is often attributed to Jespersen (1965), who notes that some verbs in English appear to have little to no lexical semantics. Since then, there has been robust description and analysis of LVs cross-linguistically, many of which have common properties. Some key characteristics of LVs include: monoclausality (Butt, 2003), no  $\theta$ -role (Grimshaw & Mester, 1988), and additional aspectual flavor to event interpretation (Butt & Ramchand, 2005; Sobolak, 2023). Consider the difference between *give* in (1) and (2). As a main verb *give* assigns two  $\theta$ -roles in (1): THEME to *toy* and GOAL to *Lennon*. However, as a LV in (2), *give* does not assign a  $\theta$ -role to its complement – *a sigh* is neither a theme nor a patient of the event. Additionally, adding the second internal argument, *Lennon*, is ungrammatical because LV *give* does not assign a  $\theta$ -role.

- (1) Katherine gave a toy to Lennon.  
 (2) Katherine gave a sigh (\*to Lennon).  $\approx$  Katherine sighed.

In Urdu, a LV co-occurs with a main verb, as shown in (3), where the LV *le* ‘take’ modifies the main verb *likH* ‘write.’ (Butt & Ramchand, 2005) show that these LV constructions in Urdu are monoclausal.

- (3) Nadya=ne khat likH **li-ya**.  
 Nadya=ERG letter.NOM write take-PFV.M.SG  
 ‘Nadya wrote a letter [completely].’ (Butt, 2003, p.21)

Butt & Ramchand (2005) propose that LVs in Hindi-Urdu vary in their position in the verbal spine. Following Ramchand (2008), they assume three verbal heads: Result, Process, and Initiate (4). They propose that some LVs in Hindi-Urdu occupy the intermediary projection (Proc) while others occupy the highest verbal head (Init).



One of Butt & Ramchand’s (2005) arguments for multiple LV positions comes from stacking LV constructions. For example, LV *de* ‘give’ and LV *le* ‘take’ can stack, but crucially only in a specific order (5a vs. 5b). Butt & Ramchand argue this stacking asymmetry is due to the LVs’ positions in the verbal spine: *de* is introduced higher than *le*.

- (5) a. Nadya=ne          Saddam=ko          khat          likH **le-ne**  
 Nadya.F.SG=ERG Saddam.F.SG=DAT letter.M.NOM write take-INF  
**di-ya.**  
 give-PFV.M.SG  
 ‘Nadya let Saddam write a letter [completely].’ (Butt, 2003, p.21)
- b. \* Nadya=ne          Saddam=ko          khat          likH **de-ne**  
 Nadya.F.SG=ERG Saddam.F.SG=DAT letter.M.NOM write give-INF.OBL  
**li-ya.**  
 take-PFV.M.SG

Butt (1995) provides a detailed review of Hindi-Urdu LVs. She reports that LV *jaa* occurs only with unaccusative verbs, as in (6). Note that the main verb appears in bare root form and the LV hosts inflectional information. This differs from the lexical verb *jaa* ‘go’ which can occur by itself (7), and the passive morpheme *jaa*, which selects for a different form of the main verb, and also hosts inflectional information (8).

- (6) baraf      **pigal gai.** (LV)  
 ice.NOM melt jaa.PFV.F.SG  
 ‘The ice melted [completely].’
- (7) Ahmed **gya.** (lexical verb)  
 Ahmed go.PFV.M.SG  
 ‘Ahmed went.’
- (8) baraf      **pigl-a-i**                      **gai.** (passive morpheme)  
 ice.NOM melt-CAUS-PFV.F.SG jaa.PFV.F.SG  
 ‘The ice was melted.’

In the Sections 2–3, we dive into the structural and functional properties of LV *jaa*, showing that it has properties unique from other LVs in Urdu.

## 2 Properties of LV *jaa*

### 2.1 Distribution

*Jaa* occurs freely with unaccusative predicates (6), see Table 1, as well as some unergative (9) and transitive (10) verbs.

- (9) Ahmed      **bHaag gya.**  
 Ahmed.NOM run jaa.PFV.M.SG  
 ‘Ahmed ran [away].’
- (10) Ahmed      seb **kHa gya.**  
 Ahmed.NOM apple eat jaa.PFV.M.SG  
 ‘Ahmed ate [up] an apple.’



	<b>Verb</b>	<b><i>jaa</i></b>
i.	<i>mar-na</i> ‘to die’	✓
ii.	<i>pigal-na</i> ‘to melt’	✓
iii.	<i>gir-na</i> ‘to fall’	✓
iv.	<i>nikal-na</i> ‘to come out’	✓

Table 1: Distribution of *jaa* with unaccusatives.

However, *jaa* cannot occur as freely with unergatives or transitives as it can with unaccusatives. There appears to be no obvious pattern, such as lexical class, which determines whether *jaa* can occur with unergatives, see Table 2. Similarly, there appears to be no pattern that determines whether *jaa* can occur with transitive verbs, see Table 3.<sup>1</sup>

	<b>Verb</b>	<b>Lexical Class</b>	<b><i>jaa</i></b>
i.	<i>bHaag-na</i> ‘to run’	Activity	✓
ii.	<i>naach-na</i> ‘to dance’	Activity	×
iii.	<i>uTH-na</i> ‘to get up’	Achievement	✓
iv.	<i>so-na</i> ‘to sleep’	State	✓

Table 2: Distribution of *jaa* with unergatives.

	<b>Verb</b>	<b>Lexical Class</b>	<b><i>jaa</i></b>
i.	<i>kHa-na</i> ‘to eat’	Activity	✓
ii.	<i>chakH-na</i> ‘to taste’	Activity	×
iii.	<i>le-na</i> ‘to take’	Achievement	✓
iv.	<i>bHag-a-na</i> ‘to cause to run’	Achievement	×

Table 3: Distribution of *jaa* with transitives.

## 2.2 Semantic contribution

*Jaa* adds a sense of completion to the event (see Butt, 1995). While in (11a), an appropriate response to the question could be ‘a little,’ in (11b), the presence of *jaa* makes the answer infelicitous. Throughout this paper, we indicate this sense of completion through the parenthetical adverb, *completely*.

- (11) a. Q: baraf **pigl-i?** A: tHoRi-si.  
 Q: ice.NOM melt-PFV.F.SG A: little  
 ‘Did the ice melt? A little.’

<sup>1</sup>Interestingly, the set of unergatives and transitives *jaa* co-occurs with differs between Hindi and Urdu. For example, while Urdu does not allow *jaa* with *bHag-a-na* ‘to cause to run’ (Table 3, viii), Hindi does, *bHag-a jaa-na* ‘to cause to run’, meaning ‘to convince to elope’. Thanks to Sakshi Bhatia for this example.



### 2.3 Case

When *jaa* occurs with unaccusatives, the case of the internal argument does not change (15). Accusative case is not available both with and without the light verb. If *jaa* was an active Voice head, we'd expect the presence of *jaa* to allow accusative case as in other active constructions.

- (15) a. baraf(\*=**ko**) pigl-i?  
ice(\*=ACC) melt-PFV.F.SG  
'Did the ice melt?'  
b. baraf(\*=**ko**) pigal gai.  
ice(\*=ACC) melt *jaa*.PFV.F.SG  
'The ice melted [completely].'

In contrast, accusative case is available in transitive constructions with (and without) *jaa*. If *jaa* was a non-active Voice head with transitives, we'd expect accusative case to be unavailable on the direct object.<sup>3</sup> This is not the case (16).

- (16) Ahmed        **seb=ko**     kHa gya.  
Ahmed.NOM apple=ACC eat *jaa*.PFV.M.SG  
'Ahmed ate [up] the apple.'

### 2.4 Summary

In summary, *jaa* with unaccusative verbs has the following properties: it can occur freely (in fact, it appears to be required), it adds a sense of completion to the event, and it does not change the case of the internal argument (i.e. it does not make accusative case available). In contrast, *jaa* with unergative/transitive verbs has the following, and crucially, different properties: it cannot occur freely and appears to have an unpredictable distribution, the resulting interpretation is more similar to a particle verb construction than simply adding an aspectual flavor, and it does not change the availability of accusative case.

These facts, especially the contrast in the presence of accusative case and the required versus optional presence of *jaa* lead us to conclude that the *jaa* which occurs in unaccusative constructions is not the same *jaa* as occurs with unergatives/transitives. We argue that the *jaa* in unaccusatives is a more functional item than the *jaa* in unergative/transitive constructions. In the next section, we outline an analysis for *jaa* and its functionality, as it occurs with unaccusatives. This analysis cannot be extended to *jaa* with unergatives/transitives - we leave this gap open for future research.

---

<sup>3</sup>This follows the analysis of so-called 'accusative-preserving passives' as underlying active constructions, as proposed by (Kidwai, 2022b, to appear), meaning accusative case is never available in true Urdu passives.

### 3 LV *jaa* introduces an EA

We have shown that *jaa* with unaccusatives shows more functional properties than other LVs, including LV *jaa* with unergatives/transitives. This suggests that *jaa* occupies a different structural position from other LVs when it combines with unaccusatives (see §4 for discussion on other LVs). In this section, we argue that *jaa* occupies the functional head which introduces the external argument (EA). We call this head Voice. We show that *jaa* introduces an argument which passes subjecthood diagnostics, indicating that it is an EA. In particular, this argument shows the same syntactic properties as *by*-phrases in passives, suggesting that the two are in the same position, Spec of non-active Voice, and that *jaa* specifically occupies non-active Voice. We also show that this non-active Voice differs from the canonical passive Voice, and in fact, shows similarity to Voice in marked anti-causatives.

In Section 3.1, we discuss the properties of the EA introduced by *jaa*, in particular, comparing it to *by*-phrases. In Section 3.2, we provide a brief analysis of the Voice head occupied by *jaa*.

#### 3.1 Properties of EA in *jaa*-constructions

Adding LV *jaa* to an unaccusative verb allows introduction of an instrumental causer (17).<sup>4</sup>

- (17) **garmi=se / Rami=se** baraf pigal gai.  
heat=INS / Rami=INS ice.NOM melt *jaa*.PFV.F.SG  
'The ice melted [completely] because of the heat/Rami.'

Such instrumental arguments cannot be added otherwise (18).<sup>5</sup>

- (18) a. guRiya (\***Rami=se**) naach-i. (unergative)  
doll.NOM (\*Rami=INS) dance-PFV.F.SG  
'The doll danced (\*because of Rami).'
- b. Rami=ne (\***Omar=se**) baraf pigl-a-i. (causative)  
Rami=ERG (\*Omar=INS) ice melt-CAUS-PFV.F.SG  
'Rami melted the ice (\*because of Omar).'

Recall that there is a strong preference to include LV *jaa* with unaccusative verbs in neutral contexts (§2.1). The instrumental causer is also possible in contexts where *jaa* is absent (19).<sup>6</sup>

---

<sup>4</sup>Interestingly, it is not possible to add an OC with an animate internal argument.

- (i) (\***Bilal=se**) Ahmed mar gya.  
(\*Bilal=INS) Ahmed.NOM die *jaa*.PFV.M.SG  
'Ahmed died [completely] (\*because of Bilal).'

<sup>5</sup>(18b) is acceptable in some dialects as an indirect causative with the reading 'Rami made Omar melt the ice' (Saksena, 1980, 1982). While acceptable, *Omar* is a causee in this interpretation, not a causer.

<sup>6</sup>Thanks to an anonymous reviewer for (F)ASAL-13 for this example.

- (19) **Ahmed=se** glass gir-a **hi** tHa keh Mina  
 Ahmed=INS glass.NOM fall-PFV.M.SG FOC be.PST.M.SG that Mina.NOM  
 chiikh-ne lag-i.  
 scream-INF start-PFV.F.SG  
 ‘The glass had only just fallen because of Ahmed that/when Mina started  
 screaming.’

Given the general preference for *jaa* with unaccusative verbs, and the fact that instrumental causers cannot occur freely in other contexts, we take it that instrumental causers are directly correlated with *jaa*, and that the absence of *jaa* in cases like (19) is due to independent factors allowing the LV to be dropped.

In the rest of this section, we discuss the syntactic and interpretive properties of these instrumental causers.

### 3.1.1 Syntactic properties

The argument introduced by *jaa* bears instrumental case and is optional (20).

- (20) **(Rami=se)** baraf pigal gai.  
 (Rami=INS) ice.NOM melt jaa.PFV.F.SG  
 ‘The ice melted [completely] (because of Rami).’

These properties are characteristic of arguments introduced in the specifier of non-active Voice (Kidwai, to appear), such as *by*-phrases and causees.<sup>7</sup> Here, we compare the argument introduced by *jaa* to *by*-phrases.

- (21) a. baraf **(Rami=se)** pigl-a-i gai. (passive)  
 ice.NOM (Rami=INS) melt-CAUS-PFV.F.SG PASS.PFV.F.SG  
 ‘The ice was melted (by Rami).’  
 b. Rami=ne **(Omar=se)** baraf pigal-va-i. (indirect causative)  
 Rami=ERG (Omar=INS) ice melt-CAUS-PFV.F.SG  
 ‘Rami had the ice melted (by Omar).’ (≈ ‘Rami made Omar melt the ice.’)

*By*-phrases have been argued to be external arguments (EAs) on the basis of their behaviour with respect to subjecthood diagnostics (Mahajan, 1995; Srishti, 2011; Davison, 2015; Kidwai, to appear). There are three commonly used subject diagnostics in Urdu: anaphor binding, control into participial clauses, and pronoun obviation (see Davison, 2015, for an overview). Subjects but not objects are able to bind the anaphor, *apna* ‘self’s’ (22a), and control into participial *kar* clauses (22b). Subjects are not able to bind non-reflexive pronominal possessors, however, while objects are (22c). These tests are strictly associated with subjecthood, and are not specific to agents or base-generated external arguments. For example, promoted objects of passives and unaccusatives are also able to pass these tests (Kidwai, to appear).

<sup>7</sup>Several works have proposed that indirect causatives have a Voice-over-Voice construction, with Bhatt & Embick (2017) specifically arguing that indirect causatives in Urdu have an embedded passive Voice.

- (22) a. Rami<sub>i</sub>=ne **apni**<sub>i</sub> baraf pigl-a-i.  
 Rami<sub>i</sub>=ERG SELF<sub>i</sub> ice melt-CAUS-PFV.F.SG  
 ‘Rami<sub>i</sub> melted his<sub>i</sub> own ice.’
- b. Rami<sub>i</sub>=ne [ **PRO**<sub>i</sub> gHar jaa kar ] baraf pigl-a-i.  
 Rami<sub>i</sub>=ERG [ **PRO**<sub>i</sub> home.LOC go do ] ice melt-CAUS-PFV.F.SG  
 ‘Upon [Rami] going home, Rami melted the ice.’
- c. Rami<sub>i</sub>=ne **us**<sub>\*i/j</sub>=**ki** baraf pigl-a-i.  
 Rami<sub>i</sub>=ERG 3SG<sub>\*i/j</sub>=GEN ice melt-CAUS-PFV.F.SG  
 ‘Rami<sub>i</sub> melted his<sub>\*i/j</sub> ice.’

*By*-phrases pass some of these tests: they can bind anaphors (23a) and control into participial clauses (23b). However, *by*-phrases do not show pronoun obviation (23c).

- (23) a. Rami<sub>i</sub>=se **apni**<sub>i</sub> kHiRki toR-i gai.  
 Rami<sub>i</sub>=INS REFL<sub>i</sub> window.NOM break.CAUS-PFV.F.SG PASS.PFV.F.SG  
 ‘His<sub>i</sub> own window was broken by Rami<sub>i</sub>.’
- b. Rami<sub>i</sub>=se [ **PRO**<sub>i</sub> zor laga kar ] kHiRki toR-i  
 Rami<sub>i</sub>=INS [ **PRO**<sub>i</sub> force put do ] window.NOM break.CAUS-PFV.F.SG  
 gai.  
 PASS.PFV.F.SG  
 ‘Upon [Rami] applying force, the window was broken by Rami.’
- c. Rami<sub>i</sub>=se **us**<sub>i/j</sub>=**ki** kHiRki toR-i gai.  
 Rami<sub>i</sub>=INS 3SG<sub>i/j</sub>=GEN window.NOM break.CAUS-PFV.F.SG PASS.PFV.F.SG  
 ‘His<sub>i/j</sub> window was broken by Rami<sub>i</sub>.’

This pattern is similar to that of dative subjects, which also bind anaphors, and control into participial clauses, but do not show pronoun obviation (Davison, 2004). To account for the behaviour of *by*-phrases, Kidwai (to appear) argues that they are generated in Spec-VoiceP, similar to ergative/nominative subjects. Hence, they are able to bind anaphors and control into participial clauses, but do not move further to Spec-TP because they have inherent case, similar to dative subjects. Therefore, they do not show pronoun obviation.<sup>8</sup> Similar explanations have been proposed for the behaviour of dative subjects by Davison (2004) and Poole (2016).

Returning to *jaa*-constructions, instrumental causers also pass two out of three subject diagnostics: they show anaphor binding (24a) and control into participial clauses (24b) (contra Bhatt & Embick, 2017, fn.22), but do not show pronoun obviation (24c). In other words, the argument introduced by *jaa* shows the same behaviour as other low subjects, namely *by*-phrases and dative subjects.

<sup>8</sup>There is a long line of literature which argues that subject properties are not associated with a single subject position, but rather are spread across multiple subject positions (see McCloskey, 1997, and following).

- (24) a. Rami<sub>i</sub>=se **apni**<sub>i</sub> kHiRki TooT gai.  
 Rami<sub>i</sub>=INS REFL<sub>i</sub> window.NOM break jaa.PFV.F.SG  
 ‘His<sub>i</sub> own window broke [completely] because of Rami<sub>i</sub>.’
- b. Rami<sub>i</sub>=se [ **PRO<sub>i</sub>** ball maar kar ] kHiRki TooT gai.  
 Rami<sub>i</sub>=INS [ **PRO<sub>i</sub>** ball hit do ] window.NOM break jaa.PFV.F.SG  
 ‘Upon [Rami] hitting the ball, the window broke [completely] because of Rami.’
- c. Rami<sub>i</sub>=se **us<sub>i/j</sub>=ki** kHiRki TooT gai.  
 Rami<sub>i</sub>=INS 3SG<sub>i/j</sub>=GEN window.NOM break jaa.PFV.F.SG  
 ‘His<sub>i/j</sub> window broke [completely] because of Rami<sub>i</sub>.’

We take the subjecthood diagnostics in conjunction with the facts about case and optional realisation to mean that instrumental causers are introduced in the specifier of non-active Voice in *jaa*-constructions.

### 3.1.2 Interpretive properties

Although the instrumental argument in *jaa*-constructions shows the same syntactic properties as *by*-phrases, its interpretive properties are significantly different.

Like ergative/nominative subjects of actives, *by*-phrases can be either volitional or non-volitional. This can be demonstrated using agency tests. *By*-phrases are compatible with purpose clauses (25), and with both volitional and non-volitional adverbs (26).

- (25) kHiRki Rami=se [ **gHar=mein daakhil ho-ne** ]=**ke** liye toR-i  
 window.NOM Rami=INS [ house=LOC enter be-INF ]=GEN for break-PFV.F.SG  
 gai.  
 PASS.PFV.F.SG  
 ‘The window was broken by Rami to enter into the house.’
- (26) Rami=se kHiRki **ghalti=se** / **jaan=ke** toR-i  
 Rami=INS window.NOM mistake=INS / know=GEN break-PFV.F.SG  
 gai.  
 PASS.PFV.F.SG  
 ‘The window was broken by Rami by mistake / on purpose.’

On the other hand, the instrumental argument with *jaa* can only be non-volitional. It is not compatible with purposes clauses (27), and is only compatible with non-volitional adverbs, and not volitional ones (28).

- (27) \* kHiRki Rami=se [ **gHar=mein daakhil ho-ne** ]=**ke** liye TooT  
 window.NOM Rami=INS [ house=LOC enter be-INF ]=GEN for break  
 gai.  
 jaa.PFV.F.SG  
 ‘The window broke because of Rami so that he could enter the house.’

- (28) Rami=se kHiRki      **ghalti=se** / #**jaan=ke** TooT gai.  
 Rami=INS window.NOM mistake=INS / #know=GEN break jaa.PFV.M.SG  
 ‘The window broke because of Rami by mistake / #on purpose.’

Thus, instrumental arguments in *jaa*-constructions are always interpreted as causers – never as agents.<sup>9</sup> Table 4 summarises the syntactic and interpretive properties of these arguments.

	<b>ERG/NOM subjects</b>	<b>By-phrase</b>	<b>INS causer</b>
a. Realisation	obligatory	optional	optional
b. Case	ERG/NOM	INS	INS
c. Anaphor binding	✓	✓	✓
d. Control into participials	✓	✓	✓
e. Pronoun obviation	✓	×	×
f. Interpretation	volitional/ non-volitional	volitional/ non-volitional	non-volitional only

Table 4: Syntactic properties of argument introduced by *jaa*.

### 3.1.3 No implicit argument

As mentioned earlier, the instrumental argument introduced by *jaa* is optional, similar to *by*-phrases. A key property of short passives (i.e. passives without *by*-phrases) is that they have an implicit argument which can be detected through syntactic and semantic diagnostics. One syntactic diagnostic is adding *by itself*, a phrase which is only possible when there is no explicit or implicit argument. *By itself* is ungrammatical in short passives (29), which has been taken as evidence for the presence of an implicit argument.

- (29) \* kHiRki      **khud-ba-khud** toR-i      gai.  
 window.NOM self-with-self break.CAUS-PFV.F.SG PASS.PFV.F.SG  
 ‘The window was broken by itself.’

The implicit argument in short passives can also be detected through truth-conditional semantics. Short passives are judged as true only if there is an agent present in the context. For example, (30) is only true if there is an agent/causer which causes the event (i.e. the ice cream melted due to the actions of an agent or due to an event, not simply from the heat of the sun).

- (30) ice-cream      dHoop=mein pigl-a-i      gai.  
 ice-cream.NOM sun=LOC melt-CAUS-PFV.F.SG PASS.PFV.M.SG  
 ‘The ice cream was melted in the sun [by someone/something].’

<sup>9</sup>Based on their non-volitionality, Bhatt & Embick (2017, fn.22) categorise instrumental causers in *jaa*-constructions as ‘manner/means adjunct[s] and not related to a syntactically present agent’. They also report that this argument cannot control into participial clauses, in contrast to the judgements presented here.



However, there is no implicit argument in *jaa*-constructions when the instrumental argument is not realised. Firstly, *by itself* is ungrammatical in such cases (31).

- (31) kHiRki        **khud-ba-khud** TooT gai.  
window.NOM self-with-self    broke jaa.PFV.F.SG  
‘The window broke [completely] by itself.’

Secondly, these sentences are judged as true regardless of whether there is an agent/causer present or not (e.g. the ice cream melted due to general weather conditions with no agent/causer present)

- (32) ice-cream        dHoop=mein pigal gai.  
ice-cream.NOM sun=LOC        melt jaa.PFV.M.SG  
‘The ice cream melted [completely] in the sun.’

Therefore, there is no syntactic or semantic evidence for an implicit agent in these constructions.

### 3.2 Analysis

The subjecthood diagnostics discussed in Section 3.1.1 show that instrumental causers in *jaa*-constructions are EAs, similar to *by*-phrases. *By*-phrases have been argued to be in the specifier of non-active Voice, Spec-Voice<sub>NACTP</sub> (Baker et al., 1989; Collins, 2005; Roberts, 2019; for Hindi-Urdu, see Mahajan, 1990; Srishti, 2011; Kidwai, to appear). Based on the shared syntactic properties of *by*-phrases and instrumental causers in *jaa*-constructions, we argue that instrumental causers are in the same syntactic position as *by*-phrases, Spec-Voice<sub>NACTP</sub>. Consequently, this is clear evidence for Voice<sub>NACT</sub> in unaccusative *jaa*-constructions.<sup>10</sup>

The logic outlined above has been used frequently in the literature on anticausatives to argue for the presence of a non-active Voice head in marked anticausatives (Kallulli, 2006, 2007). In many languages, such as Greek and Albanian, marked anticausatives share morphology with passives, and have a morphologically identical argument, introduced by the same adposition or bearing the same case. Although unmarked anticausatives or simple unaccusatives have been argued not to have Voice (Alexiadou et al., 2015), *marked* anticausatives have been argued to have non-active Voice, similar to that in passives, based on shared morphology with passives, and the shared syntactic behaviour of *by*-phrases and oblique causers.

---

<sup>10</sup>It is worth noting that while a similar instrumental phrase is also possible with unergative/transitives, it differs in three important respects, as discussed by Kidwai (to appear, pp.193–94). Firstly, it is not possible with all unergative/transitive *jaa*-constructions. Secondly, where possible, it is not interpreted as a causer but rather as a source or adjunct. Finally, and most importantly, it does not pass any subject tests, which is strong evidence that it is not an EA in Spec-VoiceP. This is unsurprising given the presence of the ergative/nominative subject, which is an EA, as is also confirmed by its behaviour with respect to subject diagnostics. These facts confirm our initial conclusion that *jaa* occupies different functional heads in unaccusatives and in unergatives/transitives.

The *jaa*-constructions examined in this paper bear a striking similarity to marked anticausatives. Firstly, *jaa* is identical to passive morphology. Secondly, it only occurs with verbs that participate in the causative alternation, that is, unaccusative verbs. (As discussed in §2, the *jaa* in question here is only found with unaccusatives, and is different from *jaa* found with unergative and transitive verbs.) Finally, *jaa*-constructions also have an argument identical to *by*-phrases, and which exhibits the same syntactic behaviour. In light of this parallel between *jaa*-constructions and marked anticausatives, we propose that *jaa* can be analysed as an anticausative marker occupying non-active Voice.

This analysis of *jaa* explains its more functional behaviour in comparison to other LVs, as seen in Section 2. As a Voice head, it selects for a particular structure rather than lexical items, hence explaining its distribution with unaccusatives. This is similar to the distribution of passive Voice, which attaches to all causative/transitive verbs, and does not vary by lexical item. Likewise, this analysis also explains why *jaa* is obligatory with unaccusative verbs – it spells out a functional head which is essential to the anticausative structure. Once again, this is similar to passive *jaa*, which is also obligatory in passive structures due to the requirement for passive Voice in these constructions.

Despite the many similarities between passive Voice and Voice-*jaa*, the two also have some key differences, indicating that they cannot be the same Voice head. Firstly, as we saw in Section 3.1, unlike *by*-phrases, instrumental causers in *jaa*-constructions are obligatorily interpreted as non-volitional. This indicates that the range of possible  $\theta$ -roles that can be assigned by Voice-*jaa* is more limited than that in passives.<sup>11</sup> Secondly, as also discussed in Section 3.1, when the instrumental argument is not realised, there is no implicit argument. This means that unlike passive Voice, Voice-*jaa* does not dispense its external  $\theta$ -role when it does not project a specifier. As a result, there is no implicit argument, syntactically or semantically. A final point in favour of Voice-*jaa* being different from passive Voice is that the two select different forms of the main verb, showing that there are two different heads in play with different morphological selection. Table 5 summarises the key properties of the two constructions, as well as actives.

	<b>Actives</b>	<b>Passives</b>	<b><i>jaa</i>-constructions</b>
a. EA	obligatory	optional	optional
b. Case of EA	ERG/NOM	INS	INS
c. $\theta$ -role of EA	all external	all external	causer only
d. No EA projected	N/A	implicit EA	no implicit EA
e. Form of main verb	–	PFV	ROOT

Table 5: Properties of actives, passives, and *jaa*-constructions.

To conclude, *jaa* is an exponent of non-active Voice with unaccusatives. This explains its more functional behaviour in comparison to other LVs. As a Voice head, *jaa* can in-

<sup>11</sup>Many works, most notably Roberts (1987), have noted that *by*-phrases in passives can have the full set of  $\theta$ -roles available in actives.

roduce an EA in its specifier. This argument receives instrumental case from non-active Voice. Finally, Voice-*jaa* is different from that in passives. A formal analysis of the properties of this new Voice head is beyond the scope of this paper (see Kidwai, to appear, for a detailed analysis).

#### 4 Comparison to other LVs

As mentioned in Section 1.1, Butt & Ramchand (2005) show that different LVs in Hindi-Urdu occupy different heads. However, the properties of *jaa* that we've discussed in this paper (especially the EA-introducing properties) do not map onto their proposed verbal spine. In this section, we review the properties of other LVs in Hindi-Urdu, showing that (i) they do not occupy the same head as *jaa*, and (ii) they have their own flavors/effects based on their relative position in the verbal spine. Thus, *jaa* is clearly in a different functional position from other Urdu LVs.

##### 4.1 *de*

In this section, we focus on permissive *de* 'give', as seen in (33b). Many works have argued that permissive *de* introduces an EA (Butt & Ramchand, 2005; Butt et al., 2008; Suliman, 2015). In (33a), the main verb *chal-a-na* 'to drive' has two arguments: AGENT and THEME. Adding permissive *de*, as in (33b), adds an additional argument for the permission-giving event.

- (33) a. Jack=**ne** gaaRi chal-a-i.  
 Jack=ERG car drive-CAUS-PFV.F.SG  
 'Jack drove the car.'
- b. **Fran=ne** Jack=**ko** gaaRi chal-a-ne di.  
 Fran=ERG Jack=DAT car drive-CAUS-INF give.PFV.F.SG  
 'Fran let Jack drive the car.'

Butt et al. (2008, p.10) suggest that *de* adds an argument because it is derived from lexical verb which is a 3-place predicate. However, we can see that LV *de* is truly a valency-increasing unit by looking at ditransitive + *de* construction. In (34), *de* occurs with a ditransitive main verb, *bHej-na* 'to send'. The number of arguments increases from three to four upon adding *de*.

- (34) a. Sana=**ne** Omar=**ko** kitaab bHej-i.  
 Sana=ERG Omar=DAT book send-PFV.F.SG  
 'Sana sent a book to Omar.'
- b. **Rami=ne** Sana=**ko** Omar=**ko** kitaab bHej-ne di.  
 Rami=ERG Sana=DAT Omar=DAT book send-INF give.PFV.F.SG  
 'Rami let Sana send a book to Omar.'

Butt & Ramchand (2005) and Butt et al. (2008) propose that *de* occupies an EA-introducing head (Init in their terms; see (4)). Upon comparing the EA of permissive *de* constructions to (unaccusative) *jaa*-constructions, we see that permissive *de* does not occupy the same Voice head as *jaa*.

Firstly, the EA of permissive *de* constructions is obligatory, unlike the EA of *jaa*-constructions, which as we saw in Section 3.1.1, is optional (20), similar to *by*-phrases. Secondly, the EA of permissive *de* constructions bears ergative/nominative case, depending on aspect, as seen in (33b) and (34b), in contrast with the EA of *jaa*-constructions, which bears instrumental case, as we saw in Section 3.1.1. These two properties together are already strong evidence against an analysis of *de* as non-active Voice unlike *jaa* – as mentioned in Section 3.1.1, EAs of non-active Voice in Urdu are consistently optional, and when present, are marked with instrumental case.

In terms of syntactic properties, the EA of permissive *de* constructions passes all the subject diagnostics, similar to ergative/nominative subjects in other constructions. It can bind anaphors (35a) and control into participial clauses (35a), and it cannot bind non-reflexive pronominal possessors (35c), hence showing pronoun obviation.

- (35) a. Fran<sub>i</sub>=ne Jack=ko **apni**<sub>i</sub> gaaRi chal-a-ne di.  
 Fran<sub>i</sub>=ERG Jack=DAT REFL<sub>i</sub> car drive-CAUS-INF give.PFV.F.SG  
 ‘Fran let Jack drive her<sub>i</sub> own car.’
- b. Fran<sub>i</sub>=ne [ **PRO**<sub>i</sub> gHar aa kar ] Jack=ko gaaRi chal-a-ne  
 Fran<sub>i</sub>=ERG [ **PRO**<sub>i</sub> house.LOC come do ] Jack=DAT car drive-CAUS-INF  
 di.  
 give.PFV.F.SG  
 ‘Upon [Fran] arriving home, Fran let Jack drive the car.’
- c. Fran<sub>i</sub>=ne Jack<sub>j</sub>=ko **us**<sub>\*i/j</sub>=**ki** gaaRi chal-a-ne di.  
 Fran<sub>i</sub>=ERG Jack<sub>j</sub>=DAT REFL<sub>\*i/j</sub> car drive-CAUS-INF give.PFV.F.SG  
 ‘Fran<sub>i</sub> let Jack<sub>j</sub> drive \*her<sub>i</sub>/his<sub>j</sub> own car.’

This is unlike instrumental causers in *jaa*-constructions which do not pass the pronoun obviation subject test. This difference is unsurprising given the difference in case – assuming pronoun obviation is associated with Spec-TP, and that only arguments which receive case from T move to Spec-TP, EAs of permissive *de* are expected to differ from EAs of *jaa*-constructions given their difference in case.

The two types of EAs also differ with respect to interpretation. Recall that instrumental causers in *jaa*-constructions can only be interpreted as non-volitional (§3.1.2). EAs of permissive *de* can be interpreted as either volitional or non-volitional. They can license purpose clauses (36) and both volitional and non-volitional adverbs (37).

- (36) Fran=ne Jack=ko [ **PRO** gHar **jaa-ne** ]=**ke** liye gaaRi chal-a-ne  
 Fran=ERG Jack=DAT [ **PRO** house.LOC go-INF ]=GEN for car drive-CAUS-INF

di.

give.PFV.F.SG

‘Fran let Jack drive the car for going home.’

- (37) a. Fran=ne **jaan=ke** Jack=ko gaaRi chal-a-ne di.  
 Fran=ERG know=GEN do Jack=DAT car drive-CAUS-INF

give.PFV.F.SG

‘Fran let Jack drive the car on purpose.’

- b. Fran=ne **ghalti=se** kar Jack=ko gaaRi chal-a-ne di.  
 Fran=ERG mistake=INS do Jack=DAT car drive-CAUS-INF give.PFV.F.SG  
 ‘Fran let Jack drive the car by mistake.’

The above facts show that the EA of permissive *de* constructions does not behave like the EA of *jaa*-constructions, suggesting that *de* does not occupy the same non-active Voice head as *jaa*. The interaction between *de* and passivisation suggests that *de* does not occupy Voice at all. Many LVs can be passivised in Urdu (see §4.2). *Jaa* cannot be passivised (38b) – this is to be expected if *jaa* itself is an instantiation of a Voice head. On the other hand, permissive *de* can be passivised (39b).

- (38) a. baraf pigal gai. (active)  
 ice.NOM melt jaa.PFV.F.SG  
 ‘The ice melted.’

- b. \* baraf pigal gai gai. (passive)  
 ice.NOM melt jaa.PFV.F.SG PASS.PFV.F.SG

- (39) a. Fran=ne Jack=ko gaaRi chal-a-ne di. (active)  
 Fran=ERG Jack=DAT car drive-INF give.PFV.F.SG  
 ‘Fran let Jack drive the car.’

- b. Jack=ko gaaRi chal-a-ne di gai. (passive)  
 Jack=DAT car drive-INF give.PFV.F.SG PASS.PFV.F.SG  
 ‘Jack was allowed to drive the car.’

This suggests that *de* is not a Voice head at all, hence allowing passive Voice to stack on top of it. Suliman (2015) proposes an analysis along these lines for passives of permissive *de*, and argues that *de* is introduced below Voice. Assuming EAs are introduced by Voice, this characterisation of *de* leads to several questions regarding its EA-introducing properties. We leave this open for future research, but emphasise the point that even a seemingly EA-introducing LV does not occupy the same functional head as *jaa*, underling its unique functionality.

## 4.2 *le*

As mentioned in Section 1.1, Butt & Ramchand (2005) argue that LV *le* occupies an intermediary head on the verbal spine. A pillar of this analysis is the fact that *le* can co-occur

with higher LVs but only in a specific order (see (5) above). Similarly, Sobolak (2023) argues that *le* occupies an Inner Aspect head between the lexical VP and the EA-introducing projection. The use of Inner Aspect is motivated by systematic aspectual changes to the event when *le* is present. Crucially, for the purposes of this paper, *le* is again reported to be in an intermediary head that does not introduce an EA.

Additionally, *le*-constructions can passivize (40). Because the passive morpheme in Voice and *le* can co-occur, *le* must not be in Voice. Therefore, *le* differs from *jaa* both in function and structural position: *Jaa* is in Voice and introduces an EA, while *le* is in a lower head (*v* or Inner Aspect) and does not introduce an EA.

- (40) khat        likH li-ya        gya  
 letter.NOM write take-PFV.M.SG PASS.PFV.M.SG  
 The letter was written [completely].

*Le* and *jaa* also have distinct selection properties. As shown in Section 2, *jaa* occurs with all unaccusative verbs, and only some unergative and transitive verbs. Crucially, Voice-*jaa* is only found with unaccusatives, and does not occur with unergatives/transitives. On the other hand, *le* only occurs with transitive predicates, as discussed by (Butt, 1995).

In following with the above literature, as well as our own observations about *le* in passives and its selectional properties, we maintain that *le* occupies a lower head than *jaa*.

### 4.3 *paR*

As observed in Butt (1995), *jaa* appears to have similar properties as LV *paR* ‘fall(en).’ Like *jaa*, *paR* can occur with unaccusative verbs (41).<sup>12</sup>

- (41) glass        gir **paR-a**.  
 glass.NOM fall fall-PFV.M.SG  
 ‘The glass fell [suddenly, accidentally].’

*PaR* with (some) unaccusative verbs can have an optional instrumental causer, similar to *jaa*-constructions. In (42), *Ahmed* is the instrumental-marked causer of the event.

- (42) (**Ahmed=se**) glass        gir paR-a.  
 (Ahmed=INS) glass.NOM fall fall-PFV.F.SG  
 The glass fell [suddenly, accidentally] (because of Ahmed).’

Instrumental causers in *paR*-constructions behave exactly like those in *jaa*-constructions with respect to the structural and interpretive properties discussed in Section 3. The properties are summarised in Table 6, with relevant examples for rows c–f shown in (43–46).

<sup>12</sup>*PaR* is also possible with some unergatives (e.g. *naach paR-na* ‘to dance suddenly, involuntarily’) and transitives (e.g. *likH paR-na* ‘to write suddenly, involuntarily’), although we do not discuss this here.

	<i>jaa</i>	<i>paR</i>
a. Realisation	optional	optional
b. Case	INS	INS
c. Anaphor binding	✓	✓
d. Control into participial clauses	✓	✓
e. Pronoun obviation	×	×
f. Interpretation	non-volitional only	non-volitional only

Table 6: Syntactic properties of argument introduced by *jaa* and *paR*

- (43) Anaphor binding (Table 6, row c):  
 Sana=se<sub>i</sub> **apni**<sub>i</sub> kursi gir paR-i.  
 Sana=INS<sub>i</sub> REFL<sub>i</sub>=GEN chair.NOM fall fall-PFV.F.SG  
 ‘Her<sub>i</sub> own chair fell [suddenly, accidentally] because of Sana<sub>i</sub>.’
- (44) Control into participial clauses (Table 6, row d):  
 Sana=se<sub>i</sub> [ **PRO**<sub>i</sub> pHisal kar ] kursi gir paR-i.  
 Sana=INS<sub>i</sub> [ PRO<sub>i</sub> slip do ] chair.NOM fall fall-PFV.F.SG  
 ‘Upon [Sana] slipping, the chair fell [suddenly, accidentally] because of Sana.’
- (45) Pronoun binding (Table 6, row e):  
 Sana=se<sub>i</sub> **us**<sub>i/j</sub>=**ki** kursi gir paR-i.  
 Sana=INS<sub>i</sub> 3SG<sub>i/j</sub>=GEN chair.NOM fall fall-PFV.F.SG  
 ‘Her<sub>i/j</sub> chair fell [suddenly, accidentally] because of Sana<sub>i</sub>.’
- (46) Non-volitional interpretation (Table 6, row f):
- # Sana=se [ **daakhil ho-ne** ]=**ke** **liye** khiRki TooT paR-i.  
 Sana=INS [ enter be-INF ]=GEN for window break fall-PFV.F.SG  
 Intended: ‘The window broke because of Sana to enter [suddenly, accidentally].’
  - Sana=se **ghalti=se** / #**jaan=ke** khiRki TooT paR-i.  
 Sana=INS mistake=INS / know=GEN window.PRO break fall-PFV.F.SG  
 ‘The window broke because of Sana [suddenly, accidentally] by mistake / #on purpose.’

However, despite these similarities, we argue that *paR* does not occupy the same non-active Voice head as *jaa* (see also Butt, 1995). First, the distribution of *paR* is much more restrictive than *jaa* – *paR* cannot occur with all unaccusative predicates (see Table 7).

Secondly, the distribution of instrumental causers in *paR*-constructions is also more restricted than in *jaa*-constructions. In (47), the instrumental-marked *Ahmed* is ungrammatical with *paR*.

	<b>Verb</b>	<i>jaa</i>	<i>paR</i>
i.	<i>mar-na</i> ‘to die’	✓	×
ii.	<i>pigal-na</i> ‘to melt’	✓	×
iii.	<i>gir-na</i> ‘to fall’	✓	✓
iv.	<i>nikal-na</i> ‘to come out’	✓	✓

Table 7: Distribution of *paR* with unaccusatives.

- (47) (??**Ahmed=se**) kapRay tHeli=se          nikal          paR-ay.  
 (??Ahmed=INS) clothes plastic-bag=INS come-out fall-PFV.M.PL  
 ‘The clothes came/fell out of the plastic bag [suddenly, accidentally] (??because of Ahmed).’

Once again, *jaa* proves to be more functional than other LV. Its distribution with unaccusative verbs does not vary by lexical item, and it is consistently able to introduce an EA.

#### 4.4 Summary

*Jaa* clearly exhibits more functional properties than other LVs in Urdu. This may or may not be linked to the number of uses *jaa* has in the language; for example, as a lexical verb (7) and passive morpheme (8), as well as a marker for (in)abilitative and necessity/prohibition reading (Davison, 1982). The tendency for grammaticalization of *jaa* in the language more generally may have influenced its functionality as a LV.

## 5 Conclusion

Variation of LVs in Hindi-Urdu is well-established in the literature. In this paper, we have shown that *jaa* occupies a different head than other LVs in Urdu – namely, the EA-introducing head, Voice. Additionally, *jaa* as a Voice head has distinct properties from both active and passive Voice heads. Therefore, we have shown both further variation across LV structural properties and variation within Voice head properties. *Jaa* as a Voice head introduces an instrumental-marked external argument, similar to anticausatives.

While not the focus of this paper, there is the obvious consideration of *jaa*’s status in the lexicon. To our knowledge, no analysis has been proposed that outlines the relationship between the different usages (lexical verb, passive, light verb). There are two most likely relationships: (i) homophony (ii) a single, underspecified entry. Butt (2010) and Butt & Lahiri (2003) support the second possibility for the connection between light verbs and lexical verbs more generally in Hindi-Urdu. Given that *jaa* has so many more functions than just light and lexical verb, this could be an especially interesting place to further investigate the connection between different functions of a verb.



## Acknowledgements

We'd like to thank the many who contributed to this project: Molly Diesing, Theresa Biberauer, participants of Cornell Syntax Circle, Cambridge SyntaxLab, Berkeley SSCircle, and participants at (f)ASAL-13. All errors are our own. Sana Kidwai's research is supported by the Harding Distinguished Postgraduate Scholarship.

## References

- Alexiadou, Artemis, Elena Anagnostopoulou & Florian Schäfer. 2015. *External Arguments in Transitivity Alternations: A Layering Approach* Oxford Studies in Theoretical Linguistics. Oxford University Press online edn. doi:10.1093/acprof:oso/9780199571949.001.0001.
- Alsina, Alex & Smita Joshi. 1991. Parameters in causative constructions. In John Doe (ed.), *Chicago linguistics society (CLS)*, vol. 27, 1–16. Chicago, IL: Chicago Linguistics Society.
- Alsina, Alex, Tara Mohanan & KP Mohanan. 2005. How to get rid of the COMP. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG05 conference*, 21–41. Stanford, CA: CSLI Publications.
- Baker, Mark, Kyle Johnson & Ian Roberts. 1989. Passive Arguments Raised. *Linguistic Inquiry* 20(2). 219–251.
- Bhatt, Rajesh. 2003. Topics in the syntax of modern Indo-Aryan languages: Causativization. Class handout.
- Bhatt, Rajesh & David Embick. 2017. Causative derivations in Hindi-Urdu. *Indian Linguistics* 78(1-2). 93–151. ISSN 0378-0759. [Draft in wide circulation since 2004].
- Butt, Miriam. 1995. *The structure of complex predicates*. CSLI Publications.
- Butt, Miriam. 2003. The light verb jungle. In Gulsat Aygen, Claire Bowerman & Conor Quinn (eds.), *Harvard working papers in linguistics: Papers from the GSAS/Dudley House workshop on light verbs*, vol. 9, 21–49. Cambridge, MA: Harvard University.
- Butt, Miriam. 2010. The Light Verb Jungle: Still Hacking Away. *Complex predicates: Cross-linguistic perspectives on event structure*.
- Butt, Miriam & Tracy Holloway King. 2006. Restriction for morphological valency alternations: The Urdu causative. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*, 235–258. Stanford, CA: CSLI Publications.
- Butt, Miriam, Tracy Holloway King & Gillian Ramchand. 2008. Complex Predication: How Did the Child Pinch the Elephant? In Linda Ann Uyeche & Lian-Hee Wee (eds.), *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, chap. 17, 231–256. Stanford: CSLI Publications. ISBN 978-15-75865-88-1.
- Butt, Miriam & Aditi Lahiri. 2003. Verbal passepartouts.
- Butt, Miriam & Gillian Ramchand. 2005. Complex Aspectual Structure in Hindi/Urdu. *The Syntax of Aspect* 117–153.

- Collins, Chris. 2005. A smuggling approach to the passive in English. *Syntax* 8(2). 81–120.
- Dalrymple, Mary. 2001. *Lexical functional grammar*. New York, NY: Academic Press.
- Davison, Alice. 1982. On the Form and Meaning of Hindi Passive Sentences. *Lingua* 58. 149–179.
- Davison, Alice. 2004. Non-nominative subjects in Hindi-Urdu: VP structure and case parameters. In Peri Bhaskararao & Karumuri Venkata Subbarao (eds.), *Non-nominative subjects I*, vol. 60 Typological Studies in Language, chap. 7, 141–168. Amsterdam/Philadelphia: John Benjamins Publishing.
- Davison, Alice. 2015. Hindi-Urdu: Central Issues in Syntax. In Artemis Alexiadou & Tibor Kiss (eds.), *Syntax - Theory and Analysis: An International Handbook*, vol. 2, chap. 42, 1478–1518. Berlin: De Gruyter Mouton.
- Dowty, David. 1991. Thematic proto-rules and argument selection. *Language* 67. 547–619. <https://doi.org/10.1353/lan.1991.0021>.
- Grimshaw, Jane & Armin Mester. 1988. Light verbs and theta-marking. *Linguistic Inquiry* 19(2).
- Harley, Heidi. 1995. *Subjects, events, and licensing*. Cambridge, MA: Massachusetts Institute of Technology dissertation. <http://hdl.handle.net/1721.1/11073>.
- Hook, Peter Edwin. 1974. *The compound verb in Hindi* The Michigan series in South and Southeast Asian Studies. Ann Arbor, MI: The University of Michigan.
- Jespersen, Otto. 1965. *A modern english grammar on historical principles, part vi, morphology*. BaGeorge Allen and Unwin Ltd.
- Kallulli, Dalina. 2006. A unified analysis of passives and anticausatives. In Christian Ebert & Cornelia Endriss (eds.), *Proceedings of the Sinn und Bedeutung 10*, vol. 44 1, 171–182. Berlin: ZAS Papers in Linguistics. doi:10.21248/zaspil.44.2006.308.
- Kallulli, Dalina. 2007. Rethinking the Passive/Anticausative Distinction. *Linguistic Inquiry* 38(4). 770–780. <https://muse.jhu.edu/article/224872>.
- Kaplan, Ronald M. & Jurgen Wedekind. 1993. Restriction and correspondence-based translation. In *Association for computational linguistics European chapter (EACL)*, vol. 6, 193–202. Utrecht, The Netherlands: Association for Computational Linguistics. <https://aclanthology.org/E93-1024>.
- Kidwai, Sana. 2022b. The Urdu Active Impersonal. In Samir Alam, Yash Sinha & Sadhwi Srinivas (eds.), *Proceedings of (Formal) Approaches to South Asian Languages 11 ((F)ASAL-11)*, Konstanz: University of Konstanz.
- Kidwai, Sana. to appear. *Voice, case and the external argument: The perspective from urdu*: University of Cambridge dissertation.
- Kratzer, Angelika. 1996. Severing the external argument from its verb. *Phrase Structure and the Lexicon* 109–137.
- Mahajan, Anoop. 1990. *The A/A-Bar Distinction and Movement Theory*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
- Mahajan, Anoop. 1995. ACTIVE Passives. In Paul Aranovich, William Byrne, Susanne Preuss & Martha Senturia (eds.), *The Proceedings of the 13th West Coast Conference on Formal Linguistics (WCCFL-13)*, 286–301. Stanford, CA: CSLI Publications.

- McCloskey, Jim. 1997. Subjecthood and Subject Positions. In Liliane Haegeman (ed.), *Elements of Grammar: Handbook in Generative Syntax* Kluwer International Handbooks of Linguistics, 197–235. Dordrecht: Springer. doi:10.1007/978-94-011-5420-8\\_5.
- Mohanan, Tara. 1988. Causatives in Malayalam. Unpublished Ms. Stanford University.
- Poole, Ethan. 2016. Deconstructing subjecthood. Ms. University of Massachusetts Amherst. <http://ethanpoole.com/papers/poole-2016-subjecthood.pdf>.
- Ramchand, Gillian. 2008. *Verb meaning and the lexicon: A first phase syntax*. Cambridge, UK: Cambridge University Press.
- Roberts, Ian. 1987. *The Representation of Implicit and Dethematized Subjects*, vol. 10 Linguistic Models. Dordrecht: De Gruyter Mouton. doi:10.1515/9783110877304.
- Roberts, Ian. 2019. *Parameter Hierarchies and Universal Grammar*. Oxford: Oxford University Press.
- Saksena, Anuradha. 1980. The Affected Agent. *Language* 56(4). 812–826. <https://www.jstor.org/stable/413490>.
- Saksena, Anuradha. 1982. Case Marking Semantics. *Lingua* 56(3–4). 335–343. doi:10.1016/0024-3841(82)90017-1.
- Sobolak, Frances. 2023. *The event structure of light verbs*: Cornell University dissertation.
- Srishti, Richa. 2011. *Hindi Verb Classes and Their Argument Structure Alternations*. Newcastle-upon-Tyne: Cambridge Scholars Publishing. ISBN 978-14-43831-48-2.
- Suliman, Faheem. 2015. Light verbs in hindi.

# A Preliminary Description of *vanthu* in Spoken Tamil

SUGANYA RAJENDRAN SCHMURA, *University of Rochester*

## ABSTRACT

Although *vanthu* has only been described as a filler word or discourse marker in spoken Tamil, this word is being used in various grammatical functions in colloquial Tamil speech. An analysis of the syntactic constraints and distribution of the occurrences of *vanthu* in one Tamil speaker shows that *vanthu* is being used primarily as a topic marker, then as a copula, a discourse marker, and a quotative. This analysis of *vanthu* suggests a change in progress occurring in colloquial spoken Tamil and confirms the necessity of more formal linguistic analysis to be done in this informal register of Tamil speech which can reveal phenomena that cannot be observed in formal or written registers of Tamil.

## 1 Introduction

This study provides a preliminary description of the Tamil word *vanthu*, previously only described as a discourse marker (Aiyer 2020), which seems to be taking on multiple grammatical functions in colloquial Tamil speech including as a copula, topic marker, and quotative.

The word *vanthu* seems to be expanding beyond its attested usage as a discourse marker to take on grammatical functions as well. Through an analysis of the occurrences of *vanthu* in the naturalistic speech of a Tamil-English bilingual woman, this study investigated the usage of this word in colloquial spoken Tamil in order to provide a preliminary analysis of its various emerging functions. This analysis revealed that *vanthu* was primarily used as a topic marker, then as a copula, then a discourse marker, and finally as a quotative. It was also found that *vanthu* is performing a different function than existing Tamil topic markers, copulas, and quotatives.

The Tamil word *vanthu* ('have come') is the past participle conjugation of *vaa* ('come') (Lehmann 1989: 204, Schiffman 1999: 154), as shown in example 1.

- (1) *na Rochester va-n-thu oru varushom aa-ch-i*  
1SG Rochester come-PST-PTCP one.INDF year become-PRS-PRF  
'It has been one year since I have come to Rochester.'

However, this word is frequently heard in colloquial Tamil speech outside of this meaning but has only been described by Aiyer (2020) as a filler word, roughly equivalent to the English 'like', which is used to bide for time to "allow the speaker to think of the right language" (3). While *vanthu* can be used in this manner as a discourse marker, new usages are emerging in colloquial spoken Tamil where it is serving various grammatical purposes as well.

In spoken Tamil, the word *vanthu* seems to be shifting from a content word to a grammatical word where it is now being utilized as a copula, topic marker, and quotative. This semantic bleaching and context broadening of *vanthu* points to a change in progress occurring in the language which has so far not been described in the literature. There is a lack of formal linguistic analysis done on colloquial Tamil speech and even less done on Tamil-English code-mixed

speech. However, this type of speech can illuminate interesting changes in progress occurring in the language that cannot be seen in formal speech or written language. The change in progress of *vanthu* will be described in this paper through utterances obtained from one Tamil speaker.

This paper has been organized as follows: Section 2 will provide more background information on existing Tamil copulas, topic markers, and quotatives, Section 3 will describe the methodology used in data gathering and analysis, Section 4 will show the results of the study through the distribution and description of *vanthu*, Section 5 will provide a discussion, and Section 6 will detail the next steps.

## 2 Background

Tamil is a highly diglossic language with the High variety (Literary Tamil), which can only be formally acquired, used primarily in writing and the Low variety (Standard Spoken Tamil), which is acquired naturally, used primarily in speech (Annamalai & Steever 2019). The phenomenon of the emerging usages of *vanthu* that will be discussed in this paper is only observed in colloquial spoken Tamil, which might explain why its description has so far not been found in grammars or linguistic analyses of written Tamil corpora or formal Tamil speech.

Another potential reason for the lack of study of *vanthu* could be due to the persistent language ideology among Tamil speakers, especially those who embody a prescriptivist view of the language, that *vanthu* is improper speech and should be avoided. This is a salient ideology that is reflected widely on online message boards like Quora, where one Tamil speaker acknowledges the pervasiveness of *vanthu* in Tamil speech by confirming that “it is a popular mannerism among Tamil speakers in casual conversation” while simultaneously denigrating its usage by saying that “it should be ignored. It has no meaning, no significance, and is used to fill up pauses while speaking and while groping for words to continue the conversation.” (Vishwanath 2021). This speaker later goes on to say that “there is no reason for using these words, and good speakers avoid these needless words and expressions.”

Another Tamil speaker believes that the usage of *vanthu* “means you lack confidence in what you speak or you are making up” (D 2021). A third speaker does recognize its function as a discourse marker, but only as a discourse marker, and explicates this by saying that usage of the word is “not required but perceived by the speaker as necessary to structure his thoughts” and that it can be used to “change the topic, or give new information about the same topic or even to contradict the listener”. This speaker believes that the “English equivalent of ‘*vanthu*’ is ‘actually’”, which suggests that it is used to emphasize or provide contrastive information (Ambrose 2021).

This cursory look at the general attitude of Tamil speakers towards *vanthu* online is often echoed in conversations with Tamil speakers offline as well. While the prevailing notion seems to be that *vanthu* is relegated to a filler word that plays no meaningful role other than as a discourse marker, this paper will show why this is not the case and how this word is actually employing various grammatical functions in its usage in Tamil speech. Sections 2.1 - 2.3 will summarize the existing copulas, topic markers, and quotatives in the Tamil language to provide

sufficient background for later explaining how *vanthu* employs these functions in a different manner in Section 4.

## 2.1 Tamil copulas

In this paper, copulas will be defined as a grammatical construct that occurs as a linking verb between a subject and a non-verbal predicate. Tamil is a zero-copula language for present tense sentences with nominal predicates (Asher 1982). Therefore, verbless clauses can exist in Tamil as exemplified below:

- (2) *naan auto-kaar-an*  
 1SG auto-NMLS-M  
 ‘I am an auto driver (male).’

There are two auxiliary verbs which can be used in their nonauxiliated forms as a copula in Tamil: *iru* (‘to be/sit/exist’) and *aaku* (‘to become’) (Lehmann 1989: 172). The first copula *iru* ‘‘predicates a temporary location or state of its subject’’ (Steever 2006: 169) which can be seen below with locational predicates (example 3) and existential predicates (example 4). In both cases, *iru* must overly agree with the subject.

- (3) *intak kiraama-tt-il muunru koovil (iru-kkir-atu)*  
 this village-OBL-LOC three temple IRU-PRES-3SN<sup>1</sup>  
 ‘There are three temples in this village.’ (Lehmann 1989: 173)

- (4) *naangkaL ungkaL aNiyaaka irunt-oom*  
 1PL.NOM 2SG.GEN team.DAT be-1PL  
 ‘We were your team.’ (Antonini 2012: 60)

However, *iru* cannot be used with nominal predicates as shown by the ungrammatical construction in example 5.

- (5) *kumaar vakkiil (\*irukkiratu/\*irukkiraan)*  
 Kumar lawyer (read-PRES.3SN/IRU.3SM)  
 ‘Kumar is a lawyer.’ (Lehmann 1989: 171)

Schiffman (1999) analyses *iru* as meaning ‘be (located)’ which can indicate possession along with the dative case as seen in example 6 or inflected with another verb as a present tense marker *-kk-* as seen in example 7.

---

<sup>1</sup> Lehmann glosses *iru* as its own category IRU, while others gloss it as ‘be’ or ‘is’.

(6) *en-akku panam irukku*  
me-DAT money is  
'I have money.'  
(Schiffman 1999: 29)

(7) *naan padi-kkr-en*  
1SG read-PRS-1SG  
'I am reading.'

Unlike *iru*, the verb *aaku*, is used to indicate a change of state of the subject, as seen below in example 8.

(8) *avan manitan a-v-aan*  
He-NOM man-NOM become-FUT-3SM  
'He will become a man.'  
(Steever 2006: 169)

The existing Tamil copulas as described above show three main behaviors: a) they must overtly agree with the subject, b) they are often marked for locative and existential function, and c) they occur at the end of a sentence or utterance. So, there are no copulas in Tamil that can serve as a purely grammatical function, such as a linking word, without having an associated lexical function.

## 2.2 Tamil topic markers

While topicalization can be achieved syntactically through the usage of marked word order by fronting the clause to be emphasized, a topic marker is a grammatical construct which explicitly marks the topic of a sentence or utterance. This phenomenon has been primarily discussed in Japanese (the thematic or contrastive topic marker *wa*) and Korean (the neutral or contrastive topic marker *nun*) as distinctive characteristics (Lee & Shimojo 2016).

In Tamil, topic markers are typically referred to as emphatic particles or markers, such as *-ee* and *-thaan*, which are used to “emphasize or focus attention on particular elements of the sentence, as well as to handle other discourse phenomena such as whether information is new, old but related to new, presupposed, and for other pragmatic functions” (Schiffman 1999: 192). Since Tamil does not employ word stress to show emphasis, it needs to use emphatic markers such as these to show emphasis in an utterance (Schiffman 1999: 192).

The examples below show the semantic difference between the usage of existing Tamil emphatic markers *-ee* (example 9) and *-thaan* (example 10).

(9) *inge-yee irukku*  
here-EMPH is  
'It's RIGHT HERE (rather than somewhere else)'.  
(Schiffman 1999: 192)

- (10) *inge-daan irukku*  
 here-EMPH is  
 ‘It’s here (and ONLY here)’. (Schiffman 1999: 192)

The main difference between these two emphatic markers is that *-ee* signifies a comparative sense of “one compared to many” while *-thaan* signifies “one and only one (compared to none)” (Schiffman 1999: 192). Schiffman (1999) mentions that these emphatic markers often “cannot be literally translated” (192) into English, even though a translation is often attempted with words like ‘only’, ‘itself’, and ‘just’. In other words, example 9 could be roughly translated as ‘it’s just here’ and example 10 as ‘it’s here only’.

### 2.3 Tamil quotatives

Quotatives are grammatical markers used to introduce reported speech (Tagliamonte & D’Arcy, 2004). The Tamil suffix *-nnu* is used as a quotative to indicate both direct and indirect reported speech, as seen in examples 11 (indirect speech) and 12 (direct speech), where the only difference in the two constructions is pronoun concord (Schiffman 1999: 152). The reported speech would appear within the square brackets in both examples.

- (11) *jaan [varraar] nnu sonnaaru*  
 John will-come QT said  
 ‘John said he would come.’ (Schiffman 1999: 152)

- (12) *jaan [naan varreen] nnu sonnaaru*  
 John [I will-come] QT said  
 ‘John said, “I will come”.’ (Schiffman 1999: 152)

This affix is also often used in constructions like “*appadi-ngraan* ‘that’s what he says’ or *appadi-mbaan* ‘that’s what he’ll say’” (Schiffman 1999: p. 56) or as the more general *appadi-ngraanga* ‘they say’ or ‘people say’ which occurs after finite verbs (152). As Tamil is a head-final language, quotatives occur immediately after the reported speech (Schiffman 1999: 151).

## 3 Methodology

In order to comprehensively investigate the behavior of *vanthu* in Spoken Tamil, I obtained recordings of naturalistic Tamil speech and analyzed the syntactic constraints and distribution of the occurrences of *vanthu* to better understand its function. Recordings (49:27 minutes of audio and 83:12 of video) of the naturalistic Tamil-English code-mixed speech of a 25-year-old woman were obtained via a discussion of the Tamil art form of *kolam*. These recordings yielded 5 regular occurrences of *vanthu* (usage as the past participle of ‘come’) and 56 non-regular occurrences of *vanthu* which were then analyzed for their syntactic constraints and distribution,



in order to help provide a more complete description of the word and its uses. Each utterance that the token was found in was transcribed in Tamil and then translated into English.

The speaker, Oviya, was born in Tamil Nadu but spent her childhood living in both the US and Tamil Nadu, so she is proficient in both Tamil and English and utilizes both Tamil and Standard American English phonology. The speaker and the researcher are both Indian Tamil-English bilinguals living in the United States. This is important to note as the researcher often code-mixed Tamil lexical items within her English speech during the interview process, which might have prompted more naturalistic, code-mixed speech from the speaker and therefore more uses of the colloquial *vanthu*.

#### 4 Results

Out of the 56 non-regular occurrences of *vanthu*, the word appeared in three main syntactic environments: a) between the subject and the subject complement, b) immediately following a spatial (e.g., here, there) or temporal (e.g., now, then) deictic marker, and c) immediately before reported speech. In the first two environments, the word employed a function as either a copula or a topic marker and in the last environment it employed a quotative function. It also appeared in its attested usage as a discourse marker as expected.

Table 1 shows the distribution of *vanthu* as it is used as a copula, topic marker, quotative, and discourse marker. Within each column, the syntactic environment that the token appears in is specified according to the aforementioned categories of a) between the subject and complement, b) following a spatial or temporal deictic marker, and c) preceding reported speech.

Copula		Topic Marker		Quotative	Disc
<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>c</i>	-
7	2	17	23	2	5
12.5%	3.6%	30.4%	41%		
16.1%		71.4%		3.6%	8.9%
<i>Total</i>					56

Table 1. Categorization and syntactic distribution of *vanthu*

From the summarized data shown in Table 1, it is clear that *vanthu* was primarily used in its topicalizing function. The second most used manner was as a copula, then as a discourse marker, and finally as a quotative. Sections 4.1 - 4.5 will further explain each of these categorizations with examples from the data.

## 4.1 Copula

Since Tamil is an agglutinative language, it is not surprising that while the lexical form of *va-n-thu* has three distinct morphemes (come-PST-PTCP) as shown in example 1 (Schiffman 1999), these morphological distinctions are lost when the word is semantically bleached and used in its grammatical functions. This is clearly shown in example 13 where *vanthu* is used as a copula to link the subject ‘one of them’ to the nominal predicate ‘mom’s school classmate’.

- (13) *One of them vanthu actually amma-oda school classmate*  
One of them COP actually mom-POS school classmate  
‘One of them was actually my mom’s school classmate.’

Like in example 13, example 14 below also shows *vanthu* being used as a copula to link the subject to its complement.

- (14) *a-ppo vanthu full-aa school*  
DIST-time COP full-ADVZ school  
‘At that time it was fully school.’

Here, it is important to note that *vanthu* is not overtly marked for tense. In the above examples where the tense of *vanthu* is interpreted as being past tense, this is gleaned from the larger discourse context of the utterance. In example 13, the speaker was talking about a story that happened in the past and in example 14, the distal temporal deictic marker *appo* prompts the past tense interpretation of *vanthu*. The word also never overtly agrees with the subject.

Example 15 shows *vanthu* acting as a copula to link the subject ‘where I was born’ with its complement ‘Erode’ while the marked SVO word order of this utterance also syntactically topicalizes ‘where I was born’ through fronting.

- (15) *na poran-tha-thu vanthu Erode-la*  
1SG born-PST-PTCP COP Erode-LOC  
‘Where I was born is in Erode.’

In the majority English utterance in example 16, *vanthu* occurs immediately before the English past tense copula ‘was’, seeming to repeat or emphasize the copular function of linking the subject ‘my second language’ to the nominal predicate ‘Hindi’. Again, *vanthu* is not marked for tense here but it is interpreted as past tense due to the usage of the past tense ‘was’ immediately afterwards.

- (16) *aana my second language vanthu was Hindi*  
but my second language COP was Hindi  
‘But my second language was Hindi.’

Since existing Tamil copulas cannot be used in constructions with nominal predicates, the usages of *vanthu* in the examples above show that it is filling this gap in Tamil by performing the function of a copula that can link a subject with a nominal predicate.

Unlike existing Tamil copulas (*iru, aaku*) which overtly agree with arguments and subjects (Antonini 2012: 60) and typically appear at the end of a sentence, following the default SOV position in Tamil, *vanthu* never appeared at the end of an utterance and it also never overtly agreed with subjects and arguments. Thus, it seems to only serve a grammatical purpose without the locative and existential functions that existing Tamil copulas typically employ. The lack of subject agreement and inflection for tense of *vanthu* shows that it is acting in a purely grammatical function as a linking verb, and the complementary distribution of *vanthu* (which never occurs utterance-finally) with Tamil copulas (which only occur utterance-finally) shows that the word *vanthu* is performing a different function than existing Tamil copulas.

#### 4.2 Topic marker

Among non-regular usages of *vanthu*, the most prevalent function was as a topic marker. Below, example 17 shows *vanthu* being used in this manner by topicalizing the immediately preceding clause ‘In Erode’, signaling to the interlocutor that the focus of this utterance is ‘Erode’. This utterance follows the canonical SOV word order in Tamil so there is no fronting or otherwise marked word order to syntactically signal topicalization, which means that only *vanthu* is performing a topicalization function to bring focus to ‘Erode’.

- (17) *na Erode-la vanthu romba naal illa*  
 1SG Erode-LOC TOP many day be.NEG  
 ‘In Erode, I was not there for much time.’

Examples 18 and 19 show two utterances that were spoken in succession. These two utterances were produced mainly in English but show *vanthu* being used to topicalize ‘*appa*’ and ‘Dad’ in both utterances.

- (18) *appa vanthu lived in Mumbai almost seven years*  
 Dad TOP lived in Mumbai almost seven years  
 ‘Dad lived in Mumbai almost seven years.’

- (19) *So Dad vanthu konjam speaks Hindi*  
 so Dad TOP little speaks Hindi  
 ‘So Dad speaks Hindi a little.’

While the existing Tamil emphatic marker *-thaan* can perform a contrastive function to contrast new information with old information (Schiffman 1999: 192), *vanthu* does not need to perform a

contrastive function along with its topicalizing function as shown in examples 18 and 19 where it is used to signal to the interlocutor to maintain focus on one topic.

Example 20 also shows *vanthu* in its topicalizing function by highlighting ‘the design’ as the emphasized topic in this utterance.

- (20) *Design vanthu i-volo fine-aa var-aa-thu*  
 design TOP PROX-much fine-ADVZ come-NEG-PTCP  
 ‘The design will not come out this finely.’

Example 21 shows how *vanthu* employs a different function as a topic marker compared to the Tamil emphatic marker *-thaan*. In this example, *vanthu* is used immediately after the third person plural pronoun as a topic marker to focus the interlocutor’s attention to the subject and the Tamil emphatic marker *-thaan* is used at the end of the utterance to emphasize the entire utterance. This shows the complementary distribution of *vanthu* and *-thaan*, since *vanthu* never appears at the end of an utterance, further suggesting that it plays a different role than the existing Tamil emphatic marker which can appear at the end of an utterance.

- (21) *avanga vanthu generation generation-aa Tamil Nadu-la iru-n-thovanga*  
 3PL TOP generation generation-ADVZ Tamil Nadu-LOC be-PST-3PL  
  
*a-ntha maathari-thaan*  
 DIST-that like-EMPH  
 ‘They were in Tamil Nadu for generations, that’s what they were like.’

One interesting function of *vanthu* as a topic marker is that it can occur more than once in one utterance, seeming to bring attention to two different topics as seen in example 22.

- (22) *a-ntha arisi maavu vanthu erumbu-kku vanthu saapaadu ah-um*  
 DIST-that rice.uncooked flour TOP ant-DAT TOP food be-FUT  
 ‘That rice flour will become food for ants.’  
 Lit. ‘That rice flour, for ants, will become food.’

It seems that the initially occurring *vanthu* highlights the most important topic (‘rice flour’), or the main topic of the entire utterance, and the second *vanthu* highlights the secondary topic (‘ant’), or the topic within the embedded clause. In other words, the fact that ‘rice flour’ serves as food is primarily important but it is also important to emphasize that it is becoming food for ‘ants’ specifically, and not for any other creature.

Unlike existing Tamil emphatic markers, *vanthu* employs a different function by allowing a non-contrastive, thematic focus on the topic of an utterance, thus functioning as a topic marker.

### 4.3 Quotative

There were two examples of *vanthu* being used as a quotative in the data. In example 23, although the utterance-final Tamil quotative *appadimbaaru* is already used to signal that the preceding clause is reported speech, *vanthu* is also used immediately before the clause to introduce the quote and therefore bookending the quote between two different quotatives.

- (23) *a-var vanthu nee i-nnai-kku enna pan-n-a a-thu Hindi-la*  
DIST-3MSG.HON QUOT 2SG PROX-day-DAT what do-PST-FV DIST-thing Hindi-LOC  
  
*soll-u a-ppadi-mb-aar-u*  
say-FV DIST-like-FUT-3MSG.HON-FV  
'He will be like, "Say what you did today in Hindi", that's what he'll say.'

In example 24, the speaker does not employ any existing Tamil quotatives, which would have to appear after the reported speech, but instead only uses *vanthu* to introduce the reported speech.

- (24) *Everyone ellarum vanthu kanna nalla iru-kkiy-a kanna*  
Everyone everyone QUOT sweetheart good be-PRS-Q sweetheart  
'Everyone will say, "sweetheart, are you doing good sweetheart?''

This inability of *vanthu* to appear at the end of an utterance unlike existing Tamil quotatives, similar to the same behavior with Tamil copulas and topic markers, suggests that it is performing a different function from existing quotatives, copulas, and topic markers in Tamil.

### 4.4 Discourse marker

As expected, *vanthu* does appear as its attested usage as a discourse marker in the data but it only appeared in this function 8.9% of the time. This suggests that a categorization as a discourse marker cannot be the primary description of *vanthu* in spoken Tamil.

The usage of *vanthu* as a discourse marker in the data was primarily determined by its syntactic position at the beginning of an utterance, meaning that there was no preceding word or clause for it to topicalize or link, or paralleling the use of the English discourse marker 'like'.

Example 25 shows an utterance that contains two usages of *vanthu* with different functions: the first occurrence is as a discourse marker which is clear as it occurs utterance-initially and therefore with nothing preceding it to topicalize or link, and the second occurrence is as a topic marker to bring attention to the subject of 'kolam powder'.

- (25) *vanthu kola-podi vanthu arisi maavu-la sei-vaa-nga*  
DISC kolam-powder TOP rice.uncooked flour-LOC do-FUT-3PL  
'They will make kolam powder using rice flour.'

Example 26 shows the speaker beginning the utterance in English with ‘but like’ and then immediately repeating this phrase in Tamil by saying ‘*aana vanthu*’ (‘but like’). This repetition suggests that *vanthu* is being used in a parallel manner to the English discourse marker ‘like’ in this scenario.

(26) *but like aana vanthu I feel like video eddu-thu a-ntha video ella paakam-bo-thu*  
 but like but DISC I feel like video take-PTCP DIST-that video all see-when-PTCP

*my parents probably took videos when I was speaking in English*  
 my parents probably took videos when I was speaking in English  
 ‘But like, I feel like when the videos were taken and we all look at the videos, my parents probably took videos when I was speaking in English.’

As shown above, *vanthu* is still used as a discourse marker in colloquial spoken Tamil but it was used in this manner relatively fewer times than its usage as a topic marker or copula.

## 5 Discussion

This preliminary analysis of *vanthu* in colloquial spoken Tamil shows that it is being used in Tamil speech in various grammatical capacities outside of its attested description as a discourse marker. The most salient function that *vanthu* performs in the speech of one Tamil speaker is as a topic marker, then as a copula, then a discourse marker, and finally a quotative. It is important to note that *vanthu* functioned significantly more as a topic marker and a copula than as a discourse marker, which suggests that a discourse marker is not its primary function as previously described in the literature. Therefore, there seems to be a change in progress happening in spoken Tamil where *vanthu* is being used in increasingly grammatical forms.

This change in progress occurring in *vanthu* could be motivated by many different factors, including syntactic influence from language contact. Silva-Corvalán (1998) has described how English syntactic influence can result in the use of parallel linguistic structures, such as in creole languages where “language contact can lead to linguistic structures which are typologically unique” (Deumert & Durrleman-Tame 2006: 2). Contact from geographically nearby Indo-Aryan languages could also influence Dravidian languages like Tamil (Sankaravelayuthan 2019), resulting in the use of parallel linguistic structures.

Within the scope of syntactic transfer, copulas, which was the second most salient way that *vanthu* was used in the data analyzed in this paper, seem to promote the emergence of parallel structures. For example, copula emergence has been described in L2 acquisition of English as one of the first morphological features acquired (Deumert & Durrleman-Tame 2006: 98) and copula deletion has been observed in spoken English due to syntactic transfer from Tamil (Herat 2005). The grammaticalization of verbs is an areal feature in many South Asian languages and there is an established pathway in Dravidian languages where verbs are grammaticalized into

auxiliaries or complementizers (Sankaraveleyathan 2019), so *vanthu* could be following a similar pathway in colloquial spoken Tamil.

A comparative linguistic approach can be taken in further studies to compare what is happening with *vanthu* in Tamil with other languages. For example, Nordoff (2010) proposes that Sri Lankan Malay is seeming to grammaticalize its word for ‘come’ (*dhaatang*) as a copula. However, Nordoff argues that this phenomenon independently developed in Sri Lankan Malay because the geographically closest contact languages (Sinhalese and Tamil) do not use a similar word as a copula. In fact, Nordoff explicitly states that *vanthu* is not used in a copular manner in Tamil, but it appears that the data he used to reach this conclusion is primarily from Lehmann’s 1989 Tamil grammar which only contains a more formal register of Tamil. This register of Tamil would not have any occurrences of *vanthu* in its grammatical functions as described in this paper as this phenomenon only occurs in colloquial spoken Tamil.

This reiterates the necessity of having more formal linguistic analysis done on colloquial Tamil speech, since many of the phenomena occurring in this register will not be found in formal speech which is often the primary register used for linguistic analysis of Tamil. Tamil-English code-mixed speech also requires more formal linguistic analysis since most of the literature on this topic focuses on their indexical functions or use as a discourse strategy (Canagarajah 1995, Krishnasamy 2015, Sanmuganathan 2020).

This study will be continued to provide a more complete description of *vanthu* in colloquial spoken Tamil through the analysis of more occurrences of the word in naturalistic Tamil speech. The grammaticalization of this word will also be investigated using the apparent time construct to simulate diachronic change using synchronic data. As research has shown, the quotative system is a prime area to observe ongoing language change (Cukor-Avila 2002, Tagliamonte & D’Arcy 2004). Therefore, special attention will be paid to the usage of *vanthu* in reported speech in future studies to explore how it performs a different function from existing Tamil quotatives and how that might signal language change in progress.

### **Acknowledgements**

I would like to thank my advisor, Professor Maya Ravindranath Abtahian, for her guidance and my consultant, Oviya, for her invaluable help and patience. Many thanks to the FASAL team for inviting me to present my work at the University of Michigan-Ann Arbor and to the audience at FASAL 13 and my colleague and friend Sreyoshi Basu for their generous feedback. Finally, I would like to thank my Amma and Appa for their never-ending support and for giving me the gift of Tamil as my *thaimozhi* (‘mothertongue’).

### **References**

Aiyer, R.K. 2020. The Use of Communication Strategies in a Multilingual Urban Workplace in Sri Lanka. *Proceedings of the Open University Research Sessions*.

Ambrose, Anbazhagan. “Why do we frequently use ‘vandhu’ (வந்து) in colloquial usage even if it’s not required most of the time?” (blog), *Quora*, June 21, 2021. [https://](https://www.quora.com/Why-do-we-frequently-use-vandhu-வந்து-in-colloquial-usage-even-if-its-not-required-most-of-the-time)

[www.quora.com/Why-do-we-frequently-use-vandhu-வந்து-in-colloquial-usage-even-if-its-not-required-most-of-the-time](https://www.quora.com/Why-do-we-frequently-use-vandhu-வந்து-in-colloquial-usage-even-if-its-not-required-most-of-the-time).

Annamalai, E. & Steever, Sanford B. 2019. Modern Tamil. In *The Dravidian Languages* (2nd ed.). Routledge.

Antonini, Lindsey Marie. 2012. *The Copula in Malayalam*. Athens, GA: University of Georgia dissertation.

Asher, Ronald Eaton. 1982. *Tamil. Lingua Descriptive Series, Vol. 7*. North Holland Publishing Co.

Canagarajah, Athelstan Suresh. 1995. Manipulating the context: The use of English borrowings as a discourse strategy by Tamil fish vendors. *Multilingua*, 4(1): 5-24.

Cukor-Avila, Patricia. 2002. She say, she go, she be like: Verbs of quotation over time in African American Vernacular English. *American Speech*, 77: 3-31.

D, Stan. “Why do we frequently use ‘vandhu’ (வந்து) in colloquial usage even if it’s not required most of the time?” (blog), *Quora*, June 21, 2021.

<https://www.quora.com/Why-do-we-frequently-use-vandhu-வந்து-in-colloquial-usage-even-if-its-not-required-most-of-the-time>.

Deumert, Ana & Durrleman-Tame, Stephanie. (Eds.). 2006. *Structure and Variation in Language Contact*. John Benjamins Publishing Company.

Herat, Manel. 2005. BE Variation in Sri Lankan English. *Language Variation and Change* 17(2): 181-208.

Krishnasamy, K. 2015. Code Mixing among Tamil-English Bilingual Children. *International Journal of Social Science and Humanity*, 5(9): 788-792.

Lee, EunHee & Shimojo, Mitsuaki. 2016. Mismatch of topic between Japanese and Korean. *Journal of East Asian Linguists* 25(81): 81-112.

Lehmann, Thomas. 1989. *A Grammar of Modern Tamil*. Pondicherry Institute of Linguistics and Culture.

Nordhoff, Sebastian. 2011. Having come to be a copula in Sri Lanka Malay: An unusual grammaticalization path. *Folia Linguistica*, 45(1).

Sankaravelayuthan, Rajendran. 2019. Grammaticalization of Verbs in Dravidian Languages. *Language in India*, 19(8): 429-491.

Silva-Corvalán, Carmen. 1998. On borrowing as a mechanism of syntactic change. In A. Schwegler et al., *Romance linguistics: Theoretical perspectives*, 225-246. Amsterdam: John Benjamins.

Schiffman, Harold F. 1999. *A reference grammar of spoken Tamil*. Cambridge University Press.

Steever, Sanford B. 2006. *The Tamil auxiliary verb system*. Routledge.

Tagliamonte, Sali, & D’Arcy, Alexandra. 2004). He’s like, she’s like: The quotative system in



Canadian youth. *Journal of Sociolinguistics*, 8(4), 493–514.

Vishwanath, Gopalkrishna. “Why do we frequently use ‘vandhu’ (வந்து) in colloquial usage even if it’s not required most of the time?” (blog), *Quora*, June 21, 2021. <https://www.quora.com/Why-do-we-frequently-use-vandhu-வந்து-in-colloquial-usage-even-if-its-not-required-most-of-the-time>.

# DOM in Kodava takk: a complex interaction among multiple factors

SAUROV SYED, *The University of Auckland*

CATHERINE LEE, *The University of Auckland*

## ABSTRACT

This paper presents novel data from Kodava takk (Dravidian), also known as Coorgi, which exhibits the well-attested syntactic phenomenon of Differential Object Marking (DOM). Crosslinguistically, objects which are differentially marked tend to be associated with features such as specificity and/or definiteness, humanness, animacy, or a combination of these. Well-known examples of specificity-driven DOM include Turkish (von Heusinger and Kornfilt 2005) and Senaya (Kalin 2018), whereas direct objects in Spanish (Ormazabal & Romero 2013) and Hindi (Dayal 2011, Bhatt & Anagnostopoulou 1996) receive differential marking on the grounds of animacy/humanness and specificity. As will be illustrated, this phenomenon is most definitely present in Coorgi, as the accusative case-marker does not always appear on direct objects. However, on the surface, there is no clear-cut featural split between objects which do and do not receive this case-marker. Instead, this differential marking is triggered by a complex interaction of multiple factors: animacy, specificity, number, humanness, and inherent lexical properties of verbs. This paper outlines the interactions which derive Differential Object Marking in Coorgi and offers a formal analysis to capture the empirical facts, which modifies Kalin's (2018) account where DOM is a result of nominal licensing. This paper not only provides complex novel data from an understudied and endangered language, but also deepens our understanding of this crosslinguistic phenomenon, and calls into question the role grammatical Number plays in Differential Object Marking.

## 1 Introduction

In this paper, we offer novel data pertaining to Differential Object Marking (DOM) as it appears in the understudied Dravidian language of Kodava takk, also known as Coorgi.<sup>1</sup> To gather this data, guided interviews and elicitation sessions were held with a native speaker consultant. Over the course of the interviews, it became apparent that DOM was present in this language, as the accusative case-marker was not obligatorily present on every direct object. However, there was no single link between any given feature and the presence/absence of the accusative case-marker. Instead, the appearance of the accusative case-marker appeared to be conditioned by the interaction between multiple factors: specificity, animacy, humanness, and grammatical Number. Furthermore, a handful of verbs appeared to override any other factors and dictate the case-marking of their objects. This paper will clearly set out how these factors interact and give rise to the presence or absence of the accusative case-marker on direct objects in Coorgi.

Differential Object Marking (DOM), a term introduced by Bosson (1985), is a crosslinguistic syntactic phenomenon whereby direct objects in a given language are marked differently. This can have syntactic or semantic motivations, and it can manifest in several different ways. In Kannada Turkish (von Heusinger & Kornfilt 2005), DOM is conditioned by specificity and there is a split between specific/definite objects which receive overt morphological case-marking (see 1), and

---

<sup>1</sup> Our native speaker consultant voiced her preference for this language to be referred to either as Coorgi or Kodava takk, and we will use Coorgi for the remainder of this paper.

non-specific objects which do not (see 2).

(1) ben    bir    kitab-ı    oku-du-m.  
I        a        book-ACC read-PST-1SG  
'I read a certain book.'

[von Heusinger and Kornfilt 2005:5]

(2) ben        bir    kitap    oku-du-m.  
I            a        book    read-PST-1SG  
'I read a book.'

[von Heusinger and Kornfilt 2005:5]

Morphological case-marking affixes are not the only way in which DOM manifests. In Senaya (Kalin 2018), this object marking takes the form of agreement markers on the verb; in Spanish, certain (usually human and specific) direct objects are marked with the preposition *a* (Rodríguez-Mondoñedo 2005, Ormazabal & Romero 2013).

As a well-attested crosslinguistic phenomenon, DOM has naturally been subject to a great deal of formal analyses. Traditionally, these have focused on DOM as a result of syntactic height or raising of the object (López 2012, Diesing 1992, Bhatt & Anagnostopoulou 1996), pseudo-incorporation of unmarked objects (Dayal 2011, Mohanan 1995), or semantic features such as affectedness (von Heusinger & Kaiser 2011, Tsunoda 1985, Fleuschhauer 2018). The relative merits and shortfalls of these approaches will be discussed in detail; this paper proposes that a modified version of Kalin's (2018) nominal licensing framework can best capture the empirical facts in Coorgi.

Section 2 will give a brief morphosyntactic background to the language of Coorgi and offer novel data displaying the phenomenon of DOM at work, outlining the five major factors – animacy, humanness, specificity, grammatical Number, and verb semantics – and presenting a number of empirical observations which can be gleaned from the novel data. Section 3 compares and contrasts existing theoretical frameworks (raising, pseudo noun incorporation) and to what extent these can be applied to the Coorgi data. In Section 4, a modified version of Kalin's (2018) nominal licensing analysis is proposed to account for the empirical facts presented in Section 2. Finally, Section 5 offers a conclusion and discusses the implications of this analysis.

## 2 Differential Object Marking in Coorgi

Coorgi is an understudied and endangered language spoken in the Coorg region of Karnataka, India. It is a Subject-Object-Verb language with a nominative-accusative case alignment. Nouns can take a variety of morphological affixes including the demonstrative prefixes *a* (distal) and *i* (proximal), the plural suffix *ya* and its allomorphs, and Case-marking suffixes which follow any plural-marking morphology. Cases in Coorgi include the nominative (null-marked), accusative, (*na* or *tna*, allomorphically *La* in the plural), dative (*əka*), genitive (*ra*, allomorphically *Da*), instrumental/locative (*lə*), and ablative (*ənjə*). Their uses are illustrated in (3)-(8) below:

(3) John        pustaka(tna)        oodəchi  
John        book(ACC)        read.PST.3SG  
'John read (a) book'

- (4) John meji-ra koDi-lə pustaka(tna) bechchatə  
 John table-GEN top-LOC book(ACC) keep.PST.3SG  
 ‘John kept (a) book on top of the table’
- (5) aŋa taanDa manæ-ənje pocchi  
 3SG 3SG(RFLX).GEN<sup>2</sup> house-ABL leave.PST.3SG  
 ‘He left his house’
- (6) John Mary- əkə pustaka(tna) koDətətə  
 John Mary-DAT book(ACC) give.PST.3SG  
 ‘John gave Mary (a) book’
- (7) John Bill-na kondatə \*(na)  
 John Bill-ACC kill.PST.3SG  
 ‘John killed Bill’
- (8) John nai-na noTəchi \*(na)  
 John dog-ACC see.PST.3SG  
 ‘John saw a dog’

Some verbs may take an object marked with the dative as opposed to the accusative, as shown in (9); per the native speaker, these can be used interchangeably.

- (9)
- a. John Bill-na kaDchatə \*(na)  
 John Bill-ACC bite.PST.3SG  
 ‘John bit Bill’
- b. John Bill- əkə kaDchatə \*(əkə)  
 John Bill-DAT bite.PST.3SG  
 ‘John bit Bill’

Looking at the above data, it is already apparent that the accusative case-marker is not always mandatorily present on direct objects in Coorgi. Per our native speaker, (3) was considered perfectly grammatically correct with or without the accusative case-marker. Examples (7)- (9), however, were considered grammatically incorrect if the direct object was not case-marked. In other words, Coorgi clearly exhibits DOM. Furthermore, looking at the difference between the direct objects, it would seem at first brush that this is DOM conditioned by animacy. However, even animate nouns are not always required to be accusatively case-marked; (10)-(11) below were considered acceptable per our speaker if there was a generic or habitual interpretation sought, such as that John were a dog-washer for a living.

---

<sup>2</sup> RFLX = Reflexive

(10) John     nai     kattuva  
 John     dog     wash.NonPST.3SG  
 ‘John washes dogs (he’s a dog washer)’

(11) John     nai     katto-injatə  
 John     dog     wash.ANT-Be.PST.3SG  
 ‘John used to wash dogs (it was his job)’

Furthermore, even though the direct object *pustaka* (‘book’) in (3) and (6) is clearly inanimate, it can optionally be case-marked without leading to the interpretation that the book is somehow animate. In other words, accusative case-marking is not linked solely to animacy. However, marking of the direct object did lead to a change in available interpretations. If the direct object *pustaka* (‘book’) was left bare, it could be interpreted as singular or plural; however, if the direct object marker *na* were used, the book could only be interpreted as singular. Here, we see a clear link between grammatical number and the presence of the accusative case-marker.

## 2.1 Number and Differential Object Marking

Bare nouns in Coorgi are number-neutral, and can be interpreted as singular or plural:

(12) John     mangæ     kattuva  
 John     mango     wash.NonPST.3SG  
 ‘John washes mango(es)’

However, this ambiguity disappears with the presence of the accusative case-marker. A nominal case-marked with *na*, as below, can only be interpreted as singular, suggesting that the singular is marked with a null morpheme as follows:

(13) naanə     pustaka-Ø-tna     oodiye  
 1SG     book-SG-ACC     read.PERF  
 ‘I have read a book’

(14) naanə     nai-Ø-na     chowTit-uLLə  
 1SG     dog-SG-ACC     kick.Ant-Be.NonPST.1SG  
 ‘I have kicked a dog’

In a similar vein, an object which is marked with plural morphology can only be interpreted as plural; furthermore, as (15) exhibits, it must be accusatively case-marked:

(15) John     nai-ya-La     kattuva     \*-(La)  
 John     dog-PL-ACC     wash.NonPST.3SG  
 ‘John washes dogs’

Examples (10)-(11) above, lacking the plural marker and any case morphology, were glossed as plural (‘dogs’) as the closest possible English translation for something approximating ‘dog-washer’. However, we argue that this bare noun is still technically number neutral; on any given

day, a dog-washer may wash one dog or multiple dogs. A parallel can be drawn to expressing in English that someone ‘reads books’ does not necessarily imply that the person reads more than one book on any day or multiple books at once; the nominal remains, semantically, unquantized, and instead forms part of a collective. For examples (10)-(11), as soon as the intended meaning is singular, referring to one dog, the accusative case-marker is required. The same can be said for (16), which our native speaker consultant advised was not clearly singular or plural. Overt plural marking morphology, or the accusative case-maker, would be required to disambiguate the grammatical number of the nominal *nai* (‘dog’) below:

- (16) John      *nai*      *kolluva*  
        John      dog      kill.NonPST.3SG  
        ‘John kills dogs’

Furthermore, looking at (8) and (10)-(11), it is further clear that this habitual/generic interpretation is not always possible, but rather is dependent on compatibility with the verb. The construction in (8) is simple past tense, referring to a single point in time - a telic construction. Examples (10), (11), and (16), however, are atelic/unbounded and compatible with multiple instances. We posit that this difference is key in explaining the differing case-marking requirements between (8) and the other examples.

As a brief interim summary, bare animate nouns are permissible only when intended in a vague, habitual context. As soon as this is not the intended interpretation, and the speaker wishes to refer to an event with a quantized object (i.e. one, or more than one), the plural marker or the accusative case-marker will be used. Either option will necessarily result in the overt accusative case-marking of the object. However, unlike animate nouns, inanimate nouns resist plural marking. Per our native speaker consultant, adding the plural marker to the direct object ‘house’ in (17) below is semantically off, and adding the plural marker to inanimate objects gives the impression that they are somehow animate. Instead, to yield a plural interpretation, it is preferred to use a quantifier such as ‘many’, as in example (18). Importantly, however, the case-marking requirement still holds over (17); if a plural morpheme is forced, the case-marker will obligatorily appear.

- (17) ?John      *manæ-ya-La*      *noTæchi*  
        John      house-PL-ACC      see.PST.3SG  
        Intended:      ‘John saw many houses’

- (18) John      *sumaar*      *manæ noTæchi*  
        John      many      house see.PST.3SG  
        ‘John saw many houses’

Comparing the above, we see two different requirements for plurality. Inanimate nominals resist plural marking, whereas animate nominals will almost always be specified for number either via the accusative case-marker (if singular) or via plural morphology (which will then introduce the accusative case-marker). We propose that animate nominals are by default marked for number, carrying the feature (+num) and projecting Num(ber)P within the nominal structure. However, +num is not present on animates only in limited scenarios – habit-like or generic occurrences where the nominal does not need to be quantized – such as (19) below:

(19) John    nai    katto-injatə  
 John    dog    wash.ANT-Be.PST.3SG  
 ‘John used to wash dogs (it was his job)’

Inanimate nominals are number-neutral by default, lacking +num and not projecting NumP. However, in limited scenarios, these nominals are quantized and will project NumP – for examples, in the causatives like (20) below. However, the object actually receives an ‘animate’ interpretation, which then naturally projects a NumP.

(20) John    seb-na/seb-ua-La                    paar-chiTTatə                    \*-(na)/-(La)  
 John    apple-ACC/ apple-PL-ACC   fly-cause.PST.3SG  
 ‘John made the apple/ the apples fly’

Already, we see an interaction between two factors: animacy and number. In addition to this, Coorgi also has DOM conditioned by specificity.

## 2.2 Specificity

Looking at (21) and (22) below, the animate direct object *nai* (‘dog’) can be interpreted as specific or nonspecific. In (23), however, a specific reading is not available; as in (19) above, the absence of the accusative case-marker is permitted only in deliberately vague and habit-like occurrences, and the direct object can only be interpreted as nonspecific.

(21) John    nai-na            katəchi    \*-(na)  
 John    dog-ACC    wash.PST.3SG  
 ‘John washed a (certain) dog’

(22) John    nai-ya-La    kattuva    \*-(La)  
 John    dog-PL-ACC wash.NonPST.3SG  
 ‘John washes (certain) dogs’

(23) John    nai    kattuva  
 John    dog    wash.NonPST.2SG  
 ‘John washes dogs’        (he’s a dogwasher)

Inanimate objects have a similar pattern, albeit with more reliance on verb semantics. The inanimate direct object *seb* (‘apple’) in examples (24) and (25), lacking the accusative case-marker, can only be interpreted as nonspecific. With the addition of the accusative case-marker in (25), the available readings are specific or nonspecific. Interestingly, though, whilst (25) can be interpreted as specific or nonspecific, (27) can be interpreted as specific only.

(24) John    seb    katəchi  
 John    apple wash.PST.3SG  
 ‘John washed apple/s’

(25) John seb-na katəchi  
 John apple-ACC wash.PST.3SG.  
 ‘John washed a (certain) apple’

(26) John seb tindatə  
 John apple eat.PST.3SG  
 ‘John ate apple/s’

(27) John seb-na tindatə  
 John apple-ACC eat.PST.3SG  
 ‘John ate a certain apple’

The speaker offered their opinion that perhaps this is due to the compatibility of the verbs ‘eat’ and ‘wash’ with animate/inanimate objects; wash could be equally used with animate or inanimate objects, whereas ‘eat’ would tend to be associated more with inanimate objects. This is an interesting point; however, as will be discussed later, the relative animacy of verbs in Coorgi is a complex issue that is not necessarily responsible for the case-marking of its objects.

Overall, some empirical generalisations can be made when taking into account both inanimate and animate direct objects. Objects that are accusatively case-marked can be interpreted as specific or nonspecific (with the exception of ‘eat’, as discussed above), and objects which lack accusative case-marking can only be interpreted as nonspecific. This can be simplified as a single statement: all specific objects in Coorgi must be overtly case-marked.

### 2.3 Humanness

As the above sections make clear, there is definitely an animacy split in Coorgi. However, there appears to be a three-way distinction between inanimate objects, nonhuman animate objects, and human objects. Whilst animate nonhuman objects are generally marked for case except when in habitual contexts, human objects cannot escape this case-licensing requirement and must always be overtly case-marked.

(28) John Bill-na chowTchi \*(na)  
 John Bill-ACC kick.PST.3SG  
 ‘John kicked Bill’

(29) John ponn-a-La chowTuva \*(La)  
 John woman-PL-ACC kick.NonPST.3SG  
 ‘John kicks women (as a hobby)’

Even where a habitual/generic context was forced for (29), such that John were a serial abuser who kicked women as a habit or a hobby (apologies for this terrible scenario), the direct object would be obligatorily case-marked and quantized. Therefore, we can generalise that all human objects in Coorgi must be case-marked.

### 2.4 Verb Semantics





‘John read a book’

(36) John Tolstoy(na) oodæchi  
John Tolstoy(ACC) read.PST.3SG  
‘John read Tolstoy(‘s work)’

(37) John Tolstoy-na oodæchi \*(na)  
John Tolstoy-ACC read.PST.3SG  
‘John read Tolstoy(‘s mind)’

Finally, as illustrated prior, the verb ‘eat’ is entirely dependent on other factors in order to yield accusative case-marking; more specifically, if the direct object is not specific or animate/human, marking it was considered ‘odd’ or unnecessary.

(38) John mangæ tindatə  
John mango eat.PST.3SG  
‘John ate a mango’

Our native speaker consultant offered up an interesting point of discussion for the role that verb semantics plays in the distribution of the accusative case-marker and suggested that the reasoning behind these different case-marking requirements could be due to the supposed animacy of their direct objects, similar to a ‘scale of preferred animacy’ (such as that proposed by von Heusinger and Kaiser, 2007). This kind of native speaker intuition is valuable and should not be discounted; however, (39)-(40) provide a problematic counterpoint for an animacy-based analysis. Working from a von Heusinger and Kaiser’s (2007) scale of preferred animacy, the verbs ‘kick’ and ‘hit’ should theoretically both be Class 1 (+human) verbs; however, whilst ‘hit’ requires its objects to be case-marked, ‘kick’ does not.

(39) John mangæ-na pojatə \*(na)  
John mango-ACC hit.PST.3SG  
‘John hit a mango’

(40) John mangæ(na) chowTchi  
John mango(ACC) kick.PST.3SG  
‘John kicked (a) mango’

Therefore, it appears that animacy cannot fully account for the difference in case-marking preference of these verbs. Another semantic option would be affectedness, as was considered by von Heusinger & Kaiser (2011) to be a compelling factor in their study of diachronic change in Spanish DOM; this was based on Tsunoda’s (1985) Affectedness Scale, as outlined below:

EFFECTIVE ACTION >> PERCEPTION >> PURSUIT >> KNOWLEDGE >> FEELING >>  
RELATION >> ABILITY

[Tsunoda 1985:388]

However, an analysis of verb semantics based on affectedness falls short on two main points. Firstly, ‘affectedness’ as a notion in itself is far vaguer than animacy; whilst animacy is a clear, definable, bivalent (+/-) feature, affectedness is traditionally analysed as more gradable, and its definition varies. Whilst Naess (2004) analysed affectedness as having links to animacy, definiteness, and saliency, von Heusinger & Kaiser (2011:593) defined affectedness as ‘a change in the direct object that is imposed by the main predicate’. Beaver (2010) proposed an alternative that ranked verbs based on quantized change (e.g. ‘kill’), non-quantified change e.g. (‘widen’), potential for change (e.g. ‘hit’, ‘kick’), and unspecified for change (e.g. ‘wait’). This ranking system shows the second problem with an affectedness-based analysis. With either Tsunoda’s (1985) or Beavers’ (2010) affectedness hierarchies, both ‘hit’ and ‘kick’ would occupy the same slot yet have different case-marking requirements in Coorgi. Furthermore, a verb such as ‘eat’, which results in the disappearance of the direct object, should be high on Beavers’ (2010) scale, yet prefers not to take a marked inanimate direct object in Coorgi. In summary, although both animacy and affectedness may play a role in object-marking preference, a more in-depth survey of a wider range of verbs would be needed to form a complete ranking system and establish any links to animacy, affectedness, or some other feature.

Differential Object Marking in Coorgi is a complex system which arises out of a multitude of factors: animacy, humanness, specificity, number, and verb semantics all play a role in the patterning of the accusative case-marker. However, from the data above, some empirical generalisations are clear:

1. An object which is accusatively case-marked is also specified for number (singular or plural).
2. If an object is animate, it must be case-marked except for habitual/generic readings.
3. Inanimate objects resist plural marking.
4. Inanimate objects are not generally case-marked except either when dictated by either verb semantics (making the inanimate object animate) or where a specific interpretation is sought.
5. Regardless of animacy, any specific object is obligatorily case-marked.
6. Human nouns are obligatorily case-marked.

Any formal framework of DOM would need to account for the above generalizations to satisfactorily account for this phenomenon in Coorgi.

### **3 Previous Theoretical Approaches to DOM**

As discussed in earlier sections, analyses based on semantic categories such as ‘affectedness’ (von Heusinger & Kaiser 2011 for Spanish) or ‘animacy’ (von Heusinger & Kaiser 2007 for Spanish;





object, and separated from the verb, a raising framework would predict that it should take obligatory case-marking; as it does not, this is evidence against a raising analysis.

- (49) John Mary-əkə or pustaka(tna) aichatə  
 John Mary-DAT one book(ACC) send.PST.3SG  
 ‘John sent Mary a book’
- (50) John or pustaka(tna) Mary-əkə aichatə  
 John one book Mary-DAT send.PST.3SG  
 ‘John sent Mary a book’
- (51) pustaka-tna John Mary-əkə aichatə \*(na)  
 book-ACC John Mary-DAT send.PST.3SG  
 ‘John sent a book to Mary’

A more likely explanation for why (51) is obligatorily case-marked is that it is due to the nominative case-marker being null; per our native speaker consultant, if the direct object were not marked, it leads to some confusion about who the subject is.

It is clear that raising makes at least two predictions which are not borne out by the Coorgi data. Following López (2012), the final prediction made by a raising-based analysis is that objects which are not case-marked must be incorporated into V. Furthermore, objects which project DP have uninterpretable Case and cannot incorporate; these should obligatorily be case-marked. In other words, in examples (49)-(50), the direct object *or pustaka* (‘a/one book’) should be case-marked, but in fact this is only optional, which is contrary to predictions. Similarly, in (52) below, *i-seb* (‘this apple’) has a demonstrative yet can optionally lack case-marking.

- (52) John i-seb(na) kattəchi  
 John PROX-apple(ACC) wash.PST.3SG  
 ‘John washed this apple’

Ultimately, due to the data points above, there is no one-to-one link between either specificity and object height, nor object height and case-marking, nor projection of DP and case-marking. Therefore, a raising analysis must be discarded when formally analysing the distribution of DOM in Coorgi.

### 3.2 Against Pseudo Noun Incorporation

Pseudo noun incorporation, or PNI, has previously been argued to account for differential object marking in Hindi (Dayal 2011, Mohanan 1995). Much like Coorgi, DOM in Hindi sees an animacy split; in Hindi, whilst inanimate objects have optional case-marking, animate objects only have optional case-marking if a determiner is lacking. If a determiner is present, animate objects are obligatorily case-marked. Objects which are not case-marked have been argued by Dayal (2011) and Mohanan (1995) to be incorporated.

- (53) Anu har kitaab/ har kitaab-ko paRhegi

	Anu	every book/ every book-ACC	read-FUT	
	'Anu will read every book'		[Dayal 2011: 127]	
(54)	Anu	*haar bacca/ haar bacce-ko	sambhaaltii	hai
	Anu	every child every child-ACC	look-after-IMP	be-PRS
	'Anu looks after every child'		[Dayal 2011:127]	

At first brush, PNI offers a tempting alternative to raising in that, per Dayal (2011), it has strong links to number neutrality. Dayal (2011) also identifies a link between telicity, aspect, and number neutrality, which we have shown to exist in Coorgi in terms of the habitual interpretations being compatible only with atelic verb constructions. Furthermore, PNI along the lines of Dayal (2011) allows scrambling, which would mean (50) above is not immediately ruled out. However, there are some outstanding issues which remain a problem for a PNI-based analysis.

Firstly, Dayal (2011) notes that there appears to be a split system in place for Hindi. As shown in (53)-(54), whilst inanimate DPs (e.g. *har kitaab* 'every book') do not have to be case-marked and can remain unincorporated, animate DPs must be case-marked and cannot be incorporated. Dayal (2011) proposes that, in order to save a PNI analysis, case-marking for inanimates must be optionally null; this causes some problems as it implies the existence of a null allomorph in free variation with the overt accusative case-marker. It is not clear why this cannot be the case for animate objects as well. In addition to this, whilst human nouns can lack case-marking and be 'incorporated' in Hindi, they are obligatorily case-marked in Coorgi; therefore, in order to properly apply this framework to the Coorgi data, there would need to be a further distinction specifying that human nouns cannot incorporate even when lacking a determiner.

Secondly, per Dayal (2011), the N+V incorporations are not possible for every verb, but rather only for 'prototypical' combinations. For example, one could have *laRkii-DhuunDhnaa* ('girl-finding') but not *laRkii-sulaanaa* ('girl-putting-to-sleep'). This has an advantage when compared to the Coorgi data as it offers a possible explanation for the impact of verb semantics; for example, one could argue that 'dog-washing' is a prototypical activity, which is why (23) can remain unmarked. However, to apply this to Coorgi, one would need to explain why 'dog-killing' in (56) is prototypical (allowing unmarked subjects) whereas 'dog-hitting' in (55) is not.

(55)	John	nai-na/nai-ya-La	poyyuva	*-(na)/(La)
	John	dog-ACC/dog-PL-ACC	hit.NonPST.3SG	
	'John hits a dog/dogs'			

(56)	John	nai	kolluva
	John	dog	kill.NonPST.3SG
	'John kills dogs'		

Finally, according to Dayal (2011), Hindi allows plural-marked nouns to incorporate, whereas in Coorgi, any plural-marked direct object must be accusatively case-marked. Overall, as explored

above, a pseudo noun incorporation analysis cannot satisfactorily capture the empirical facts seen in the Coorgi data.

#### 4 Nominal Licensing and Number: Proposed Analysis

In discounting the above frameworks, we instead argue that the empirical generalisations in Coorgi are best captured by a nominal-licensing-based analysis, as posited by Kalin (2018) for Senaya. In observing that specific nominals are banned in perfective aspect in Senaya, Kalin (2018) proposed a novel DOM framework based on the concept of nominal licensing. The crux of this nominal licensing framework is that DOM arises from two factors: the types of nominals that need case-licensing, and the location/identity of nominal licensors. Essentially, not all nominals are case-licensed, and nominals can go unlicensed for Case without crashing the derivation. Instead, Kalin (2018) follows Pesetsky & Torrego (2007) in drawing a distinction between unvalued, interpretable Case and unvalued, uninterpretable Case. Only the latter type will cause a crash in the derivation. This uninterpretable Case is introduced by certain features within the nominal structure; for example, in Senaya, this uninterpretable Case comes from the feature +specific. In addition to uninterpretable Case, Kalin (2018) posits two types of case-licensors. Obligatory licensors must license either their closest nominal ('inherent licensing') or the highest nominal which they c-command ('structural licensing'); here, a major advantage of Kalin's (2018) framework is the ability to explain why, in languages like Coorgi, subjects are always licensed for nominative Case, without drawing any distinction between subjects and objects. Secondary licensors, on the other hand, are only activated by the presence of uninterpretable Case, where the derivation would otherwise crash. This is implemented for Senaya (Kalin 2018) in order to explain the ban on specific objects within the so-called 'perfective' aspect by asserting that, whilst the +specific feature introduces uninterpretable Case, only the -perfective Aspect head can act as a secondary licensor; a +perfective Aspect head cannot. Therefore, specific objects in perfective constructions will introduce uninterpretable Case which cannot be licensed, causing a crash in the derivation. Per Kalin (2018), this nominal licensing analysis holds great crosslinguistic applicability as it allows for variation regarding the identity of the obligatory/secondary licensors and the features which introduce uninterpretable Case.

Here, we implement Kalin's (2018) nominal licensing framework, but with some adjustments. We assert that, in Coorgi, the obligatory and secondary licensors are  $T^0$  and  $v^0$  respectively; this accounts for why subjects always receive nominative case regardless of the features of the nominal in question. The features which introduce uninterpretable Case (uCase) are +specific, +human, and +number. Any one of these can introduce uCase, which percolates up to become a feature on the object DP; this will activate  $v^0$  in order to license the nominal and avoid a crash of the derivation. Whilst specificity and humanness seem relatively straightforward as features linked to DOM, the link between number and uninterpretable Case is contra Kalin (2018), and is proposed as a needed modification to the mechanism in Kalin's analysis to capture the Coorgi facts. In Kalin's (2018) analysis, all nominals should project NumP (in Kalin's proposal) and if NumP is associated with uCase, all objects should have uCase and receive case-marking as a result – this is not desired and Kalin thus rules out the possibility of a link between NumP and uCase. However, we have argued Coorgi bare nominals to be number-neutral, and we instead follow Déprez (2005) and Wiltschko (2005) in asserting that the NumP is not always projected in Coorgi nominals; we propose that only nominals with the feature +number will project NumP (thereby introducing uCase). The



animacy split is accounted for by stating that animate nominals will by default have the feature +number; therefore, they will almost always be case-marked regardless of specificity. The exception to this is in habitual/generic constructions, in which the nominals can remain number-neutral and will not be case-marked. Inanimate nominals, on the other hand, will lack +number by default; this not only explains why they are less likely to be case-marked than animates, but also establishes a link between animacy and the plural marker which has already been shown to exist (namely, that it is semantically ‘off’ to add a plural morpheme to inanimate nominals in Coorgi).

Applying this to the Coorgi data, we see that the empirical generalisations naturally fall out from this approach. In a scenario where the object is +specific, the nominal will have uCase introduced in its nominal structure. This will activate the secondary licenser  $v^0$ , which will license the object DP; as a result, this object DP will be differentially marked with the accusative case. This predicts that we should see obligatory DOM with all specific nominals, both -animate and +animate; this prediction is borne out by the Coorgi data. In a scenario where the nominal is animate, it will by default be specified for number; this +number feature will introduce uCase into the nominal structure, activating the secondary licenser and resulting in a case-marked object DP. This predicts that animate objects should receive DOM in specific and nonspecific interpretations – this, again, is borne out. In a habitual construction where the animate object is unquantized, it will lack the feature +number, and nothing will introduce uninterpretable Case or trigger case-licensing. This predicts that animate nominals will not be accusatively marked in habitual contexts; again, this is proven true. When the nominal is inanimate, it will by default be number-neutral. In the absence of other factors such as specificity, it should therefore not introduce uninterpretable Case. In this situation our analysis correctly predicts that these nominals can remain un-case-marked. In cases where the inanimate nominal is marked with the accusative case-marker, its interpretation is set as singular; therefore, in these cases, the nominal has been quantized and specified for number, introducing the feature +number which then introduces uninterpretable Case. The one remaining issue is verb semantics, and the exact nature of the case-marking requirement with certain verbs. We speculate that inanimate objects may not be number-neutral for certain verbs. This, again, requires more in-depth investigation to construct a full paradigm. However, whatever the link may be, what is certain is that these marked nominals have their number interpretation set in stone by the presence of the accusative case-marker: it is therefore clear that these nominals have been quantized, and therefore carry the feature +number, which has introduced uCase and activated the secondary licenser.

Finally, if a nominal is +human, it will have uCase introduced into its structure. This will activate the secondary licenser  $v^0$ , and the DP will be marked with the accusative case, regardless of whether it is specific or whether it is in a habitual/generic context. This prediction is borne out.

## 5 Conclusion

Over the course of this paper, we have presented novel data pertaining to DOM as it appears in the understudied and threatened language of Coorgi. We have shown that DOM arises not from one single feature, but from the complex interaction of multiple factors including animacy, specificity, number, humanness, and verb semantics. Due to this complexity, we have shown that previous theoretical approaches such as raising or pseudo noun incorporation, or analyses which rely on semantic categories such as animacy or affectedness, fail to account for the patterning of the

accusative case-marker in Coorgi. However, a modified version of Kalin's (2018) nominal licensing analysis, in which uninterpretable Case is introduced by nominals carrying the features +specific, +human, and +number, can yield the empirical generalisations presented in this paper. The findings presented here not only contribute to our understanding of the crosslinguistic phenomenon of DOM through the presentation of novel empirical data displaying its complexities, but also open up an interesting new avenue of investigation: the link between grammatical Number and DOM. A consideration of DOM as an interaction of multiple factors (for example, animacy and grammatical Number) may hold ramifications for other languages which display complex patterning of DOM and have so far resisted theoretical approaches.

## References

- Beavers, John. 2010. On affectedness. *Natural Language & Linguistic Theory* 29. 335-370.
- Bhatt, Rajesh & Elena Anagnostopoulou. 1996. Object shift and specificity: Evidence from ko-phrases in Hindi. *Chicago Linguistics Society (CLS)* 32. 11-22.
- Bossong, Georg. 1985. *Empirische universalienforschung: Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Gunter Narr Verlag.
- Dayal, Veneeta. 2011. Hindi pseudo-incorporation. *Natural Language & Linguistic Theory* 29. 123-167.
- Déprez, Viviane. 2005. Morphological number, semantic number and bare nouns. *Lingua* 115. 857-883.
- Diesing, Molly. 1992. *Indefinites*. Cambridge: MIT Press.
- Fleuschhauer, Jens. 2018. Animacy and affectedness in Germanic languages. *Open Linguistics* 4. 566-588.
- Heusinger, Klaus von & Jaklin Kornfilt. 2005. The case of the direct object in Turkish: Semantics, syntax and morphology. *Turkic Languages* 9. 3-44.
- Heusinger, Klaus von & Georg A. Kaiser. 2007. Differential object marking and the lexical semantics of verbs in Spanish. In Georg A. Kaiser & Manuel Leonetti (eds.), *Definiteness, Specificity and Animacy in Ibero-Romance Languages*, 85-110. Konstanz: Fachbereich Sprachwissenschaft der Universität Konstanz.
- Heusinger, Klaus von & Georg A. Kaiser. 2011. Affectedness and differential object marking in Spanish. *Morphology* 21. 593-617.
- Kalin, Laura. 2018. Licensing and differential object marking: The view from Neo-Aramaic. *Syntax* 21(2). 112-159.
- Lidz, Jeffrey. 2006. The grammar of accusative case in Kannada. *Language* 82(1). 10-32.
- López, Luis. 2012. *Indefinite objects: Scrambling, choice functions, and differential marking*. Cambridge: MIT Press.
- Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language & Linguistic Theory* 19(1). 153-197.
- Mohanan, Tara. 1995. Wordhood and lexicality: Noun incorporation in Hindi. *Natural Language & Linguistic Theory* 13(1). 75-134.
- Naess, Åshild. 2004. What markedness marks: The markedness problem with direct objects. *Lingua* 114(9). 1186-1212.
- Ormazabal, Javier & Juan Romero. 2013. Differential object marking, case and agreement. *Borealis: An International Journal of Hispanic Linguistics* 2(2). 221-239.

- Pesetsky, David & Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In Simon Karimi, Vida Samiian & Wendy K. Wilkins (eds.), *Phrasal and clausal architecture: Syntactic derivation and interpretation*, in honour of Joseph E. Emonds, 262-294. Amsterdam: John Benjamins.
- Rodriguez-Mondoñedo, Miguel. 2007. *The syntax of objects: Agree and differential object marking*. Storrs, CT: University of Connecticut dissertation.
- Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of Linguistics* 21. 385-396.
- Wiltschko, Martina. 2008. The syntax of non-inflectional plural marking. *Natural Language & Linguistic Theory* 26. 639-694.

## Processing of stacked NPs in embedded sentences in Malayalam

REVATHI SURESH

*St. Joseph's University, Bangalore*

### ABSTRACT

The paper examines the effect of stacked NPs in centre-embedded sentences in Malayalam and uses the experimental results to compare two theories of processing difficulty: Gibson (2000)'s Dependency Locality Theory (DLT) and Hale (2001)'s Surprisal Theory. Crucially, the study also looks at the definite vs. non-definite NP distinction in Malayalam in a stacked NP context. Non-definiteness in Malayalam is indicated by having the determiner *oru* before the noun; a non-definite NP in a centre-embedded sentence disrupts NP stacking. A self-paced reading task is run using single and double-embedded sentences in Malayalam with the embedded NP alternating between definite and non-definite conditions. The test is designed to determine if (i) stacked NPs in a centre-embedded sentence result in a processing difficulty (ii) having *oru* preceding an NP affects processing. The results showed that processing difficulty increases with the addition of each NP; NPs from the embedded clauses take longest to be read. When the embedded NP is preceded by *oru*, reaction time significantly drops at *oru* and the following NP, indicating that the determiner facilitated integration of the NP, disrupting NP stacking. The results were compared against the predictions of DLT and Surprisal models. We found that the anticipation based Surprisal account best accounted for the results for Malayalam.

### 1 Introduction

Sentence processing studies have established that Case is crucial in driving sentence parsing in head-final languages like Malayalam. Case indicates the relationship between a noun phrase and its role in relation to verbs. Malayalam being a head-final language uses Case-marking suffixes to express this relationship. Case affixes provide information that enable the parser to build structures that incorporate the Case-carrying NPs as well as predict incoming verbs and their argument structures (Konieczny, 2000; Konieczny & Döring, 2003). However, it is still not very clear as to whether their predictive capacity holds in embedded structures, where three or more noun phrases are stacked one after the other, and where the subject of the main clause is separated by the embedded clause (1).

- (1) Radha Raju-vinooDə [novel vaayikk-aan] paranj-u  
Radha.NOM Raju-DAT [novel.NOM read-NF] tell-PAST  
'Radha told Raju to read the novel' [Centre-embedded]

Contexts like (1) may either (i) overload working memory by having to store too many items or handle interference due to similar items or (ii) further increase the predictive capacity of the parser by providing more information about what items are yet to come. In this paper, I will also look at (i) stacked NP contexts in embedded sentences, and (ii) what happens when a determiner (*oru*) interrupts the stacking and it is followed by a non-definite NP. Experimental data from the study will be used to examine how Case-marking functions and determine if it still retains its predictive capabilities in these contexts. The broader aim of the study is to use the experimental findings to evaluate existing sentence processing models to identify some aspects or criteria that are crucial in accounting for processing data from SOV languages like Malayalam.

## 2 Embedded structures, definiteness and the processing models

This section introduces the two factors examined in the study, embedded structures and definitions, and crucially how they may help in determining a best fit model for Malayalam. The two approaches in consideration here are (i) memory-based models and (ii) anticipation-based models.

### 2.1 Processing of Embedded Structures in Malayalam

An embedded sentence contains an independent main clause and one or more dependent or subordinate clauses. A simple example would be (2), where the complement clause within square brackets is embedded on the right of the main clause.

(2) Radha said [<sub>CP</sub> that she wants to buy a new book].

Embedded sentences can be more complex when there is more than one level of embedding. Compare sentences (3a) and (3b), which have one and two levels of embedding respectively.

(3) a. *Single-embedded*

Radha found the ball [<sub>1</sub> that the boy who juggles] lost

b. *Double-embedded*

Radha found the ball [<sub>1</sub> that the boy who juggles [<sub>2</sub> bottles that have yellow stripes on it]] lost

The possibility of embedding technically allows for recursion with infinite levels of embedding which are grammatical but may be difficult to comprehend after a point (4).

(4) a. The rat [that the cat [that the dog bit] chased] died.





What are the implications of definiteness on sentence or phrase processing? Is there a difference between non definite and definite NPs? What motivates this difference, a structural factor or a pragmatic factor? Vasishth (2003) looked at processing of definiteness in Hindi, where the definiteness NP has an overt objective case marking *-ko*, and the nondefinite is indicated with a null marking. Vasishth (2002) found definite NP marked with *-ko* to be more difficult to process and thereby less acceptable than a bare indefinite NP in the same position. He attributes this processing difficulty to the position in which the NP occurred, the direct object position; Vasishth suggests that a definite NP in the subject or oblique position would not have been difficult to process.

- (8) a. Sita      kitaab              paDh      rahii      hai  
       Sita      book                    read      CONT      is  
       ‘Sita is reading a book’
- b. Sita      kitaab-ko              paDh      rahii      hai  
       Sita      book-ACC              read      CONT      is  
       ‘Sita is reading \*a/ the book’
- (Vasishth, 2003)

Vasishth (2003), generalising from Aissen (2003), suggests that cross-linguistically Case marking indicates markedness and that Case marking becomes obligatory as the degree of markedness increases. He evaluates corpora to show that marked objects require Case marking and that indefinites occur more frequently in the direct object position than definite NPs. Non-definites occurring more frequently in direct object position concurs with Givón’s hypothesis (1978) which says that direct object positions usually introduce new discourse referents<sup>1</sup>. A definite NP, which is ‘old information’ in this position is thus marked, and markedness has to be expressed using overt Case marking. Since a definite NP is not expected in the direct object position, their occurrence is non-canonical and unexpected, and thus increases the processing load.

Malayalam does not implement DOM to indicate definiteness as in (7); does it imply that there will be no processing cost associated with definite NPs in Malayalam? It is also possible that a non-definite NP introduces a new discourse referent forcing the parser to allocate memory resources to identify and encode this new referent. In contrast, a definite NP is presupposed and requires less memory load for processing. This hypothesis is independent of the position in which the non-definite NP occurs and predicts that a definite NP will be easier to comprehend. The possible outcomes regarding the effect of having non-definite NPs in an embedded sentence provide an environment where a determiner provides cues for an incoming item in the absence of Case-like features.

---

<sup>1</sup> A discourse referent is an entity that exists in a spatio-temporal location, which can be referred to in speech using a pronoun or NP if it is an object, or using tense on a verb if it is an event (Gibson, 2000). A new discourse referent is an entity that has not been presupposed or previously mentioned in a context; it is usually introduced using non-definite NPs.



### 2.3 Memory and anticipation-based processing models

Computational theories of language processing difficulty argue for either a memory or expectation-based approach. For the present study, we will examine one representative theory from each group: Dependency Locality Theory, or DLT (Gibson, 2000) and Surprisal theory (Hale, 2001; Levy 2008). This section provides a brief introduction to these approaches. A detailed account regarding their architecture and metrics is presented in Section 6.

DLT posits two separate components of a sentence's processing cost: storage and integration costs. Storage or memory cost depends on the number of syntactic heads required to complete the grammatical structure and seems to be independent of the amount of time that an incomplete dependency is held in memory (Gibson, 2000). Integration cost, on the other hand, is locality-based, i.e. the cost is based on the distance between the dependent and its head; this distance is based on the number of new intervening discourse referents (Gibson, 2000). DLT accounts for a wide range of sentence processing data from English (Gibson, 1998), Japanese (Babyonyshev & Gibson, 1999; Nakatani & Gibson, 2010) and Hindi (Vasishth, 2003; Agrawal, Agarwal, and Husain, 2017; Husain, Vasishth, & Srinivasan, 2015), to name a few.

The Surprisal framework attempts to provide a unified account for ambiguity resolution as well as predictive parsing in the case of other syntactic complexities (Hale, 2001; Levy, 2008). When there is an unexpected input, there is a forced (and costly) discarding of a highly probable structure, indicating a higher Surprisal. The processing cost at a particular word is estimated as the effort involved in discarding the current analysis or triggering a reanalysis at that word (Hale, 2001). The Surprisal model has found cross-linguistic support from German (Levy & Keller, 2013), Hindi (Husain, Vasishth, and Srinivasan, 2014; Agrawal et al., 2017) and English (Staub, 2007).

### 3 Malayalam Data

The Malayalam sentence processing study was a self-paced reading task which had single and double centre-embedded Malayalam sentences with definite and non-definite NPs. The study was non-cumulative and had 32 target sentences and 32 non-target sentences (fillers). Tables 1 and 2 show the region-wise division in the self-paced reading task, for single and double embedded sentences with a definite NP.

<i>Region</i>	1	2	3	4	5
<i>Single embedded</i>	Radha NP1.NOM	amma NP2.NOM	paalkaaran-e NP3-ACC	viLikkunna -atə VP2 -NZ	keeTT -u VP1 -PAST

‘Radha heard mother calling the milkman

**Table 1**

*Region-wise representation of a single-embedded sentence with a definite NP*

Region	1	2	3	4	5	4	5	
Double	Radha	Raju-vinoodə	amma	paalkaaran-e	viLikkunna	-atə	keLkk -aan	paranj -u
embedded	NP1.NOM	NP2-DAT	NP3.NOM	NP4-ACC	VP3	-NZ	VP2 -NF	VP1 -PAST

‘Radha told Raju to listen to mother calling the milkman’

**Table 2**

*Region-wise representation of a double centre-embedded sentence*

Tables 3 and 4 show the region-wise division in the self-paced reading task, for single and double embedded sentences with a non-definite NP, preceded by a determiner *oru*.

Region	1	2	3	3	4	5	
Single	Radha	amma	oru	paalkaaran-e	viLikkunna	-atə	keeTT -u
embedded	NP1.NOM	NP2.NOM	a.DET	NP3-ACC	VP2	-NZ	VP1 -PAST

‘Radha heard mother calling a milkman’

**Table 3**

*Region-wise representation of a single-embedded sentence with a non-definite NP*

Region	1	2	3	4	5	6	7	8	
Double	Radha	Raju-vinooDə	amma	oru	paalkaaran-e	viLikkunna	-atə	keLkk -aan	paranj -u
embedded	NP1.NOM	NP2-DAT	NP3.NOM	DET	NP4-ACC	V3	-NZ	V2 -NF	V1 -PAST

‘Radha told Raju to listen to mother calling a milkman’

**Table 4**

*Region-wise representation of a double-embedded sentence with a non-definite NP*

The region-wise layout of a single-embedded sentence is *NP1 [NP2 NP3 V2] V1* and that of a double embedded sentence is *NP1 NP2 [NP3 NP4 V3 V2] V1* (Table 5). Note that in these embedded structures, the NPs are stacked at the beginning and verbs at the end. The region manipulated to examine the definiteness factor is NP3 for single and NP4 for double embedded sentences. These NPs are preceded by the determiner *oru*. The presence of the determiner disrupts NP stacking.

	<i>Region 1</i>	<i>Region 2</i>	<i>Region 3</i>	<i>Region 4</i>	<i>Region 5</i>	<i>Region 6</i>	<i>Region 7</i>
<i>SE</i>	NP1	NP2	NP3	V2	V1	---	---
<i>DE</i>	NP1	NP2	NP3	NP4	V3	V2	V1

**Table 5**

*Layout and regions of analysis for single and double-embedded sentences*

## 4 The experiment

The test sentences were presented using version 3.2.5 Kafkaesque Koffka of OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) in a controlled environment. There were 64 experimental sentences: 32 target sentences and 32 non-target (fillers) sentences. The sentences were randomized on the OpenSesame software<sup>2</sup>. The self-paced sentences were presented one word at a time. Participants had to press the spacebar key to go to the successive word. The software recorded every key press response and the reaction time (time duration between the appearance of stimuli on the screen and the time of response). Examples (5a-b) illustrate the structure of test sentences used in the study. See (9) for example of a filler sentence.

### (9) Cleft

[Ummar aaN-ə] Sheela-yooDə kutti-ye viLikk-aan paranj-atə  
 Ummar.NOM be-PRES Sheela-DAT child-ACC call-NF tell-NZ  
 ‘It is Ummar who told Sheela to call the child’

The main experiment was preceded by a practice session consisting of ten trials. Each experimental trial consisted of a self-paced sentence which had 5-6 parts, and at the end of the sentence, a key press would lead to a comprehension question, which served as a distractor task (see example 10).

### (10) a. Sentence

Radha [amma paalkaaran-e  
 Radha.NOM mother.NOM milkman-ACC  
 ‘Radha heard mother calling the milkman’

### b. Comprehension question

<sup>2</sup> The study reported here is one section of a larger study that looked at levels of embedding, position of embedding, effect of same Case on adjacent NPs and definiteness. The data and results reported here are associated with two factors: levels of embedding and definiteness.

Ar-aaNə paalkaaran-e viLich-a-tə

Who-is milkman-ACC call-PAST-NZ

‘Who called the milkman?’

c. Options

(a) Radha (b) amma

For the analysis, data from the self-paced task were used for region-wise analysis for the experimental conditions and results from the comprehension question were used for examining the accuracy of interpretation.

The reaction time data was trimmed to minimize the effect of outliers using the interquartile method (Jones, 2019). Statistical analysis of data was run on R and used log transformed values of reaction times (ms). The analysis involved determining areas of interests for each condition and doing contrast analysis using ANOVAs.

**Participants:** 26 native speakers of Malayalam within the age range of 30 - 85 years answered the questionnaire. The subjects could speak and read Malayalam without any difficulty. They used Malayalam in informal contexts, at home and with friends, and therefore had a sound knowledge of the language.

## 5 Results

The results sections contains results for (i) stacked NPs in centre-embedded sentences and (ii) non-definite vs. definite NP comparison in a stacked NP context.

### 5.1 NP Stacking

Table 6 given below, presents the mean latencies for each region for single and double embedded sentence types. Figures 1a-b graphically illustrated the data from Table 6.

	<i>Region 1</i>	<i>Region 2</i>	<i>Region 3</i>	<i>Region 4</i>	<i>Region 5</i>	<i>Region 6</i>	<i>Region 7</i>
<i>SE</i>	817.88	943	1156	---	---	845.4	864
<i>DE</i>	888.85	1026.8	1211.98	1133.99	875	815.5	890

**Table 6**

*Mean RTs for each region of single and double-embedded sentences*

Note: SE stands for single-embedded and DE for double-embedded. Mean RTs (ms) are given for each region in single and double-embedded sentences; the shaded regions form the embedded part. Regions 4

and 5 for single-embedded sentences are left empty to visually align its embedded verbs with that of double-embedded sentences; regions 5 and 6 are regions 4 and 5 respectively in single-embedded sentences.

Values from Table 4.4 and Figure 4.1 clearly show that there was an increase in processing cost starting from the first NP till the final NP followed by a decrease towards the end at the verbs. For single embedded sentences (see Figure 4.1 (a)), the processing time was shortest at NP1 and longest at NP3, then shorter at V2 and V1. For the double-embedded sentence (Figure 4.1 (b)), the processing time increased starting from NP1 and is the longest at NP3, and fell slightly at NP4. There was also no remarkable increase in reaction time between the embedded verbs V3 and V2, and the main verb V1.

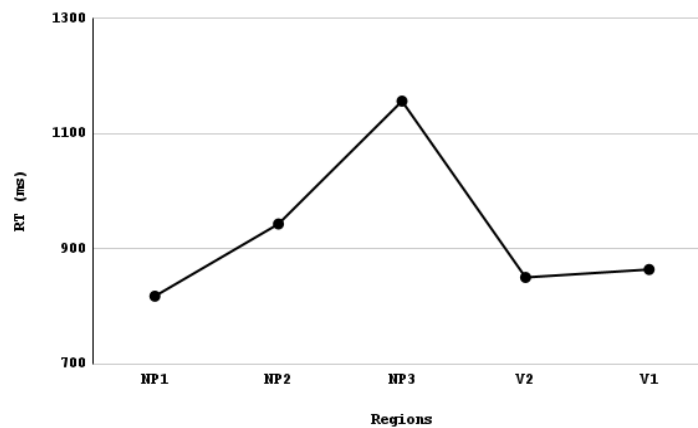


Figure 1 (a) Mean RTs (ms) for single-embedded sentences

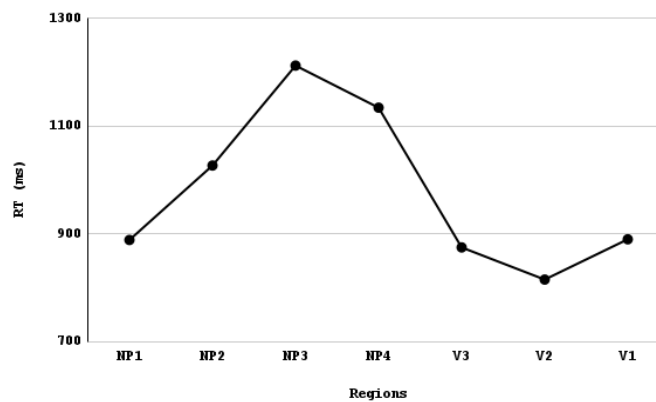


Figure 1 (b) Mean RTs (ms) for double-embedded sentences

The above-given data shows that the observed trend for embedded sentences was that the RT increased as more NPs were added and decreased as the corresponding verbs were encountered. As the data shows, the contributing factor towards the increase of processing time was the number of NPs at the beginning of the embedded clause.

A one-way ANOVA was run to check whether the number of NP in the stacking affected the reading time for each NP. The analysis compared the RTs for NP1, NP2 and NP3 for single embedded and NP1, NP2, NP3 and NP4 for double embedded sentences. Table 7 presents the mean reading times (in milliseconds) for the NPs along with the results of the analysis for comparison of the reading time for NPs in a sentence.

	<i>NP1</i>	<i>NP2</i>	<i>NP3</i>	<i>NP4</i>	<i>F-value</i>	<i>p-value</i>
<i>SE</i>	817.46 <sup>c</sup>	943.17 <sup>c</sup>	1156.33 <sup>a,b</sup>	---	8.6	P < 0.0001
<i>DE</i>	888.84 <sup>c,d</sup>	1027.59	1211.13 <sup>a</sup>	1132.88 <sup>a</sup>	6.78	P < 0.0001

**Table 7**

*Comparison of RTs (ms) across stacked NPs for single and double-embedded sentences*

The results showed a significant difference between the reaction times of the stacked NPs for both single- ( $F(2, 44) = 8.6, p < 0.0001$ ) and double-embedded sentences ( $F(3, 43) = 6.78, p < 0.0001$ ). The reading time increased with each additional NP in both types of sentences. The first NPs had the shortest reading times (single: 817.46 ms; double: 888.84 ms), and the innermost NPs had longer RTs, at least by 250-300 ms. Post hoc Tukey tests showed that in single-embedded sentences, RTs were significantly longer for NP3 when compared to NP1 ( $t = 4.5, p < 0.0001$ ) and NP2 ( $t = 2.6, p < 0.0001$ ). However, there was no difference between the time taken to read NP1 and NP2. Significant differences were found between NP1 and NP3 ( $t = 4.6, p < 0.0001$ ) and NP4 ( $t = 3.75, p < 0.0005$ ) in double-embedded structure. There was no significant difference between NP3 and NP4.

The results show that for single-embedded sentences, NPs arranged according to RT values give the order:  $NP1 < NP2 < NP3$ . And, for double embedded sentences, the arrangement gives the order:  $NP1 < NP2 < NP3, NP4$ .

## 5.2 Definite and Non-definite NP

We compared (a) whether non-definite (and definite) NPs were read in the same time in single and double-embedded sentences (comparison across a number of embeddings), and (b) whether the non-definite NPs were read in the same time as the definite NPs in embedded sentences (comparison of definite and non-definite status). The region-wise layout for non-definite NPs and definite NP in the embedded sentences are represented in

Table 8<sup>3</sup>. The shaded regions were the points of interest. We also looked at the spillover regions (V2 in single embedded sentences and V3 in double embedded sentences). Regions other than the target regions were not analysed.

	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7	Region 8
NON-DEFINITE								
SE	NP1	NP2	oru	NP3	V2	V1	—	—
DE	NP1	NP2	NP3	oru	NP4	V3	V2	V1
DEFINITE								
SE	NP1	NP2	NP3	V2	V1	—	—	—
DE	NP1	NP2	NP3	NP4	V3	V2	V1	—

**Table 8**

*Layout and regions of analysis for single and double-embedded sentences with definite and non-definite NPs*

	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Region 7	Region 8
SE	883.6	980	913.5	816.4	812.6	873		
DE	793	1093	1288.7	911.05	806.9	803.9	776	806

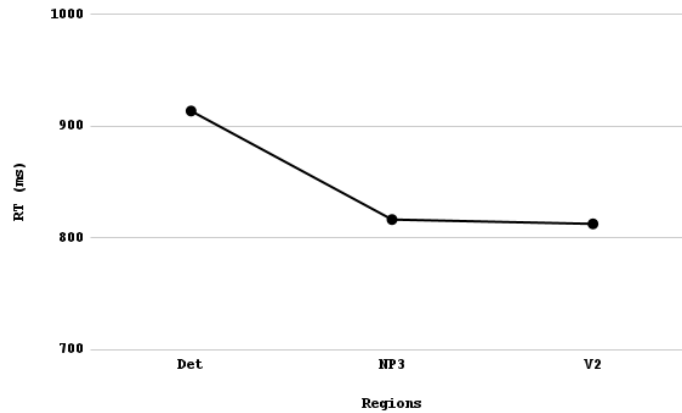
**Table 9**

*Region-wise mean RTs of single and double-embedded sentences with a non-definite NP*

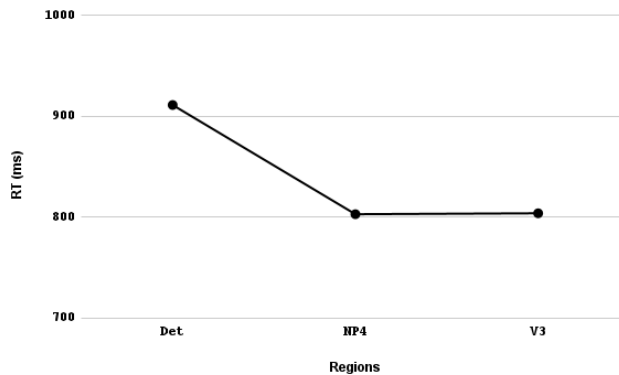
Table 9 presents the comparisons of RT at the three regions: determiner, NP and the following verb across single- and double-embedded sentences; the values showed that ‘*oru*’ which took (approximately 900 ms) helped in anticipating an NP and therefore brought down the reading time of the consequent NP by 100-125 ms. However, the number of embedded clauses made little difference to the reading time of ‘*oru*’ or the NP. The differences were not statistically significant, which is evident from Figures 2a-b given below. The V2 and V3 comparison (spillover regions) did not show a dramatic

<sup>3</sup> Initially, the study intended to study definiteness as a semantic factor, for which the determiner and following the NP would form a single segment for analysis. However, in the region-wise segmentation, the sentences were segmented at the level of word, and therefore the non-definite NP *oru paalkkaran* were presented as two segments rather than at a phrasal level. The intention was also to check whether *oru* allows the parser to anticipate an NP. However, at the time of analysis, a direct comparison of DP was not possible, since the determiner and the NP formed separate regions. Thus a comparison was made between the two NPs (one which had a preceding *oru* and the other which had a null determiner).

change in reading time in either of the embedded contexts. Therefore, the number of embeddings did not seem to affect the processing of the embedded object with respect to definiteness.



**Figure 2 (a)** Mean RTs (ms) for *oru*-NP3-V2 sequence from single-embedded sentences

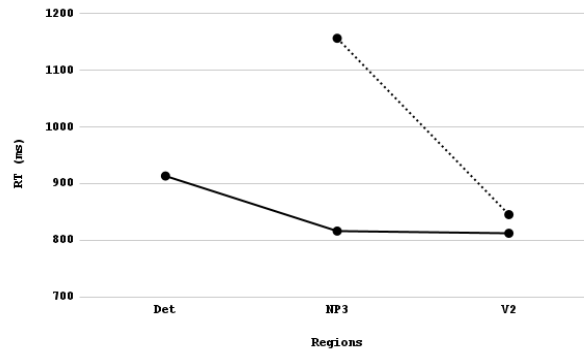


**Figure 2 (b)** Mean RTs (ms) for *oru*-NP4-V3 sequence from double-embedded sentences

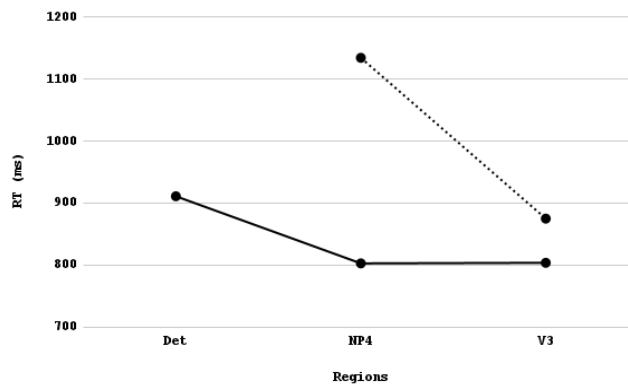
To check whether the definiteness status affected the processing of the NPs, a comparison was made between the RT of NPs (when they were (not)/preceded by *oru*). For the comparison, we looked at single and double-embedded sentences individually.

Figures 3 (a) and (b) show that determiner *oru* lowered the processing cost for the non-definite NP when compared to the definite sentences. The analysis attempts to determine if the determiner *oru* significantly facilitated the addition of NP by predicting it; if it allowed the parser to anticipate and allot memory resources for storage of the incoming NPs.





**Figure 3 (a):** Comparison of RTs for non-definite and definite NPs: *oru*-NP3-V2 (non-definite; solid) and NP3-V2 (definite; dotted) sequence from SE sentences



**Figure 3 (b):** Comparison of RTs for non-definite and definite NPs: *oru*-NP4-V3 (non-definite; solid) and NP4-V3 (definite; dotted) sequence from DE sentences

	<i>Non-definite (ms)</i>	<i>Definite (ms)</i>	<i>F-value</i>	<i>p-value</i>
<i>Single-embedded</i>	816.4	1156.05	13.8	p<0.001
<i>Double-embedded</i>	806.9	1133.9	31.23	p<0.001

**Table 10**

*Effect of definiteness in sentences with definite and non-definite NPs*

Table 10 shows that there was a significant difference between the reading times at definite and non-definite NPs for single ( $F(1,45) = 13.8, p < 0.001$ ) and double-embedded ( $F(1,45) = 31.23, p < 0.001$ ) sentences. These results indicate the determiner *oru* facilitates the processing of the following (non-finite) NP. The final NPs (NP3/NP4) were easier to process because the determiner *oru* helps the parser to

anticipate the incoming verb. There was also no significant difference between non-definite and definite conditions at the innermost verb; the definiteness of the NP did not affect the processing of the following verb (V3/V2). The results showed that the determiner *oru* that precedes a definite NP helps the parser in predicting the incoming NP, making the storage easier.

## 6 Discussion

To summarise the results from the previous section, processing difficulty increases with the addition of each NP in a centre-embedded sentence. If arranged according to RT measurement, we can come up with the following NP sequences: *NP1<NP2<NP3* for single-embedded sentences and *NP1<NP2<NP3, NP4* for double embedded sentences. In this case, all NPs are definite. When the embedded clause has a non-definite NP, (*oru palkkaaran*), the reading times significantly decrease at the determiner and the following NP. The determiner thus facilitates the faster integration of the NP by creating an anticipation for it. Following sections will look at how the two introduced earlier: Dependency Locality Theory, or DLT (Gibson, 2000) and Surprisal theory (Hale, 2001; Levy 2008), will account for the results.

### 6.1 Dependency Locality Theory (Gibson 2000)

The locality metric of the Dependency Locality Theory (Gibson, 2000) predicts that the processing of a verb should be easier if its arguments are closer. DLT's version of locality metric counts new discourse referents to measure distance. The integration component of DLT counts the number of intervening discourse referents between the item being integrated and the structure into which it is being integrated. Dependency Locality Theory, also assigns a separate storage cost for the memory load incurred by keeping unresolved dependencies or predictions active. Storage cost is a measure of the number of heads required to complete a structure that is currently being constructed. We will first look at how DLT's metrics can be used to estimate the processing cost for nouns and verbs in embedded sentences in Malayalam.

*Integration cost:* In DLT, nouns and verbs are discourse referents, pronouns and adjectives are not. Each new discourse referent gets 1 unit of integration cost. The Dependency Locality measures the distance from the item being integrated to its dependent, which has already been stored in the working memory. Table 11 demonstrates how integration cost is calculated for a single-embedded sentence in Malayalam<sup>4</sup>.

(11) Radha [amma paalkaaran-e viLikkunna-atə] keeTT-u

---

<sup>4</sup> The calculation of Integration Cost here is very basic and not explored further here. DLT model is evaluated on the basis of the experimental data regarding the processing of stacked NPs. This requires a detailed look at the Storage component of DLT, which is provided in the following segments.

Radha.NOM amma.NOM milkman-ACC call-NZ hear-PAST  
 “Radha heard mother calling the milkman”

<i>Regions</i>	1	2	3	4	5
<i>Category</i>	NP1.NOM	NP2.NOM	NP3.ACC	V2 -NZ	V1 -PAST
<i>New Discourse Referents</i>	1	1	1	1 0	1 0
<i>Distance of Integration</i>	0	0	0	2 0	4 0
<i>Integration cost</i>	1	1	1	1.5	2.5

**Table 11**

*Integration cost for a Malayalam single-embedded sentences*

*Note:* For estimating the integration cost (and storage cost in subsequent tables), I have split the verb into stem and tense inflection (V2, -NZ) to capture the parsing capabilities based on inflectional morphology of the verb following the method adopted in (Gibson & Nakatani, 2010). The total integration cost (or storage cost) at a word is then the average of the storage cost of the stem (V2) and the inflection (-NF).

*Storage Cost:* Storage cost is the count of the number of heads (nouns or verbs) minimally required to complete the current structure (Gibson, 2000). This predictive feature of the storage component can be evaluated on the basis of whether it can account for (i) the processing difficulty at nouns (NP stacking and Similarity-based interference), (ii) lower processing cost at non-definite NPs and (iii) increased processing difficulty at the final verb when it follows a nominalised verb.

Table 12 shows the storage cost for single-embedded sentences given in (12). The values show that the storage component of the models predicts that the memory load increases as the NPs are added to the sentence and decreases as verbs are encountered. As NPs are added, they have to stay in the memory till their head (verb) is encountered and integrated with the subject. As they are being integrated, memory load is taken off. This is not explicitly treated as a distance problem but having a separate storage cost component is the consequence of defining locality as a linear-order metric where processing requires that these predictions stay active over a particular distance, until their retrieval is triggered.

- (12) a. Radha [amma paalkaaran-e viLikkunna-atə] keeTT-u  
 NP1.NOM NP2.NOM NP3-ACC VP2-NZ VP1-PAST  
 “Radha heard mother calling the milkman”

<i>Regions</i>	1	2	3	4	5
<i>Categories</i>	NP1.NOM	NP2.NOM	NP3.ACC	V2 -NZ	V1 -PAST
<i>Predicted heads</i>	V1	V1, V2	V1, V2	V1 V1	0 0
<i>Memory cost</i>	1	2	2	1 1	0 0
<i>Total cost</i>	1	2	2	1	0

**Table 12**

*Storage cost for Malayalam single-embedded sentences*

For the single-embedded sentence, the storage cost is 1 MU at region 1 because it needs only 1 finite verb (V3) to complete the sentence. As another nominated NP is encountered, the parser predicts an embedded verb V2 and a finite final verb V1; the storage cost is revised to be 2 MUs. Similarly, as more NPs are encountered, the parser predicts more verbs depending on the Case information it receives from the NPs. As verbs are encountered the storage cost reduces because memory is freed up as dependencies are resolved; in addition, no new items are predicted at these verbs.

The storage component of DLT rightly predicts the increase in RT due to repeated encounter and storage of nouns. However, it does not account for how the parser is able to anticipate a non-definite NP as there is no mechanism for accounting for how predictions reduce processing cost. Table 13 presents the results against storage cost predictions.

<i>Storage results</i>	<i>Level of embedding</i>	<i>DLT's prediction</i>	<i>Data results</i>	<i>Remarks</i>
<i>NP stacking</i>	Single	NP1 < NP2, NP3	NP1 < NP2 < NP3	Confirmed
	Double	NP1 < NP2, NP3, NP4	NP1 < NP2 < NP3, NP4	Confirmed
<i>Definiteness</i>	Single & double	NP <sub>ND</sub> = NP <sub>D</sub>	NP <sub>ND</sub> < NP <sub>D</sub>	Not confirmed

**Table 13**

*Evaluation of the storage component of DLT*

## 6.2 Surprisal Model

The Surprisal model (Hale, 2001; Levy, 2008) is an expectation-based probabilistic model which claims that reading times are lower at the head when there is more preverbal (intervening) material in the clause. In addition to this, the Surprisal model also claims

that the intervening items help accurately predict the final verb's identity and location, and that expectation for a verb becomes stronger as the length of the sentence increases.

The models measure processing difficulty with a Surprisal cost. Surprisal cost is the processing difficulty at any point, taken as the effort or cognitive load involved in disconfirming a highly probable structure at that point. When the parser encounters a nominative NP, a verb that governs it is predicted; this would be the simplest and possibly the most predicted structure. If the next input is another NP, the parser needs to discard the prediction and posit either (i) a transitive verb if the new NP is dative or accusative marked or (ii) an embedded clause with a non-finite verb if the NP is nominative.

The three steps in (13) illustrate how expectation based parsing works for a single embedded sentence (12). At each step, the parser encounters a new NP: in (12a) with NP1 (Radha), a VP (*urangi*) is predicted; in (13b) with NP1 (Radha) and NP2 (*amma*), a nominalized VP2 and the main verb (VP2) is predicted; in (13c) with NP1, NP2 and NP3 (*paalkkaaran-e*), a non-finite/nominalized VP2 and the main verb (VP2) is predicted. Therefore, in (13a) the predictions are not met, leading to a Surprisal cost; in (13b) too, there is a Surprisal cost. However, in (13c), there is no or marginal Surprisal cost expected since the prediction of the embedded clause given in (12) is successfully met.

(13) a. **Radha**

Radha            urang-i  
 Radha.NOM    sleep-PAST  
 'Radha slept'

b. **Radha amma**

Radha            [amma            urangunna-atə] kaND-u  
 Radha.NOM    mother.NOM    sleep-NZ            see-PAST  
 'Radha saw mother sleeping'

c. **Radha            amma            paalkkaaran-e**

Radha            [amma            paalkkaaran-e    viLikkunna-atə] keeTT-u  
 Radha.NOM    mother.NOM    milkman-ACC    call-NZ            hear-PAST  
 "Radha heard mother calling the milkman"

Example (12) shows how NP stacking, despite temporarily resulting in processing load,

ultimately provides information that allows easier integration of the verbs. In case of definite NPs, the Surprisal model predicts that when the parser encounters the determiner *oru*, it anticipates a noun, which was seen in the Malayalam results, where the NP showed a faster reading time when compared to a definite NP (where the definiteness is null). The determiner essentially functions as a highly reliable expectation for an incoming NP. Table 5.10 presents the results against predictions of the Surprisal model.

<i>Position of Embedding</i>	<i>Level of embedding</i>	<i>Expectation-based prediction</i>	<i>Data results</i>	<i>Remarks</i>
Storage				
<i>NP stacking:</i>	Single	NP1 < NP2 < NP3	NP1 < NP2 < NP3	Confirmed
	Double	NP1 < NP2 < NP3 < NP4	NP1 < NP2 < NP3, NP4	Confirmed
<i>Definiteness</i>	Single & double	NP <sub>ND</sub> < NP <sub>D</sub>	NP <sub>ND</sub> < NP <sub>D</sub>	Confirmed

**Table 14**

*Evaluation of the expectation-based Surprisal model*

## 7 Conclusion

In the present study, we focused on the processing of nouns, stacked NPs specifically, in centre-embedded sentences. The intention was to test if stacked NPs incur a processing load. This prediction was proven, as reaction times increased with each NP. The result matched predictions of DLT, according to which NPs are items in an unresolved dependency. It predicts the incoming head which would resolve the expectation, but at a cost; the cost increases with the distance (and time) between the NP and its verb. Surprisal theory accounts for the difficulty associated with NP stacking as discarding or inhibiting inaccurate predictions regarding what is to come. The predictions made by the two models do not seem to be very different. It is also important to note that Case information does not facilitate the storage of NPs; or rather, Case information is not fully accessible to the parser. When it comes to definiteness, DLT and Surprisal make different predictions. While Surprisal theory can correctly predict how the determiner *oru* creates an anticipation for an NP. DLT does have a prediction-based Storage component, but it does not explain how that reduces processing load.

At one level, the processing difficulty at NPs seems like it is arising out of delayed resolution or inaccurate predictions. Another way to think of it would be in terms of similarity-based storage interferences. It is difficult to efficiently differentiate between and allocate memory resources to items of a similar type, i.e., NPs; this results in a storage-related cost. However, when a different kind of item is encountered, *oru*, the storage is easier, so is processing of the item predicted by *oru*. This analysis also directs

one to look at more activation-based accounts (Lewis and Vasishth, 2005) for further research.

## References

- Agrawal, Arpit, Agarwal, Sumeet and Husain, Samar. (2017). Role of expectation and working memory constraints in Hindi comprehension: An eyetracking corpus analysis. *Journal of Eye Movement Research*, 10(2), 1-15.  
<https://doi.org/10.16910/jemr.10.2.4>
- Aissen, Judith. (2003). Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory*, 21(3), 435–483.  
<https://doi.org/10.1023/A:1024109008573>
- Chomsky, Noam. (1957). Logical structures in language. *Journal of the American Society for Information Science*, 8(4), 284.
- Gibson, Edward. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O’Neil (Eds.), *Image, language, brain: Papers from the First Mind Articulation Project Symposium* (pp. 95-126). Cambridge, MA: MIT Press. Retrieved July 12, 2021, from  
[https://www2.bcs.rochester.edu/sites/fjaeger/teaching/LabSyntax2006/readings/Gibson\\_2000.pdf](https://www2.bcs.rochester.edu/sites/fjaeger/teaching/LabSyntax2006/readings/Gibson_2000.pdf)
- Givón, Talmy. (1978). Definiteness and referentiality. In J.H. Greenberg, C.A. Ferguson, & E.A. Moravcsik (Eds.), *Universals of human language*, Volume 4 (Syntax) (pp. 291–330). Cambridge, MA: Stanford University Press, Stanford, California.
- Hale, John. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*.  
<https://doi.org/10.3115/1073336.1073357>
- Husain, Samar, Vasishth, Shravan and Srinivasan, Narayanan. (2014). Strong expectations cancel locality effects: Evidence from Hindi. *PLoS ONE*, 9(7), e100986. <https://doi.org/10.1371/journal.pone.0100986>
- Konieczny, Lars. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, 29(6), 627–645. <https://doi.org/10.1023/A:1026528912821>
- Konieczny, Lars and Döring, Philipp. (2003). Anticipation of clause-final heads: Evidence from eye-tracking and SRNs. In *Proceedings of iccs/ascs* (pp. 13-17).
- Levy, Roger P. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>

- Levy, Roger P. and Keller, Frank. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, 68(2), 199–222.  
<https://doi.org/10.1016/j.jml.2012.02.005>
- Lewis, Roger L. and Vasishth, Shravan. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.  
[https://doi.org/10.1207/s15516709cog0000\\_25](https://doi.org/10.1207/s15516709cog0000_25)
- Nakatani, Kentaro and Gibson, Edward. (2010). An on-line study of Japanese nesting complexity. *Cognitive Science*, 34(1), 94–112.  
<https://doi.org/10.1111/j.1551-6709.2009.01067.x>
- Staub, Adrian. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 550–569.  
<https://doi.org/10.1037/0278-7393.33.3.550>
- Vasishth, Shravan. (2003). Working memory in sentence comprehension: Processing Hindi center embeddings. (Doctoral dissertation), The Ohio State University.
- Yngve, Victor H. (1960). A model and an hypothesis for language structure. *Proceedings of the American philosophical society*, 104(5), 444-466.  
<https://dspace.mit.edu/handle/1721.1/4453>



# Gender Agreement in a Tamil-Hindi Bilingual Situation: The Role of Feature Valuation

MADHUSMITHA VENKATESAN, *Indian Institute of Technology Delhi*

## ABSTRACT

Within the Generative paradigm, variation is understood in terms of features. A crucial mechanism for this is the Borer-Chomsky Conjecture, explained in Baker (2008) as: ‘All parameters of variation are attributable to differences in the features of particular items in the lexicon.’ This paper is an attempt to understand a case of asymmetric bilingualism of Hindi-Urdu (Indo-Aryan) and Tamil (Dravidian) and to explain the resultant changes in gender agreement in terms of featural configurations of functional heads. The empirical core is set in New Delhi, with simultaneous bilinguals acquiring Tamil (L1) at home and Hindi-Urdu (L2) from the external environment. While both grammars mark gender information on the verb, Hindi-Urdu has grammatical and biological gender. In Tamil, on the other hand, gender is purely semantic. The Hindi-Urdu grammar of this bilingual population appears to find gender agreement challenging.

This paper adopts a representational approach; the loss of Gender is analysed as the deletion of an uninterpretable valued feature on a functional head. The first approach is to posit the uninterpretability of the feature as the cause of deletion. The inability of this claim to hold up empirically is then taken to mean that the explanation for the deletion of the feature lies in its other property: Value. Reanalysis of the change in contact situations reveals that losing valued features could simply be a strategy adopted by languages in an effort to be more parsimonious.

## 1 Introduction

Contact situations are one of the most significant contributors to language variation and change. Such changes are crucial in enriching our existing knowledge on the shape of Universal Grammar, especially with regard to its extent of plasticity. Within the generative paradigm, understanding variation in terms of features has been a rewarding exercise as it affords us a microscopic view of the way in which languages undergo change. This paper attempts to contribute to the discussion by introducing the factor of a contact situation. The objective of this paper is to understand the features underlying gender agreement in Hindi-Urdu when placed in the bilingual context of simultaneous bilinguals of Hindi-Urdu and Tamil.

## 1.1 Gender as a Category

Gender is primarily a system of nominal categorization; it is a way of organising nouns into different categories based on some inherent attributes that the items may possess (Corbett, 1991). There are multiple aspects of gender which make its study essential; the acquisition of Gender is intricately connected to structure, the presence of gender in the world's languages is highly varied, and its omnipresent involvement in agreement relations reveals much about the role of features in agreement operations. All this combined offers an insightful window into the mental architecture of Language.

Gender in languages is broadly divided into two kinds: Natural/Biological or Semantic gender (BG), and Grammatical Gender (GG). The former is a system in which meaning plays a central role. Only animate nouns are allotted a gender, and their gender value correlates with the biological sex of the referent. Thus, nouns denoting a male human or animal are assigned [MASCULINE] gender, and nouns denoting female humans or animals are assigned [FEMININE] gender. Crucially, these are nouns where the gender value of the noun is grounded in its meaning. Grammatical Gender, on the other hand, should be understood as a more complex and abstract concept than just the co-referent of biological sex. GG may or may not be congruent with natural/biological gender, animacy or other related semantic properties. Grammatical Gender is often referred to as arbitrary gender (Kramer, 2015), as neither is there consistency in the assignment of gender to inanimate objects nor do the intrinsic properties of the noun have any role to play in its gender assignment.

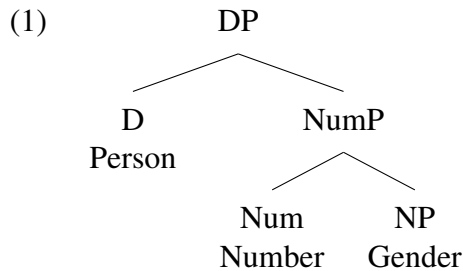
## 1.2 Gender and Language Contact

Whenever the context is bilingual, grammatical gender seems to be vulnerable. Loss of gender is cross-linguistically common in language contact situations. This is well recorded in the literature (Sánchez et al., 2022; Igartua, 2019; Lohndal & Westergaard, 2016; Kramer, 2014; Karatsareas, 2009; Oliphant, 1998; Gumperz & Wilson, 1971). These are all cases of typologically unrelated languages losing grammatical gender, and the reason has canonically been attributed to sustained contact with a gender-less language. This brief but telling literature survey forms the basis for this study: Empirical evidence points towards gender being vulnerable in different kinds of language contact situations. What could possibly motivate such a tendency in gender agreement? How can we capture this using the theoretical mechanisms of generative grammar?

## 1.3 Gender in the Generative Theory: An Uninterpretable and Valued Feature

Features have been central to generative theory almost from its inception (Chomsky, 1965). Since then, features have evolved as crucial for operations such as MERGE, MOVE and AGREE; it is the features of a lexical item that decide which other lexical items it may combine with, and features are integral for forming syntactic dependencies between a Probe and a Goal, leading to agreement. This naturally leads to the next question: How do we represent gender as a feature? According to the Universal DP structure (1), as proposed

by Polinsky (2016) and Carstens (2000), each projection inside the DP is associated with different phi-features. Person features are assumed to be a property of the determiner D, number features are contained in the Num head, and gender features are considered a lexical property of the noun itself.



Sigurðsson (2019) proposes an analysis wherein the interpretable but unvalued Gender feature on D acts as a Probe and looks for a value. It finds its Goal in the uninterpretable but valued Gender feature on the NP. A relation of Agree is established between the two and this is how D gets its valuation for Gender. Such a relation with the Interpretable feature being the Probe and the Uninterpretable feature being the Goal is not permitted in the Chomskyan system (Chomsky, 2000). However, Pesetsky & Torrego (2007) provides a framework within which the interpretable status of a feature does not prevent it from being a Probe; as long as a feature requires valuation through syntax, it can act as a Probe.

When we approach Gender as a feature on a functional head, it opens up the possibility of analysing the changes taking place in the contact situation from the perspective of the Borer-Chomsky Conjecture. The Borer-Chomsky Conjecture given by Borer (1984), Chomsky (2001) and explained in Baker (2008) states that all parameters of variation are attributable to differences in the features of particular items (eg, functional heads) in the lexicon. This approach affords us a microscopic view of how languages show variation and will have interesting implications when extended to contact situations, which differ from diachronic change in having to factor in external forces such as the influence of another language.

Following the Borer-Chomsky Hypothesis which considers even large-scale language change as an alteration of the featural composition of functional heads, the loss of grammatical gender in a language can be characterised as the loss of the uninterpretable [uGen:FEM] feature on n.

## 2 The Current Problem

The empirical core of this study is formed by the grammar of Tamil-Hindi/Urdu bilinguals residing in New Delhi. Here we present a brief overview of the gender systems of the two grammars.

## 2.1 Gender in Hindi-Urdu

Hindi/Urdu is an Indo-Aryan language with semantic as well as formal systems of gender (Corbett (2013)). All nouns, regardless of their status with respect to animacy, are assigned a gender value. There are two gender categories in Hindi, Masculine and Feminine, and every noun in the language is allotted one of the two. Additionally, Hindi/Urdu also has a rich agreement paradigm; gender (and other phi features) are marked on a range of functional heads: D (2a , 2b ) , ADJ (2c , 2d ) , T (2e , 2f ) and *v* (2g , 2h ) .

- (2) a. mer-ii kitaab  
my-F book.F  
'My book'
- b. mer-aa betaa  
my-M son.M  
'My son'
- c. lamb-ii ladki  
tall-F girl.F  
'Tall girl'
- d. baD-aa ghar  
big-M house.M  
'Big house'
- e. wah ladkaa roTii khaat-aa hai  
DEM boy.M bread.F eat.PRS-3MS be.PRS  
'That boy eats bread'
- f. wah lakkii ghar jaa rah-ii th-ii  
DEM girl.F house.M go.INF PROG-3FS be.PST-3FS  
'That girl was going home'
- g. vijay-ne roTii khaa-yii  
Vijay.M-ERG bread.F eat.PST-3FS/\*3MS  
'Vijay ate (the) bread'
- h. vijay-ne roTii khaan-ii chaah-ii  
vijay.M-ERG bread.F eat.INF-3FS want-PST-3FS  
'Vijay wanted to eat bread'

## 2.2 Gender in Tamil

Tamil is a Dravidian language with a gender system that is entirely semantic , i.e., based on the natural gender of the noun's referent. Consequently, there exists a sharp divide between [+HUMAN] and [-HUMAN] nouns. Thus, all nouns denoting human beings are encoded with a gender value that corresponds to the biological sex of the referent (MASC (3a) or FEM (3b)), and nouns denoting animate non-human (3c), as well as inanimate entities (3d), exhibit the NEUTER or non-gendered marker. The gender value of the noun then manifests

in verbal morphology, as seen in (3). Agreement in Tamil is displayed only on verbs. No other functional heads such as adjectives, determiners, etc participate in agreement.

- (3) a. anda paiyan va-nd-aan  
DEM boy.M come-PST-3MS  
'That boy came'
- b. anda ponnu va-nd-aa  
DEM girl.F come-PST-3FS  
'That girl came'
- c. maadu va-nd-udu  
cow.N come-PST-3NS  
'(a/the) cow came'
- d. seidi va-nd-udu  
news.N come-PST-3NS  
'(the) news reached (us)'

As we can see from the examples above, there are two key differences between the gender systems of (L2) Hindi-Urdu and (L1) Tamil:

- Hindi-Urdu has grammatical as well as biological gender; Tamil has only biological gender.
- Hindi-Urdu has multiple agreeing functional heads- determiners, adjectives, verbs, whereas only verbs agree in Tamil.

### 2.3 The Target Population

The target population comprises early childhood bilinguals who have spent all or most of their lives in Tamil-speaking families in New Delhi. The rationale behind selecting this group is that by growing up in a bilingual environment (Tamil at home and Hindi/Urdu in the larger society), they would have had access to input from both languages during their initial stage of language acquisition, thus making them early-childhood bilinguals. These speakers could be simultaneous or sequential bilinguals. A

Data on L2 Hindi/Urdu was collected from L1 speakers of Tamil who have grown up in New Delhi. Biological gender mismatches were found in abundance in the data, as seen in (4):

- (4) a. \*mer-aa maa  
my-M mother.F  
'My mother'
- b. \*usk-aa betii  
his/her-M daughter.F  
'His/her daughter'

- c. \*acch-aa ladkii  
good-M girl.F  
'Good Girl'
- d. \*ach-aa kitaab  
good-M book.F  
'Good book'
- e. \*sast-aa ghaDii  
cheap-M watch.F  
'Cheap watch'

The examples in (4) demonstrate the absence of gender agreement within the DP, despite the gender value of the nouns 'mother', 'daughter' and 'girl' being readily available to the speakers via external context. We also observe this effect indiscriminately on biological (4a, 4b, 4c) and grammatical (4d, 4e) gender-marked nouns.

The question to ask here is: Why is gender agreement so vulnerable in the grammar of these simultaneous bilinguals? What underlying mechanism of gender agreement is causing this effect?

### 3 Analysis

This section proceeds to address the question of why gender may be affected so severely in contact situations. The approach adopted here is layered: It is well established (Kramer, 2014, 2015) that gender as a feature is uninterpretable and valued. Our first step would be to posit the uninterpretability of the feature as the cause for its vulnerability. This claim is then held up against cross-linguistic empirical evidence, which then leads us to assess the other attribute of the feature: its value.

#### 3.1 On the Vulnerability of Interpretable Features

Features are deemed uninterpretable when they cannot be meaningfully interpreted at the C-I component or the LF interface. Such features are also considered to be more vulnerable and likely to be lost when a language undergoes change. Van Gelderen (2019) provides an account where a reanalysis of contact-induced change reveals that it is indeed the uninterpretable features that are lost. Supporting evidence for this claim comes from two phenomena: Negation Concord and Pro-Drop, both of which tend to be discarded from a grammar when it enters into contact with another language.

In certain varieties of English, a clause contains more than one negation element, but only one of them is interpretable and meaningful (5). The other negation element is considered to be superfluous and can be reinterpreted as an uninterpretable NEG feature [uNEG].

- (5) When my granddaddy dead, I ain't had no children  
When my grandfather died, I did not have any children yet.

It is reported that Negation Concord is not found in contact languages. Van Gelderen (2019) analyses the inability of Negation Concord to be retained in a contact situation as an instance of the loss of the [uNEG] feature in contact. The lexical item with the second [NEG] feature is deleted because the [NEG] here is uninterpretable and thus, vulnerable.

The second source of examples comes from Pro-Drop in language contact. Pro-Drop or Null Subject is traditionally considered to be a category-defining structural parameter. But in terms of features, it can also be described as an uninterpretable feature [uD] on T. Adger (2003) states that the [uD] feature on T is a strong feature, a requirement for a DP that drives movement of the structurally most appropriate DP to [Spec, TP]. This feature on T, too, is uninterpretable. Van Gelderen's analysis of the loss of Pro-Drop in contact situations draws upon this fact: Pro-Drop does not feature in contact situations because the feature controlling it is uninterpretable and hence vulnerable.

The interim conclusion is that all uninterpretable features are vulnerable and therefore must be let go of in contact situations, i.e., situations which force languages to change. The next section takes this idea further, and we immediately encounter a challenge while looking at another uninterpretable feature in action: Agreement controlled by T.

### **3.2 Not all Uninterpretable Features are Vulnerable**

Negation Concord and Pro-Drop in contact situations are certainly indicative of an analysis that banks on the vulnerability of uninterpretable features in general. However, such an account falls short when faced with an obstacle in the form of the other uninterpretable features on T. T is the locus of a bundle of phi-features. The phi features on T, which are responsible for agreement on the tensed verb, are all uninterpretable. According to the idea developed by Van Gelderen, they must all be vulnerable. What we see in contact situations, on the contrary, is that this prediction does not bear out.

Subject-Verb agreement (controlled by T) is not reported as a vulnerable phenomenon in contact situations. It is resilient despite being operated by uninterpretable features. This paves the way for a revision of the claim that uninterpretable features are vulnerable in contact. In the next section, we therefore proceed to look closely at the other defining property of gender features: Inherent Value.

### **3.3 The Role of Valued Features**

Let us revisit all the features discussed so far: [uD] (Pro-Drop), [uNEG] (Negation Concord), [uPhi:] (T Agreement) and [uGen:+FEM] (Gender). Van Gelderen (2019) presented the argument that uninterpretable features are vulnerable in language contact situations and therefore are lost in those contexts. The loss of grammatical gender could be analysed in the same manner. However, such an account fails to explain why T-Agreement, which is also operated by uninterpretable features, is not affected in contact situations.

We then turn to the other property of all these features: Value. Out of the four features mentioned above, three are deleted in language contact situations (Negation Concord, Pro-Drop and Gender Agreement), and one is not (T-Agreement). While uninterpretability is

not helpful in explaining this divide, Valuation is. Upon closer inspection we find a thread of commonality across those features that do get deleted: either they are inherently valued [uGen:+FEM], or they are features that do not require any value [uNEG], [uD]. Both types imply that these features do not need to participate in any structural relation such as Agree in order to get a value. The feature that does not get deleted (T-Agreement), on the other hand, is one that requires a value and will Probe the derivation in search of a Goal. This marks a crucial difference between the two types of features: a. Features that are not at all in need of Valuation. b. Features that will look for Valuation.

Gender Features belong to the former category. By virtue of being inherently valued, gender features simply do not need to act as probes. Since such features are not going to participate in structural relations, they have no function to perform in tying the structure together. Hence, they can be let go of, and languages do end up deleting them in contact situations, especially when there is a catalyst in the form of another language that does not have this feature to begin with.

This paper makes the suggestion that the grammar of Hindi-Urdu, in contact with Tamil, is willing to let go of its inherently valued gender features in an effort to economise its own system

#### **4 Conclusion**

The central prediction made by this paper was that valued features are dispensable in a contact situation, as they are not useful for the structure. This goes on to strengthen the claim that the behaviour of languages in a contact situation is no different from a system in isolation, at least with respect to the core property of economy. This claim needs to be explored further, both empirically, as well as in terms of the conceptual implications of value in a feature.

#### **References**

- Adger, David. 2003. *Core syntax: A minimalist approach*, vol. 20. Oxford University Press Oxford.
- Baker, Mark. 2008. The macroparameter in a microparametric world. *The Limits of Syntactic Variation* 351–374.
- Borer, Hagit. 1984. Restrictive relatives in modern hebrew. *Natural Language & Linguistic Theory* 2(2). 219–260.
- Carstens, Vicki. 2000. Concord in minimalist theory. *Linguistic inquiry* 31(2). 319–355.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. MIT press.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework (mitopl 15). *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* 89–155.
- Chomsky, Noam. 2001. Derivation by phase. *Ken Hale: A Life in Language* 1–52.
- Corbett, Greville G. 1991. *Gender*. Cambridge University Press.



- Corbett, Greville G. 2013. Systems of gender assignment (v2020.3). In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Zenodo. doi:10.5281/zenodo.7385533. <https://doi.org/10.5281/zenodo.7385533>.
- Gumperz, John J & Robert Wilson. 1971. Convergence and creolization. *Pidginization and creolization of languages* 151. 167.
- Igartua, Iván. 2019. Loss of grammatical gender and language contact. *Diachronica* 36(2). 181–221.
- Karatsareas, Petros. 2009. The loss of grammatical gender in cappadocian greek. *Transactions of the Philological Society* 107(2). 196–230.
- Kramer, Ruth. 2014. Gender in amharic: A morphosyntactic approach to natural and grammatical gender. *Language sciences* 43. 102–115.
- Kramer, Ruth. 2015. *The morphosyntax of gender*. Oxford University Press.
- Lohndal, Terje & Marit Westergaard. 2016. Grammatical gender in american norwegian heritage language: Stability or attrition? *Frontiers in psychology* 7. 344.
- Oliphant, Katrina. 1998. Acquisition of grammatical gender in italian as a foreign language. *Canadian modern language review* 54(2). 239–262.
- Pesetsky, David & Esther Torrego. 2007. The syntax of valuation and the interpretability of features. *Phrasal and clausal architecture* 262–294.
- Polinsky, Maria. 2016. Agreement in archi from a minimalist perspective. *Archi: Complexities of agreement in cross-theoretical perspective* 184–232.
- Sánchez, Liliana, José Camacho, Elisabeth Mayer & Carolina Rodríguez Alza. 2022. Gender agreement in a language contact situation. *Languages* 7(2). 81.
- Sigurdsson, Halldór Ármann. 2019. Gender at the edge. *Linguistic Inquiry* 50(4). 723–750.
- Van Gelderen, Elly. 2019. Are uninterpretable features vulnerable? *Theoretical Linguistics* 45(3-4). 275–282.