Proceedings of Formal Approaches to South Asian Languages 5

Edited by Rahul Balusu and Sandhya Sundaresan

October 2016

Contents

Adjectival Expressions:

Focus, Ellipsis and Sluicing:

(Anti-)causatives and middles:

Pronouns, anaphora, and polarity:

Processing and Acquisition:

Decomposing Color Expressions in Malayalam[1]

Mythili Menon & Roumyana Pancheva

University of Southern California

1. Expression of color

Although adjectival structure and meaning have been widely studied, there have been relatively few formal investigations of color adjectives specifically (Kennedy & McNally 2010, McNally 2011, McNally & de Swart 2011, Alexiadou 2013, Moltmann 2013). An examination of color expressions is warranted because they differ from other adjectives in at least two ways.

First, color adjectives have both gradable and non-gradable meanings (as in (1)-(2)). Gradable adjectives are compatible with overt degree modifiers, without a meaning change to the adjective itself (cf. the two uses of *green* in (1) vs. (2)).

(1) Gradable:
The leaves are green but they aren't green enough.

(2) Non-gradable:
The traffic light is green. # It is greener than the traffic light on that corner.

Other adjectives do not show this pattern. Relative (*tall, expensive*) and absolute (*full*, *clean*) adjectives only have gradable meanings (putting aside idiomatic expressions such as *high tea*). Non-gradable adjectives such as *Czech, chemical, female* need to be coerced to a new meaning in order to be used with degree modifiers, and can do so with varying success (e.g., *Mary is more Czech than Susana* vs. # *Mary is more female than Susana*).

[1] We would like to thank Artemis Alexiadou, Elena Anagnostopoulou, Itamar Francez, Andrew Koontz-Garboden, Carmen Dobrovie-Sorin, Anna Szabolsci, an anonymous reviewer, audiences at FASAL 5, 38th GLOW, and Syntax+ at USC for valuable comments and suggestions. Any errors are our own.

Second, color adjectives can appear in nominal positions, unlike other adjectives, as seen in (3)-(4) and discussed in McNally and de Swart (2011), Alexiadou (2013), a.o.

(3) Green suits you
(4) *{Big/short/expensive/wet} suits you

This paper contributes to our understanding of the grammar of color terms in Malayalam in comparison to other expressions used for predications, attributive modification and comparison in this language. The behavior of color terms prompts a partial revision to the analysis of Menon and Pancheva (2014). In our earlier work, we analyzed the internal structure of two classes of 'adjective-like' complex expressions in Malayalam, arguing that both are gradable. Here, however, we note that the properties of color terms require a more nuanced analysis. In particular, some color expressions show mixed properties, behaving like non-color expressions of one class in attributive position but like non-color expressions of the second class in predicative position. We offer a revised analysis that treats some color and non-color expressions in Malayalam as inherently non-gradable; further composition with possessive predicates contributes gradability.

More generally, the paper provides support for a gradable/non-gradable ambiguity in color expressions (Kennedy & McNally 2010, McNally 2011), although we do not analyze the different meanings as a case of a lexical ambiguity but as the result of different representations constructed by syntax below and above the word level. We also show evidence that color terms in Malayalam can behave as nominals, which we attribute to their combination with a null noun COLOR. Thus, Malayalam color expressions show similar behavior with respect to the phenomena illustrated in (1)-(2) and (3)-(4) and the analysis we offer in this paper contributes to the understanding of the cross-linguistic patterns.

2. Previous account of property denoting expressions in Malayalam

In previous work, we have shown that Malayalam lacks an adjectival category and uses syntactically complex expressions for predication, attributive modification, and comparison (Menon 2013, Menon & Pancheva 2014). Specifically, we have argued that such structurally complex 'adjective-like' expressions are built on the basis of property-concept denoting roots, i.e., roots with meanings such as *height, beauty, intelligence* (Chierchia and Turner 1988, Koontz-Garboden and Francez 2010, Francez and Koontz-Garboden 2015). Lexemes formed on the basis of such roots belong to two classes which differ in their syntactic category, and correspondingly, exhibit different syntactic behavior when they build the bigger structures used for predication, attribution and comparison. The ultimate semantics of the complex attributive and predicative expressions belonging to the two classes is, however, the same. We will review here the basic analysis of Malayalam property concept expressions in Menon & Pancheva (2014). We will then offer a revised account, partly in response to data from Malayalam color expressions.

Some examples of Class 1 and Class 2 property concept expressions are given in (5) and (6). Class 1 expressions are *-a*-ending reduced (participial) relatives (*-a* being the Proto-Dravidian relative marker)[2] and Class 2 expressions are *-am*-ending nominals (*-am* being a nominal marker). Class 1 expressions are formed on the basis of native roots whereas Class 2 roots are borrowed from Sanskrit.

(5) *valiy**a*** 'big', *čeriy**a*** 'small', *puthiy**a*** 'new', *pazhay**a*** 'old', *nall**a*** 'good'
(6) *santosh**am*** 'happiness', *sankaṯ**am*** 'sadness', *prayas**am*** 'difficulty'

In Menon & Pancheva (2014) we suggested that both Class 1 and Class 2 property concept expressions start out as roots denoting abstract mass substances (notated by *Π*), which, following Chierchia and Turner (1988), have the type of entities.

(7) a. $[\![\sqrt{nall}]\!]$ = the property of goodness (Class 1)
b. $[\![\sqrt{santosh}]\!]$ = the property of happiness (Class 2)

---

[2] A reviewer points us to an alternate analysis as explicated in Balusu (2014) where the *–a* ending is analyzed as the genitive marker in similar Telugu Class 1 forms.

(i) a. pillalu 'children' ~ pillal-a 'children's'
b. pustakaalu 'books'~ pustakaal-a 'books'

In Malayalam, however, unlike in Telugu, the only genitive marker is *–inte/-uṯe* and it can suffix to the *–a* ending forms after they are turned into light headed relatives, such as in (ii), shown attaching to a Class 1 property concept expression, and in (iii), the counterpart of the Telugu examples in (i), where there is no *–a* marker.

(ii) a. nalla-van-te 'good man's'
b. nalla-vaḷ-te 'good woman's'
c. nalla-t-inte 'goodness's'

(iii) a. kuṯṯi-kaḷ 'child-pl' ~ kuṯṯi-kaḷ-uṯe 'children's'
b. pustakaŋ-ŋaḷ 'book-pl' ~ pustakaŋ-ŋaḷ-uṯe 'books'

If indeed *–a* is a genitive marker in Malayalam, the *–a* ending forms in (5) should be able to appear predicatively without the help of pronominalization, which does not happen. Note that the counterpart forms of Telugu can appear in the predicative position without being turned into nominals (see Menon, *in prep* for an explanation). The possessive genitive forms can appear as a complement to a copula directly as seen in (iv).

(iv) itə kaḷḷan-te aaŋə
this robber-GEN EQ-COP
'This is the robber's.'

Thus, we maintain our analysis that *–a* in Malayalam is a relativizer and not a genitive marker. A further piece of evidence suggesting this comes from Judeo-Malayalam, the traditional language of the Cochin Jews in Kerala, now spoken primarily in Israel. As shown in Gamliel 2013, in Judeo-Malayalam the participial form is *-e* instead of *-a*, and, as we predict if the two are the same morpheme, so is the suffix on Class 1 roots. Thus, in Judeo-Malayalam, the word for 'good' is *nall-e* rather than *nall-a*. (Itamar Francez p.c).

The core idea is that these roots are turned into predicates of individuals through possessive predication, and to have *Π* means to have an instance, a certain amount, of *Π,* i.e., possessive predicates formed on the basis of property-concept roots are gradable.

2.1 Class 1 property concept expressions

The Class 1 roots are verbalized first, with a null possessive *v,* before the addition of the relative marker *–a*. Support for the presence of the null verbal head comes from the distribution of the relative marker *-a*: *-a* only merges with verbs. Hence, Class 1 roots are turned into nonfinite verbal expressions by the addition of a null *v*, which, as we suggested in Menon and Pancheva (2014), has possessive semantics, as in (8).[3]

(8) $[\![\varnothing_{v\_poss}]\!] = \lambda\Pi.\ \lambda d.\ \lambda x.\ \exists y\ [y \text{ is an instance of } \Pi \text{ and } x \text{ has } y \text{ and } \mu(y) \geq d]$

The degree argument is bound by the non-overt positive morpheme *POS*. The meaning of forms such as those in (9) is norm-related – they are interpreted as making reference to a standard, as would be expected if *POS* is binding the degree variable word-internally.

(9) $[\![nalla]\!] = \lambda x.\ \exists d\ \exists y\ [y \text{ is an instance of goodness } \&\ x \text{ has } y\ \&\ \mu(y) \geq d\ \&\ d > d_s]$
$\approx \lambda x.\ \exists d\ [x\text{'s goodness} \geq d \text{ and } d > d_s]$

Thus, Class 1 expressions are similar in meaning to a positive gradable adjective *good* in English in many semantic accounts. Their syntax of reduced participial relatives and their semantics of predicates of individuals makes them appropriate attributive modifiers (as in (10)).

(10) nalla kuṭṭi
good child

The equative (EQ) copula completes the predication structure, see (11). The EQ copula is the canonical predication strategy in Malayalam, as illustrated in (12) – it is the structure used with predicate nominals. The example in (12b) is particularly relevant, since it has the same structure as the property concept predicates in (11) – a verb ('to hear') is relativized by *-a* and turned into a participle; the pronominal then changes the participle into a light-headed relative – an appropriate nominal to be a complement to the EQ copula. An analysis that treats Class 1 expressions as morphologically simple, or assigns *–a* the status of an adjectival suffix, clearly misses the commonality between the forms in (11) and (12b) (and ex. (17) later, where *–a* relativizes the non-finite possessive copula, resulting in the participial form *uḷḷa*[4]). This is an important point in light of

[3] Cross-linguistically, possession could be encoded by some other functional head, not necessarily a *v*. In Malayalam the functional head must be verbal due to the restriction imposed by the relativizer. What is crucial for our analysis is not the category of the null element, but the fact that it encodes possession, and that possession plays a key role in encoding gradability in property concept expressions.

[4] A reviewer finds the null possessive verb objectionable on the grounds that postulating it amounts to suggesting that Malayalam has a null *have* while lacking an overt *have.* But notice that Malayalam does have an overt possessive copula, as seen in (15) and (16). Thus, the null possessive verb is the covert word-

criticism in Francez and Koontz-Garboden (this volume) that positing a null verb in Class 1 is unmotivated. If *–a* in *nalla* in (11) is the same relative marker as *–a* in (12b) and (17), then it is attaching to a null verb, since relative *-a* only ever attaches to verbs.

(11) a. avaḷ nalla-vaḷ aaṇə (Class 1)
she having-goodness-F.SG EQ-COP
'She is good.' (lit. 'She is one having goodness.')

b. avan nalla-van aaṇə
he having-goodness-M.SG EQ-COP
'He is good.' (lit. 'He is one having goodness.')

(12) a. avan kolayali aaṇə
he murderer EQ-COP
'He is a murderer.'

b. avaḷ keḷkkun-a-vaḷ aaṇə
she hear-REL.F.SG EQ-COP
'She is one who can hear.' (lit. 'She is one hearing.')

We see that Class 1 property concept roots participate in canonical predication – with the EQ copula. Thus, the *external* syntax and semantics of Class 1 expressions are compatible with the proposal made in Koontz-Garboden and Francez (2010), Francez and Koontz-Garboden (2015) concerning adjectives cross-linguistically. Our contribution in Menon and Pancheva (2014) is to show that the *internal* syntax and semantics of 'adjective-like' expressions also conforms to a generalization Koontz-Garboden and Francez (2010), Francez and Koontz-Garboden (2015) made concerning property-concept nominals cross-linguistically: creating attributive and predicative structures on the basis of nominals requires the use of verbs denoting possessive relations. The Class 1 property-concept denoting expressions of Malayalam are built on the basis of entity-denoting property concept roots, and thus they too use a possessive strategy, but this happens covertly below the word level.

2.2. Class 2 property concept expressions

Class 2 property concept roots end in *–am*, a Proto-Dravidian nominal marker. Thus, we take *–am* to be the spell out of a nominalizing head. In Menon and Pancheva (2014) we gave this nominalizing head the semantics in (13), suggesting that it turns abstract property concepts into measured instances of the property.

(13) $[\![\text{-}am_{n}]\!] = \lambda\Pi.\ \lambda d.\ \lambda x\ [x \text{ is an instance of } \Pi \text{ and } \mu(x) \geq d]$

---

internal counterpart of the overt possessive copula. Also, we posit additional evidence for the null possessive verb from our analysis of Complex color expressions in Section 4. In color terms, we find two versions – the Simple color term and the Complex color term – the latter showing a spell-out of the *v_poss* head.

The degree argument can be bound by an existential degree quantifier, as in (14), a measure phrase, or a comparative quantifier, resulting in structures without norm-related semantics (unlike the case of Class 1 forms, where, before the addition of *–a*, i.e., below the word level, *POS* binds the degree argument). Class 2 nouns denote predicates of individuals that are instances of *Π*, in contrast to Class 1 forms, which, as we suggested in Menon and Pancheva (2014), are predicates of individuals that possess instances of *Π*. The form in (14) can be the complement to a non-finite (*uḷḷa* in (15)) or finite (*uɳṭə* in (16)) existential (EX) copula which encodes possession. The individual argument is existentially closed off, as in regular possessive/existential predication.

(14) a. $[[[\sqrt{}\text{pokk} + am_{n}]_{n} + \exists D]$ (Class 2)
Lit. 'being an instance of tallness measuring to some degree'
b. $[[\ pokkam\ ]] = \lambda x\ \exists d\ [x$ is an instance of tallness and $\mu(x) \geq d]$

(15) pokkam uḷḷa kuṭṭi (Class 2)
tallness having child
'tall child.' (lit. 'tallness-having child.')

(16) avaḷkkə pokkam uɳṭə (Class 2)
she.DAT tallness EX-COP
'She is tall.' (lit. 'To her there is tallness.')

We see that Class 2 property concept roots participate in overt possessive predication, with the existential copula, as is to be expected from nominals on the account of Koontz-Garboden and Francez (2010) and Francez and Koontz-Garboden (2015). But we also see that with the help of the same syntactic mechanisms available to Class 1 forms – relativization with *-a*, and the creation of a light-headed relative with the help of pronouns – Class 2 forms can also participate in canonical predication, with the equative copula. This is illustrated below.

(17) a. avaḷ pokkam uḷḷa-vaḷ aaɳə (Class 2)
she tallness having-F.SG EQ-COP
'She is tall.' (lit. 'She is one having tallness.')

b. avan pokkam uḷḷa-van aaɳə (Class 2)
he tallness having-M.SG EQ-COP
'He is tall.' (lit. 'He is one having tallness.')

To summarize, the main proposals in our previous account (Menon and Pancheva 2014) are as follows, (i) possession, either covert or overt, is the basis for encoding property concept predication; (ii) property concepts universally lexicalize as category-less roots, and they denote abstract substance-like individuals, requiring possessive predication; (iii) variation in property concept predication is rooted in the morphosyntax and semantics of the functional vocabulary that categorizes property concept roots.

3. Revising the previous account

Our previous analysis posits two different heads which introduce gradability, namely the *v_poss* head for Class 1 and the *n* head spelled out as *-am* for Class 2. While for Class 1 expressions gradability is tied to the semantics of property possession, localized to *v_poss*, this is not so for Class 2 expressions. Uniformity will be theoretically preferable. Moreover, we need to posit two different nominal heads that are spelled out as *-am*, since the nominal *–am* appears also on nouns that do not have gradable semantics (e.g., the nominals in (18)). Then the question arises of why a verbal head could not combine with Class 2 roots, given the freedom of roots like the ones in (18a,b) to combine with both a verbal and a nominal head. The nominal nature of the functional head that combines with Class 2 roots is accidental as the same semantics could correspond to a verb.

(18) a. chaaṯ-uka ‘to jump’ – chaaṯ-am ‘a jump’
b. ooṯ-uka ‘to run’ – ooṯ-am ‘a run’
c. sneh-ikk-uka ‘to love’ – sneh-am ‘love’

Here we revise our analysis, arguing that both Class 1 and Class 2 roots compose first with verbal heads, but not the same verbal head. We retain our previous analysis for Class 1. We now suggest that Class 2 roots compose with a covert *v* head without possessive semantics. There are several implications to this new analysis. Since the covert *v* head doesn’t have possessive semantics, it does not introduce a degree argument. Only the covert *v_poss* has a degree argument. Crucially, this suggests that only Class 1 expressions are gradable. We will provide evidence for this when we look at how Class 1 and Class 2 expressions encode comparatives in Section 4.2. This new account is also desirable over the previous account since possession is expressed overtly in Class 2 expressions using the EX copula, and gradability can be encoded in the relation of possession of a property concept nominal. Thus, gradability, in our revised account, is tied directly to property possession, with both Class 1 and Class expressions. Finally, there is no need to posit two nominal *–am* morphemes in the language.

3.1. Revised Class 2 expressions

The Class 2 property concept roots first compose with the null *v* head without possessive semantics. The nominal marker *–am* then nominalizes this expression. Unlike the previous account, the null *v* head does not incorporate a degree argument. The existential copula then turns the nominal into a gradable predicate.

(19) a. $[[[\sqrt{\text{pokk}} + \varnothing_{v}] + am_{n}]_{n}$ (revised, Class 2)
b. $[\![\varnothing_{v}]\!] = \lambda \Pi.\ \lambda x\ [x \text{ is an instance of } \Pi]$
c. $[\![\textit{pokkam}]\!] = \lambda x.\ [x \text{ is an instance of tallness}]$

Class 2 expressions are not gradable. They are made gradable optionally overtly using the comparative marker or a measure phrase. An existential degree quantifier, without

norm-related semantics (unlike the case of Class 1 forms, where POS binds the degree argument), binds the degree argument introduced with possessive predication.

After the null *v* head composes with the Class 2 root and turns it into a verbal element, one could ask why the relative marker *–a* does not turn these forms into reduced relative clauses. We are aware that this is an issue (as also pointed out by a reviewer), but we do not have a good answer to this question. We could say that the morpheme in (19b) that turns property concept roots into predicates is not verbal but nominal in category. This would explain why the *–a* does not attach to it; recall that *–a* only attaches to verbs. However, as we will soon see, we need a null verbal head with the same semantics as in (19b) to account for the form and meaning of certain color expressions (see the discussion around ex. (23)). Once that head is available in the inventory of Malayalam, we will have to stipulate that it does not apply to Class 2 roots. Now we have to stipulate that Class 2 verbal forms have to be nominalized. At least in the color expressions (next section), we will see that the same root can compose with a verbal head (with different semantics) to form a verb (e.g., the two forms of 'white'). This suggests to us that Class 2 forms too are made verbal first and are then nominalized.

Additionally, one can ask why the Class 2 roots do not combine with the null *v_poss* and then with *–a.* In fact, in other related Dravidian language this is indeed what happens. In Kannada, Class 2 borrowed roots are turned into reduced relatives using *–a*. The semantics of these forms suggests that they incorporate a null *v-poss* before the addition of *–a.*

(20) a. santosha 'being happy' i.e., 'having happiness'
b. dukka 'being sad' i.e., 'having sadness'

As a reviewer points out, the Kannada forms in (20) can be *–am* ending nominal forms truncated to *–a* phonologically. We acknowledge that this may indeed be true given forms such as (21a), where '*santosha*' behaves similar to 'happiness' with the help of the past participial form of the "be" verb, namely '*agi*'. However, forms such as (21b) are also attested. These are similar to the pronominalization seen with Class 1 *–a* ending forms in Malayalam (11), suggesting that the forms in (20) show both nominal and adjectival properties.

(21) a. ii huḍuga santosha-agi idd-ane
this boy happiness-BE.PST BE-3P.SG
'This boy is happy.'

b. huḍuga santosha-vanu/ huḍugi santosha-vaḷu
boy happy-3M.SG/ girl happy-3F.SG
'The boy is happy.' 'The girl is happy.'

In Malayalam, as well as in Tamil, however, forms such as the ones in (20) do not occur. This could be a language-internal morphological fact: Class 2 roots can only combine with the non-possessive *v* just like Class 1 forms in that language only combine

with the possessive *v*. This is a stipulation, but we think that the insight into the link between possession and gradability that a compositional analysis of these forms allows us, makes the stipulation worthwhile. Note that an alternative lexicalist account similarly involves a stipulation: adjectival lexemes are based on native roots and borrowed roots are lexicalized as nominals.

3.2 Intermediate summary

The two classes of property concept roots undergo different syntactic derivations, but crucially start with, and end with, the same meaning. The possessive relation is expressed at the level of the word, through a covert possessive verbal morpheme, with Class 1 roots, and at the phrasal level, through an overt possessive verb, with Class 2 roots. Gradability is directly related to property possession. Only Class 1 lexemes are gradable, Class 2 expressions become gradable only at the phrasal level, through combination with a copula expressing possessive semantics.

4. Color expressions in Malayalam

In Menon & Pancheva (2014), we subsumed color expressions under Class 1, since they are *–a* ending terms. In this section, we will look in detail at color expressions in Malayalam suggesting that they exhibit both Class 1 and Class 2 properties, even though morphologically they belong to Class 1 property-concept expressions.

The revised analysis in the previous section has implications for the analysis of color terms. Color terms are *–a* ending, like Class 1 expressions. However, they exhibit two different forms with different syntactic behavior. We will refer to these forms as *Simple* and *Complex*. The term *Simple* is meant simply in opposition to *Complex*, the Simple form has internal morphological composition, similar to Class 1 expressions (except with a non-possessive *v*, as we will see shortly). The Complex form, although *–a* ending too, differs from Class 1 expressions in that it is overtly more complex. An exhaustive list of color terms in Malayalam is given in (22) below.

(22) *Color expressions in Malayalam*

ROOT	SIMPLE	COMPLEX	NOMINALIZATION	MEANING
√weɭ	weɭɭa	weɭɭutta	weɭɭuppə	‘white’[5]
√kaɾ	--	kaɾutta	kaɾuppə	‘black’
√kem	--	čuvanna	čuvappə	‘red’
√pačč	pačča	--	paččappə[6]	‘green’

[5] Only the root for ‘white’ exhibits both simple and complex forms. We have to say that the lack of one form or the other for the remaining color roots is an accidental gap. However, we do note that there exist words for ‘red’ and ‘black’ that look simple but semantically mean different things, such as *kara* ‘stain’ and *čuva* ‘taste’. These two terms are nouns and not reduced relatives as seen by the presence of overt case marking.

√niil	niila	--	--	'blue'
√maɲ	maɲɲa	--	--	'yellow'
√čaar	čaara	--	--	'ash grey'
√uut	uuta	--	--	'violet'

Among the color roots, the roots for 'white', 'black', 'red', and 'green' (√weɭ, √kaɾ, √kem, √paččˇ) can be identified as having Proto-Dravidian origins. These roots have cognates in other Dravidian languages. The roots for 'blue' and 'violet' are borrowings from Hindi-Urdu, and the root for 'yellow' is derived from the Tamil word for turmeric 'maɲɲal'. Apart from these, English terms for colors such as 'cream', 'rose', 'pink', 'orange' have been borrowed into the Dravidian lexicon and they are pronounced with morphology used in borrowing, such as the epenthetic vowel (ə in Malayalam, u in Telugu etc).

4.1 Syntactic and semantic behavior of color expressions

4.1.1. Simple color terms

Simple color terms morphologically resemble Class 1 expressions. However, they depart in their syntactic behavior exhibiting similarities with Class 2 expressions. Our proposal is to treat the Simple color terms similar to Class 2 forms up to a point. They compose with a *v* head with no possessive semantics and no degree argument. This is intact with our earlier observation about the composition of *–a*, the Proto Dravidian relative clause marker, with verbal elements. Thus, Simple color terms are turned into participial verbal expressions by the addition of a null *v* as in (23), the same null *v* that derives Class 2 expressions according to (19b). Recall that for Class 2 forms, positing the null non-possessive *v* was stipulative, at least as far as the syntactic category of the functional element is concerned. For Simple color terms though, the *v* is justified on the ground that the relative marker *–a* attaches next.

(23) $[\![\varnothing_v]\!] = \lambda\Pi\, \lambda x\, [x \text{ is an instance of } \Pi]$

The *v*Ps that are created by the merge of the null non-possessive *v* are further relativized by the verbal relative marker *–a*, see (24). This changes the syntactic category, as the structure is now participial; the semantic type remains unchanged. Note that the Simple color term *weḷḷa* means 'being an instance of whiteness', which is a predicate of individuals but of semantically different sortal type than the NP *kuppayam* 'dress' with which the color expression seemingly combines in (25a). This sortal mismatch disallows a Simple color form as in (24c) to combine with a noun phrase in the attributive position directly. Our proposal is to posit a null covert color expression, COLOR (à la Kayne 2005), which the Simple color terms modify, and a null expression of possession, (25b). Note

[6] Note, however, that *paččappə* means 'greenery'. All the other nominalizations refer to the color itself.

that both null elements posited in the attributive structure (24b), which surfaces as (24a), can be seen overtly in (25c): the nominal *niram* 'color' and the possessive non-finite copula *ull-a*, the same strategy employed in attributive modification in Class 2 expressions. The null possessive element POSS plays the role of linking together the Simple color nominal expression 'white color' with the noun phrase it modifies. We leave the precise formulation of this structure for future work.[7]

(24) a. $[[\sqrt{\text{wel}} + \varnothing_v]_v$ (Simple color)
Lit. 'be an instance of whiteness'
b. $[[\sqrt{\text{wel}} + \varnothing_v]_v + \textit{-a}\ ]_{rel}$
Lit. 'being an instance of whiteness'
c. $[\![\text{wel-a}]\!] = \lambda x.$ [x is an instance of whiteness]

(25) a. wel-a kuppayam (Simple color)
being-whiteness dress
'a white dress' (lit. #'being an instance of whiteness dress')

b. wel-a COLOR POSS kuppayam
being-whiteness color having dress
'a white dress' (lit. 'being a white color having dress')

c. wel-a niram ul-a kuppayam
being-whiteness color having-REL dress
'a white dress' (lit. 'being a white color having dress')

Simple color terms can appear in the predicative position with the EQ copula without the help of a bound pronominal morpheme, see (26a). This suggests that in this structure too, just like in the case of the attributive structure, Simple color terms combine with a null nominal COLOR; the presence of this nominal makes the pronominal forms unavailable. The structure behind the surface predication in (26a) is as in (26b). Note that the posited covert nominal COLOR can also be overt (27). The presence of the covert nominal makes the use of bound pronouns with the EQ copula (as in Class 1 predication) unnecessary, by providing the EQ copula with a nominal complement.[8]

(26) a. kuppayam wela aaṇə (Simple color)
dress being-whiteness EQ-COP
'The dress is white.'

[7] The question of the nature of the possession-encoding linker in (25a/b) – or even its presence – is tied to the structure of the predicative use of Simple color terms, as in (26a,b) and (27a).

[8] A question arises as to whether there is a possession-encoding linker in (26a/b) in addition to the null COLOR. We would expect that to be the case, given the discussion concerning the attributive use of Simple color terms as in (25). The EQ copula can also take PPs, given its use in locatives, so a complement like 'of a white color' could be possible in (26a/b). However, (27a), with overt *niram* 'color' shows no such linking element. Similarly the interpretation of (28a) suggests the absence of a possession-encoding element in these predicative structures. We leave this question open for the future.

b. kuppayam weḷḷa COLOR aaɳə
dress being-whiteness color EQ-COP
'The dress is white.'

(27) a. kuppayam weḷḷa niram aaɳə
dress being-whiteness color EQ-COP
'The dress is of a white color.'

b. kuppayat-inə weḷḷa niram unʈə
dress-DAT being-whiteness color EX-COP
'The dress is white color.'

We see that Simple color terms morphologically belong to Class 1 property concept expressions but exhibit peculiar syntactic behavior in predicative position. We suggested that Simple color terms start out as roots that are made verbal using the non-possessive *v* head (the same head we suggested combines with Class 2 non-color roots); then they are relativized by *-a*. In attributive position, Simple color terms modify a null covert nominal COLOR. They also combine with additional structure which encodes possession, enabling modification between the Simple color term and the NP. Simple color terms also exhibit canonical predication with the help of the equative copula, however, unlike Class 1 expressions, they do not require a pronominal element. We capture this by positing the same null covert nominal COLOR in the predicative position as in the attributive position, which the Simple color term composes with. This covert nominal plays a similar role to the bound pronominal in Class 1 predication, converting the participial *–a* form into a nominal expression. This null nominal is likely behind the ability of color terms in languages like English to appear in nominal positions, as seen in (3).

Importantly, Simple color terms are non-gradable (as shown in (28)). Gradability is introduced by the overt existential copula in predicative position, just as is the case with Class 2 property concept expressions.

(28) a. traffic light paccca COLOR aaɳə
traffic light green color EQ-COP
'The traffic light is green.'

b. # traffic light-inə paccca niram unʈə
traffic light-DAT green color EX-COP
'The traffic light is green.' (Lit. 'The traffic light has green color')

c. # ii traffic light aa traffic lightin-e kaaḻ-um kuuʈuttal paccca COLOR aaɳə
this traffic light that traffic light-ACC than more green color EQ-COP
'This traffic light is greener than that traffic light.'

As Kennedy and McNally (2010) note and as shown in (2), it is the traffic light's classificatory property to be green in color; the property either obtains or it does not. Thus, as we can see in the Malayalam examples, when the Simple color term is used in (28a), only a non-gradable, classificatory reading is available. Gradability cannot be overtly introduced here, using the possessive copula, as in (28b), which results in infelicity. The comparative is also disallowed (similar to # *more female*).

Moreover, Simple color terms can be used as classificatory modifiers, as in (29), which are non-gradable (Kennedy and McNally 2010).

(29) a. paccca we[[am
green water (lit. 'fresh water')

b. we[[a wine
white wine (in fact, yellow in color)

In the next section, we will see that Complex color terms syntactically and morphologically behave differently from Simple color terms.

4.1.2 Complex color terms

Complex color terms also morphologically resemble Class 1 expressions, in that they are *–a* ending, but they depart from Class 1 non-color and Simple color expressions in exhibiting more complex overt morphology. Despite their overt morphological complexity, they pattern similar to Class 1 non-color expressions in their syntactic behavior. We analyze them accordingly. Complex color terms start out as roots that compose with a null *v_poss* with possessive semantics (30) (repeated from (8)). We analyze the *–utt* morpheme as a spell-out of the *v_poss*, it is the overt morpheme counterpart of the null possessive *v* head[9] (the possessive copula, as we discussed earlier, is another overt possessive verbal form, morphologically free rather than bound).

(30) $[\![\varnothing_{v\_poss}]\!] = \lambda \Pi\, \lambda d\, \lambda x\, \exists y\, [y \text{ is an instance of } \Pi \;\&\; x \text{ has } y \;\&\; \mu(y) \geq d]$

The meaning of Complex color forms is norm-related – they are interpreted as making reference to a standard, as would be expected if POS is binding the degree variable rather than a regular existential degree quantifier. The meaning given in (31b) is similar to the meaning assigned to positive gradable adjectives such as 'good' in English by many semantic accounts.

(31) a. $[[[\sqrt{\text{we[}} + \text{-utt}_{\text{v\_poss}}]_{\text{v}} + \text{POS}]_{\text{v}} \text{-a}]_{\text{rel}}$ (Complex color)

[9] There is some variability in the phonological realization of this morpheme. As seen, complex 'black' and 'white' have the *–utt* morpheme, whereas complex 'red' has an *-ann* morpheme. We take this to be a phonological fact depending on the coda position of the root, lateral ending for *–utt* and nasal for *–ann*. They are both spell-outs of the null *v_poss* head.

Lit. 'having an instance of whiteness measuring to a degree that exceeds the standard'

b. $[\![$we$\underline{ɭ}$-utt-a$]\!]$ = $\lambda x.\ \exists d\ \exists y$ [y is an instance of whiteness and x has y and $\mu(y) \geq d$ and $d > d_s$]
$\approx \lambda x.\ \exists d$ [x's whiteness $\geq$ d and $d > d_s$]

The participial *-a*-forms can be used in attributive position – they have the appropriate participial syntax as well as semantics to be interpreted through predicate modification with nominals such as *dress*. Unlike Simple color expressions, there is no null COLOR in attributive position in the case of Complex color expressions.

(32) a. we$\underline{ɭ}$-utt-a kuppayam (Complex color)
having-whiteness dress
'a white dress' (lit. 'having whiteness dress')

b. not: we$\underline{ɭ}$-utt-a COLOR kuppayam (Complex color)
having-whiteness color dress

The participial *-a*-forms can also be used in predicative position, after they are turned into light-headed relatives, i.e., DPs, through the merge of bound pronouns. Given the absence of the null nominal COLOR with Complex color terms, the obligatory presence of the pronominal is expected.

(33) itə we$\underline{ɭ}$-utt-a-tə aaɳə (Complex color)
this having-whiteness-neut EQ-COP
'it which has whiteness' (lit. 'it having goodness')

Complex color expressions morphologically and syntactically behave like Class 1 non-color expressions. They use a covert possessive strategy and display canonical predication using the equative copula. They also show overt evidence for the existence of a possessive *v* head, which spells out as the *–utt/-ann* morpheme.

4.2. Implications for comparatives

The analysis we have sketched out predicts an asymmetry in comparison. Simple color terms should behave similar to Class 2 expressions in allowing the comparative marker 'more' in structures with the possessive predicate needed to introduce gradability. Complex color terms should behave similar to Class 1 non-color expressions in disallowing an overt comparative marker, as the degree variable introduced by the possessive *v*, with which *–a* combines, is closed off by POS. The color terms behave exactly as predicted by the analysis. In both predicative and attributive forms, *kuuʈuttal* 'more' (a form that only appears in comparatives) is optionally allowed with Simple color terms and disallowed with Complex color terms.[10] In the case of Simple color

[10] The issue of optionality of *kuuʈuttal* 'more' is addressed in Menon 2015, in prep.

terms, the attributive form provides further evidence for our covert nominal COLOR. It has to be overtly pronounced.

(34) a. Simple color: predicative

ii	kuppayam	aa	kuppayathin-e	*kaaɭum*	*(kuuʈuttal)*	weḷḷa	aaṇə
this	dress	that	dress-acc	than	more	white	EQ-COP

'This dress is whiter than that dress.'
(lit. This dress is white color more (so) than that dress.')

b. Simple color: attributive

Anil	Komalan-e	*kaaɭum*	*(kuuʈuttal)*	weḷḷa	niram	uḷḷa	kuppayam	iʈʈu
Anil	Komalan-acc	than	more	white	color	having	dress	wore

'Anil wore a whiter dress than Komalan.'
(lit. 'Anil wore a dress having whiteness more than the dress Komalan was wearing.')

c. Complex color: predicative

ii	kuppayam	aa	kuppayathin-e	*kaaɭum*	*(*kuuʈuttal)*	weḷḷutta-tə	aaṇə
this	dress	that	dress-acc	than	more	white-pron	EQ-COP

'This dress is whiter than that dress.'
(lit. This dress is having whiteness more than that dress.')

d. Complex color: attributive

Anil	Komalan-e	*kaaɭum*	*(*kuuʈuttal)*	weḷḷutta	kuppayam	iʈʈu
Anil	Komalan-acc	than	more	white	dress	wore

'Anil wore a whiter dress than Komalan.'
(lit. 'Anil wore a dress having whiteness more than Komalan.')

In the comparatives in (34), the semantics of comparison is encoded in the standard marker *than*. We leave a detailed analysis of these comparatives for future work (See Menon 2015 and Menon in prep).

5. Summary and conclusions

We give below an updated summary of the structures for Class 1, Class 2, Simple color, and Complex color forms. The two classes of property concept roots participate in different structures, but both start with, and end with, the same meaning. Simple color terms are similar to Class 2 forms in that they are non-gradable and do not have a *v*_poss in their composition; Complex color terms are similar to Class 1 expressions in that they are formed with *v_poss* and are thus gradable.

(35) Class 1: native roots, non-color

a. $[[[_{vP}\ \sqrt{1}+ \varnothing_{v\_poss}\ ] + POS\ ] + a_{rel}\ ]$ (attributive)
'having an instance of Π that exceeds the standard'

b. $[_{vP}\ [_{DP}\ [[[_{vP}\ \sqrt{1} + \varnothing_{v\_poss}\ ] + POS\ ] + a_{rel}\ ]\ pron\ ]\quad EQ.COP\ ]$ (predicative)
'be someone having an instance of Π that exceeds the standard'

(36) Class 2: borrowed roots

a. $[[_{vP}\ [_{DP}\ [_{vP}\ \sqrt{2} + \varnothing_{v}\ ] + am_{n}]\quad EX.COP_{non\text{-}finite}\ ] + a_{rel}\ ]$ (attributive)
'having an instance of Π'

b. $[_{vP}\ [_{DP}\ [_{vP}\ \sqrt{2} + \varnothing_{v}\ ] + am_{n}]\quad EX.COP\ ]$ (predicative)
'have an instance of Π'

c. $[_{vP}\ [_{DP}\ [[_{vP}\ [_{DP}\ [_{vP}\ \sqrt{2} + \varnothing_{v}\ ] + am_{n}]\ EX.COP_{non\text{-}finite}\ ] + a_{rel}\ ]\ pron\ ]\ EQ.COP\ ]$
'be someone having an instance of Π' (predicative)

(37) Simple color

a. $[[_{vP}\ \sqrt{SC} + \varnothing_{v}\ ] +\ a_{rel}\ ]$ (attributive to COLOR)
'being an instance of Π_color'

b. $[_{DP}\ [[_{vP}\ \sqrt{SC} + \varnothing_{v}\ ] +\ a_{rel}\ ]\quad COLOR\ ]$ (attributive to nouns)
'being a Π_color'

c. $[_{vP}\ [_{DP}\ [[_{vP}\ \sqrt{SC} + \varnothing_{v}\ ] + a_{rel}]\quad COLOR\quad]\quad EQ.COP\ ]$ (predicative)
'be an instance of Π_color'

d. $[_{vP}\ [_{DP}\ [[_{vP}\ [_{DP}\ [[_{vP}\ \sqrt{SC} + \varnothing_{v}]\ +\ a_{rel}]\ COLOR]\ EX.COP_{non\text{-}finite}\ ] + a_{rel}\ pron]$
EQ.COP] (predicative)
'be someone having an instance of Π_color'

(38) Complex color

a. $[[[_{vP}\ \sqrt{CC} + \varnothing_{v\_poss}\ ] + POS\ ] + a_{rel}\ ]$ (attributive)
'having an instance of Π_color that exceeds the standard'

b. $[_{vP}\ [_{DP}\ [[[_{vP}\ \sqrt{CC} + \varnothing_{v\_poss}\ ] + POS\ ] + a_{rel}\ ]\ pron\ ]\ EQ.COP\ ]$ (predicative)
'be someone having an instance of Π_color that exceeds the standard'

In this paper, we demonstrated that in Malayalam, a language that does not have a category of adjectives, adjective-like meanings for attributive modification and predication involving color terms are expressed by complex structures built from roots denoting property concepts. We provided an analysis of color expressions, by updating our previous account of the internal structure of one class of property concept expressions (Class 2 nominals). We argued that possession, either covert or overt is the basis for encoding property concept predication, including with color terms. Variation in property concept predication is rooted in the morphosyntax and semantics of the functional vocabulary that categorizes property concept roots. Color expressions come in two avatars, a simple form that needs overt possessive predication to be gradable and a complex form, which encodes possession covertly, is gradable, and exhibits canonical predication. We thus provided evidence from Malayalam for a distinction between gradable and non-gradable color expressions, posited by Kennedy & McNally (2010) and

McNally (2011) as a lexical ambiguity, and argued here to be a distinction rooted in the morphology rather than the lexical semantics of the color terms.

An alternative analysis such as the semantic variation hypothesis, in Francez and Koontz-Garboden (this volume), also needs to stipulate why Class 1 roots end up behaving like adjectives and Class 2 roots end up as nominals. Our analysis hinges on possession both below and above the word level. Below the word level, possession is introduced by a functional head that also introduces a degree argument. This head turns out to be a verbal head in Malayalam, although cross-linguistically this functional head could have a different category. Above the word level, possession is encoded using the possessive copula, which also contributes gradability. Our account posits that the source of variability in the behavior of property-concept expressions, within Malayalam, and likely cross-linguistically as well, is morpho-syntactic variation, a consequence of structure building processes, and not variability encoded in the lexicon.

References

Alexiadou, A. 2013. Deriving color adjectival nominalizations. *Linguistica 8*, 143-158.

Balusu, R. 2014. Are there adjectives in Telugu? Handout of paper presented at ICOLSI 36, University of Kerala, December 2014.

Francez, I. and A. Koontz-Garboden. 2015. Semantic variation and the grammar of property concepts. *Language*. 91: 3.

Francez, I. and A. Koontz-Garboden. this volume. Malayalam property concept sentences and the locus of variation. In R. Balusu and S. Sundaresan (eds.) *Proceedings of FASAL V*.

Gamliel, O. 2013. Voices yet to be heard: On listening to the last speakers of Jewish Malayalam. Journal of Jewish Linguistics 1, Leiden: Koninklijke Brill NV, 135-167.

Kayne, R. 2005. *Movement and Silence*. New York: Oxford University Press, pp. 369.

Koontz-Garboden, A. and I. Francez. 2010. Possessed properties in Ulwa. *Natural Language Semantics*. 18: 197-240.

Kennedy, C & L. McNally. 2010. Color, Context and Compositionality. *Synthese* 174.1: 79-98.

McNally. 2011. Color terms: A case study in natural language ontology. Talk given at the *Workshop on the Syntax and Semantics of Nounhood and Adjectivehood*, UAB/CSIC, Barcelona, 25/3.

McNally, L. & H. de Swart. 2011. Inflection and derivation: How adjectives and nouns refer to abstract objects. *Pre-proceedings of the 18th Amsterdam Colloquium*, 425-434.

Menon, M. 2013. The apparent lack of adjectival category in Malayalam and other related languages. In N. Goto, K. Otaki, A. Sato, K. Takita (eds.). *Proceedings of GLOW-in-Asia IX 2012*, 157-171.

Menon, M. and R. Pancheva. 2014. The grammatical life of property concept roots in Malayalam. In U. Etxeberria, A. Fălăuș, A. Irurtzun, B. Leferman eds., *Proceedings of Sinn und Bedeutung* 2013. pp. 289-302.

Menon, M. 2015. *Than* encodes comparative semantics. Talk given at CUSP 8, Stanford University. November 06 - 07 2015.

Menon, M. in prep. Predication, Modification, and Comparison without Adjectives. Ph.D dissertation. University of Southern California.

Moltmann, F. 2013. On the Distinction between Abstract States, Concrete States, and Tropes. In A. Mari/ C. Beyssade/Del Prete, F. (eds.): *Genericity*. Oxford University Press, Oxford, pp. 292-311.

Mythili Menon & Roumyana Pancheva
mythilim@usc.edu
pancheva@usc.edu

Malayalam property concept sentences and the locus of variation*

Itamar Francez & Andrew Koontz-Garboden

University of Chicago & University of Manchester

1. Introduction

In Francez & Koontz-Garboden (2015), we call attention to an intra- and crosslinguistic generalization about the form of what we call (in the spirit of Dixon 1982, Thompson 1989) PROPERTY CONCEPT SENTENCES, translational paraphrases of sentences whose main predicate is an adjective in some language, such as (1).

(1) Krishna is wise.

We observe that both internal to a single language and crosslinguistically, such sentences can be either **predicative** or **possessive**. For example, internal to English, the canonical predicative adjective sentence in (1) can be paraphrased with (2), which takes the form of an English predicative possessive sentence.[1]

(2) Krishna has wisdom.

The possessive pattern is restricted in English; relatively few property concept propositions can be encoded with such sentences. In other languages, however, it is much more robust and sometimes practically exclusive, as is the case in Ulwa, a Misumalpan language spoken in Nicaragua, the focus of Francez & Koontz-Garboden (2015). The general question we are concerned with there is what the source of variation in the morphosyntactic form of property concept sentences is. What determines whether a property concept sentence is morphosyntactically possessive or predicative? Two kinds of answers suggest themselves.

*We thank Mythili Menon for engaging with us, both empirically and theoretically, on this material. We also benefited from audience comments at FASAL. This work has been supported financially by Arts and Humanities Research Council Grant AH/H033645/1.

[1]By *canonical predication* we mean the morphosyntactic form used in ordinary non-verbal predication, as with predicative nominals and predicative adjectives. See Francez & Koontz-Garboden (2015) for further discussion.

One kind of answer links the differences in morphosyntax to **semantic variation**, specifically to variation in the lexical semantics of what we term the *property concept lexeme*—the lexical item in a property concept sentence responsible for introducing the 'adjectival' descriptive content (*wise* in (1), *wisdom* in (2)). On this view, the distribution of possessive and predicative form reflects a contrast between possessive and predicative semantics. The semantic choice (between possession and predication) is governed by what is required, semantically, to express the truth conditions of a property concept sentence. This line of explanation, linking form directly to meaning, is similar to that pursued by Talmy (1985) in relation to the motion typology, and by Chierchia (1998) in relation to variation in the distribution of numeral classifier morphosyntax.

An alternative line of analysis is to assume that the semantic components are uniform across property concept sentences, viewing the variation in their morphosyntactic form as reflecting **syntactic variation**, specifically variation in how the semantic components are linked to syntactic elements in particular languages. That semantic components are universal and uniform within and across languages is the null hypothesis according to the strong view argued for by Matthewson (2001) and those following her. In particular, on this view, the distribution of surface possessive and predicative forms is an artifact of language specific facts about the inventory and phonology of functional material.

In this paper, we examine the consequences of choosing between these two alternatives, drawing on the expression of property concept sentences in Malayalam, as described and analyzed by Menon and Pancheva (2014). In Francez & Koontz-Garboden (2015) we argue for a semantic variationist explanation for the distribution of possessive and predicational forms, linking it to the lexical semantics of property concept lexemes. In contrast, Menon and Pancheva argue, based on Malayalam, for a syntactic variationist explanation. Specifically, they argue that property concept sentences are universally possessive. Whether they are possessive or predicational on the surface depends on the inventory and phonological realization of functional morphemes. Our goal is to reexamine Menon and Pancheva's data, arguing that they are compatible with the semantic variationist explanation, and in fact lend more support for it. We show that the syntactic variationist position leads to missed generalizations and to crosslinguistic expectations that do not seem to be met. Our conclusion is therefore that overall the observed data are best explained by semantic variation, and, more broadly, that at least some morphosyntactic variation in the form of translationally equivalent sentences[2] in some is ultimately due to semantic distinctions (viz., Chierchia 1998; Matthewson 2001 for discussion).

We begin with discussion of the semantic variationist analysis of the phenomenon in Francez & Koontz-Garboden (2015), and then turn to the Malayalam data that Menon and Pancheva bring to bear on the discussion. After laying out the facts as described by Menon and Pancheva, we describe their analysis, and point out the problems it faces. We then show how a semantic variationist analysis can capture the facts without running into any of these problems. We conclude with some broader observations.

[2]Translational equivalence is a problematic concept at best (see e.g., Keenan 1973 for some discussion). However, for current purposes, we follow standard practice in the field and assume, perhaps somewhat optimistically, that this intuitive concept corresponds to some theoretically viable notion. See Francez & Koontz-Garboden (In prep, Chapter 1) for further justification.

2. The (lexical) semantic variationist view

The simple intuition underpinning our analysis of variation in the form of property concept sentences in Francez & Koontz-Garboden (2015) is that possessive morphosyntax reflects possessive semantics and predicative morphosyntax reflects predicative semantics. What underlies the choice of strategy is the semantics of **property concept lexemes**, i.e. the simplest lexemes contributing the "adjectival" content in a sentence expressing property concept predication. Property concept lexemes come in two semantic varieties. There are those which are **individual characterizing**, like *wise*, which characterizes the set of wise individuals, and those which are **substance characterizing**, like *wisdom*, which characterizes the set of portions of the substance wisdom. In informal terms, the proposal in Francez & Koontz-Garboden (2015) is that substance characterizing lexemes require a semantics of possession to express the truth conditions of property concept predication. This is because predicating a predicate of substances of an individual does not yield the relevant meaning. For example, (3a) simply does not have the truth conditions of (3b).

(3) a. Kim is wisdom. $\neq$
b. Kim is wise.

It is a basic empirical observation, illustrated by (3), that direct predication with substance characterizing lexemes does not generate translational equivalents of property concept sentences. However, possessively relating individuals to substances *does* yield a proposition that is true whenever the relevant property concept proposition is true:[3]

(4) Kim is wise. $\Rightarrow$ Kim has wisdom.

The main point in Francez & Koontz-Garboden (2015) is that whether possession or canonical predication is used for expressing a property concept proposition is entirely predictable from the denotation of the basic property concept lexeme involved, a denotation that, at least in most cases, can be independently diagnosed, for example by investigating what kind of truth conditions are generated with the lexeme in direct predication.

The formal details of this theory are developed in Francez and Koontz-Garboden (2015, 2016, In prep), building on the observation that, in familiar languages, substance characterizing property concept lexemes behave in many ways like mass nouns. In Francez & Koontz-Garboden (2015), they are given denotations similar in key respects to those of other mass nouns, building on Link's (1983) foundational work. The idea is that individual and substance characterizing property concept lexemes denote in different domains. Individual-characterizing lexemes denote in the domain of ordinary individuals (in some way or another, depending on the theory of gradability assumed), whereas substance-characterizing lexemes denote in a separate domain, the domain of portions. This domain is totally pre-ordered by a relation called 'size'. Substance-characterizing property concept

[3]The reverse direction does not always hold, at least for English, which is one of the reasons why it is not obvious how to formulate a notion of translational equivalence in truth conditional terms.

lexemes denote *substances*, which are mutually disjoint, partially ordered subsets of the domain of portions. A theory of substance possession is then outlined in which saying that an individual "has" a substance is true if and only if it bears a possessive relation to some portion of that substance.

With this as background, we now turn to Menon and Pancheva's challenge to this picture from Malayalam.

3. Malayalam: The descriptive facts

Menon and Pancheva (2014) lay out a detailed description of the facts of property concept sentences in Malayalam, which we summarize in this section. According to them, there are two classes of property concept lexeme in the language, which they call Class 1 and Class 2. At a purely descriptive level, the first of these is canonically predicating while the other is possessively predicating. Both classes are robustly attested in the language.

3.1 Class 1

Class 1 roots are those which become free words when suffixed with *–a*, as shown by the data in (5).

(5) Malayalam property concepts words in *–a* (Menon & Pancheva 2014, 290)
valiya 'big'; *čeriya* 'small'; *puthiya* 'new'; *nalla* 'good'; *paččа* 'green'; *niila* 'blue'

As discussed by Menon and Pancheva, the suffix *–a* is, diachronically, a relativizer. Whether it should be taken to be a relativizer synchronically or not is a matter of debate: Asher & Kumari (1997, 116–117, 350) reject this idea, while others, including Menon & Pancheva (2014, 290) argue that it is indeed synchronically productive. We are not in a position to contribute to a resolution of this issue, but nothing we say here hinges on it.

For reasons that seem to be mysterious to everyone, and will remain so here, these relativized forms, in order to be used as predicates, must be turned into light-headed relatives, using suffixes which Menon and Pancheva claim are bound pronouns (Menon & Pancheva 2014, 292). Examples are given in (6) (the glosses throughout are Menon and Pancheva's).

(6) a. *nalla-va⌊*
good-F.SG
'she who is good'
b. *nalla-van*
good-M.SG
'he who is good' (Menon & Pancheva 2014, 292)

These light-headed relatives serve as the main predicates in property concept sentences in which these Class 1 roots feature, as illustrated by the data in (7).

(7) a. *avaɭ nalla-vaɭ aaɳə*
she good-F.SG EQ-COP
'She is good.'
b. *avan nalla-van aaɳə*
he good-M.SG EQ-COP
'He is good.' (Menon & Pancheva 2014, 292)

Such sentences feature what Menon and Pancheva call the "equative copula" *aaɳə*, the copular element generally used for non-verbal predication in Malayalam, as shown by the data in (8).

(8) a. *avan kolayali aaɳə*
he murderer EQ-COP
'He is a murderer.'
b. *avaɭ keɭkkun-a-vaɭ aaɳə*
she hear-REL-F.SG EQ-COP
'She is one who can hear.' (Menon & Pancheva 2014, 293)

3.2 Class 2

Class 2 roots are those that combine with the suffix *–am* to form nouns. Some property concept words in this class are given in (9):

(9) *santosham* 'happiness'; *sankaṭam* 'sadness'; *madhuram* 'sweetness'; *prayasam* 'difficulty'; *santam* 'quietness'; *pokkam* 'tallness' (Menon & Pancheva 2014, 290)

The suffix *–am* is "a productive nominal marker in Malayalam" (Menon & Pancheva 2014, 293), as evidenced by the fact that it forms nouns not only from property concept roots, but also from roots that form verbs, as in (10), and other ordinary nouns, as in (11).

(10) a. *chaaṭ-uka* 'to jump'; *ooṭ-uka* 'to run'; *snek-ikk-uka* 'to love'
b. *chaaṭ-am* 'a jump'; *ooṭ-am* 'a run'; *sneh-am* 'love'
(Menon & Pancheva 2014, 293)

(11) *paz-am* 'banana'; *vell-am* 'water'; *kall-am* 'theft' (Mythili Menon, p.c.)

In contrast with Class 1 roots, Class 2 roots combining with *–am* form property concept sentences with possessive morphosyntax. Ordinary possessives in Malayalam are created with an existential construction, using a special copula (which Menon and Pancheva call the "existential copula") and with a dative marked possessor, as shown in (12).

(12) *avaɭkkə mookutthi uɳṭə*
she.DAT nose.pin EX.COP
'She has a nose pin.' (Menon & Pancheva 2014, 294)

Precisely the same construction is used to create property concept sentences with Class 2 property concept roots suffixed by *–am*, as shown by the data in (13).[4]

(13) *avaɭkkə pokkam uɳʈə*
she.DAT tallness EX.COP
'She is tall.' (Menon & Pancheva 2014, 294)

3.3 Summary of the descriptive facts in Malayalam

To summarize, descriptively speaking, there are two classes of property concept lexemes in Malayalam. Property concept lexemes in what Menon and Pancheva call Class 1 appear in property concept sentences suffixed with relative clause forming morphology and with morphosyntax otherwise used for canonical non-verbal predication in the language. Class 2 lexemes appear in property concept sentences suffixed with nominalizing morphology and with possessive morphosyntax.

Notwithstanding this contrast in morphosyntactic behavior, Menon and Pancheva argue for an analysis which, at a deeper level, treats the two classes as a single, semantically uniform class of roots, with property concept sentences based on them also having an identical semantics, but involving different functional heads. We turn next to the details of this analysis.

4. Menon and Pancheva's analysis of Malayalam

Menon and Pancheva (2014) analyze the Malayalam facts within a Distributed Morphology framework, with the goal of maintaining a uniform lexical semantics for property concept lexemes, and for property concept sentences generally. Their analysis locates the variation in the morphosyntactic form of property concept sentences, in Malayalam and, by hypothesis, crosslinguistically, in syntax and morphophonology. The key assumption of the analysis is that property concept lexemes are universally precategorial roots, and universally denote *properties* (in the property-theoretic sense of Chierchia & Turner 1988, following Koontz-Garboden & Francez 2010). Property concept sentences built on these roots always

[4]Comparatives based on Class 2 are also built on a possessive morphosyntax, by contrast with those in Class 1, as shown by the data in (i) and (ii). Menon and Pancheva make the interesting observation that comparatives based on class 1 property concept roots, disallow *kuuʈuttal* 'more', whereas those based on class 2 roots optionally allow it, as (i) shows.

(i) a. *Anil-inə Komalan-e kaaɭ-um (kuuʈuttal) pokkam uɳʈə*
Anil-DAT Komalan-ACC than-UM more tallness EX.COP
'Anil is taller than Komalan.' (Menon & Pancheva 2014, 299)

b. *Anil Komalen-e kaaɭ-um nalla vidhyarthi aaɳə*
Anil Komalan-ACC than-UM good student EQ-COP
'Anil is a better student than Komalan.' (Menon & Pancheva 2014, 299)Why this is the case is unclear to us.

express the proposition that an individual possesses an "instance" of the property denoted by the root. In cases where there is canonical predication on the surface, as with Malayalam Class 1 roots, possession is introduced covertly by a phonologically null categorizing head. In the case of Malayalam Class 1 roots, the null categorizing head is a *v* head, assumed to have possessive semantics. When possessive morphosyntax *is* seen on the surface, this is because possessive semantics has not been introduced by the head categorizing the root. This is the case with Malayalam Class 2 roots, which are categorized by a *n* head, spelled out as *–am*, which does not introduce possessive semantics. With such words, possessive semantics is instead introduced overtly by a possessive construction. In what follows, we detail how this analysis works with the two classes of Malayalam property concept roots, following this by consideration of how they intend their analysis to be seen in a crosslinguistic context. We then consider the plausibility of the predictions that this analysis gives rise to.

4.1 Class 1 roots

Class 1 roots are turned into possessive predicates by a functional verbalizer *v* with possessive semantics, which also introduces a degree argument. Formally speaking, as shown in (14), this phonologically null *v* head takes a property denoting root and creates a degree predicate from it— a function from degrees to a function from ordinary individuals to truth values which when predicated of an individual a and a degree d_1 is true just in case there is some instance of the root property[5] that a has and the measure of the instance of the root property that a has is greater than or equal to d_1.

(14) $[\![\emptyset_{v-poss}]\!] = \lambda\Pi\lambda d\lambda x\exists y[y \text{ is an instance of } \Pi \;\&\; x \text{ has } y \;\&\; \mu(y) \geq d]$

As elsewhere in the degree-based literature on property concept sentences, a key question concerns the source of the degree argument in any particular sentence with a gradable predicate in it. In comparatives, for example, the degree argument is saturated by comparative degree morphology. In ordinary predicative contexts like (15), the assumption is that there is a "positive" degree which is responsible for the vagueness of such sentences and which is often phonologically null.

(15) *avan nalla-van aaŋə*
he good-M.SG EQ-COP
'He is good.' (Menon & Pancheva 2014, 292)

Menon and Pancheva avail themselves of such an analysis, proposing that in such contexts, the degree argument of the verbalized property concept root is saturated by the positive degree morphology with the denotation in (16).

[5]Menon and Pancheva are not explicit about what an instance of a property is, but it is clear that they do not intend by this an individual instantiating the property. Rather, the notion of an instance of a property they intend seems to us to be something like the instantiation of a property in an individual, a notion similar to Moltmann's *trope* (e.g. Moltmann 2009)

(16) $[\![\text{POS}]\!] = \lambda g_{<d,<e,t>>} \lambda x \exists d[g(d)(x)\ \&\ d > d_s]$

This null morpheme takes a degree predicate (the denotation of e.g., the verbalized property concept root), and creates a predicate of individuals, true of an individual *a* iff the degree above which *a*'s instance of the root property measures is higher than some contextually given standard (d_s in (16)).

In schematic fashion, then, their treatment of a property concept word like *nalla* 'good' in the context of a sentence like that in (15) is as in (17). It is first verbalized by the phonologically null verbalizer, which introduces the possessive semantics required in order to relate a property to individuals. The null POS saturates the degree argument, and introduces a context-sensitive degree of comparison, as described above. Finally, relativizing morphology is affixed, for reasons which, as described above, are not well understood, but are assumed (by all) to be morphosyntactic in nature, and have no impact on the lexical semantics relevant for our considerations. This gives to *nallavan* 'good-M.SG', a denotation like that in (17b).

(17) a. $[[[\sqrt{nall} + \emptyset_{v-poss}]_v + POS]_v$–*a-van*$]_{rel}$
b. $\lambda x \exists d \exists y[y$ is an instance of goodness & x has y & $\mu(y) \geq d$ & $d > d_s]$

4.2 Class 2 roots

Class 2 roots are turned into nouns by a nominalizing functional head realized as the *–am* suffix appearing on such roots. On Menon and Pancheva's analysis, this categorizing *n* can have different meanings, one of which is a function that takes a property and returns a relation between degrees and "instances" of the property, as in (18).

(18) $[\![\text{–am}]\!] = \lambda \Pi \lambda d \lambda x[x$ is an instance of Π & $\mu(x) \geq d]$

This suffix, unlike the verbalizing morphology above, does not relate the property denoted by the root to the individuals possessing it; this has to come from some where else (overt possession, as discussed below). This nominalizer *–am* does introduce a degree argument, which as above, needs to be saturated. In comparative constructions, this is done by the comparative morphology, and in positive constructions like (19), it is accomplished in the same manner as with the Class 1 property concept sentences, through composition with phonologically null positive degree morphology.

(19) *avaɭkkə pokkam uɳṭə*
she.DAT tallness EX.COP
'She is tall.' (Menon & Pancheva 2014, 294)

Schematically, this gives the structure in (20a) to a property concept word like *pokkam* 'tall'. Its denotation in a positive context like (19), with a saturated degree argument, is

as in (20b). It denotes a set of instances of tallness whose measure is greater than some contextually given standard.

(20) a. $[[[\sqrt{pokk} + am_n]_n + POS]$
b. $\lambda x \exists d[x \text{ is an instance of tallness } \& \ \mu(x) \geq d \ \& \ d > d_s]$

Crucially, (20b) is not a predicate of ordinary individuals, and therefore cannot compose directly with an ordinary individual in order to express the translational equivalent of a property concept sentence. This is why, on Menon and Pancheva's analysis, possessive morphosyntax surfaces with such lexemes in property concept sentences. The overt possessive morphosyntax seen in (19) introduces the semantics that was introduced by the verbalizer in the case of class 1 roots.

5. Malayalam and the syntactic variationist view

Descriptively, as discussed in §3, Malayalam shows two strategies of property concept predication. Class 1 roots give rise to canonical predication, as illustrated in (21), while Class 2 roots give rise to possessive property concept sentences, as in (22).

(21) *avaɭ nalla-vaɭ aaɳə*
she good-F.SG EQ-COP
'She is good.'

(22) *avaɭkkə pokkam uɳʈə*
she.DAT tallness EX.COP
'She is tall.'

On Menon and Pancheva's analysis, however, this descriptive generalization is an illusion created by morpho-phonological accidents. At a deeper level, there is in fact no difference in the mode of predication observed with the two classes. Rather, all roots give rise to possessive strategies of predication. The possession is simply covert with the Class 1 roots, since the morpheme contributing possessive morphology is phonologically null (i.e., $v - poss$). With Class 2 roots, the possession is observed on the surface, in the form of the existential copula+dative construction. Furthermore, their assumption is that this kind of contrast governs the surface variation between predicative and possessive property concept sentences not only in Malayalam, but universally (Menon and Pancheva, 301). Specifically, they claim that "property concepts universally lexicalize as roots, and they denote substance-like individuals, requiring possessive predication." Crosslinguistic variation arises as a consequence of (i) morphophonological accidents, such as that observed internal to Malayalam, where a possessive *v* is phonologically null, and (ii) differences in the inventory of functional heads. Generally, property concept sentences are predicative when possessive semantics is introduced by a null morpheme, and overt when it is introduced by an overt one. Adjectives, in languages that have them, are assumed on this analysis to be "syntactically derived categories that too use a possessive strategy of predication, a

covert one." The adjectivizing categorizer of roots in languages like English, presumably a phonologically null head *a*, on such an analysis, introduces possessive semantics. In short, on such a theory, as Menon and Pancheva say, the observed variation in the morphosyntactic form of property concept sentences, both language internally and crosslinguistically, is *not* a consequence of the nature of the inventory of roots, but rather a consequence of the nature of the inventory of functional heads and their realization. In the sections that follow, we discuss a range of undesireable consequences that this view has.

6. Problematic predictions of the syntactic analysis

The intuition underlying Menon and Pancheva's analysis is that there is a universal lexical semantics for property concept lexemes and that the variation in the form of property concept sentences is a consequence of morpho-phonological accidents and variation in the inventory of functional heads. We see two kinds of problems with such an analysis. The first kind has to do specifically with overgeneralization in Malayalam. The second kind is more general, and concerns problems that arise when the crosslinguistic implications of an analysis pinning variation on language specific inventories of functional morphemes are considered in more detail. We treat these in turn.

6.1 Problems specific to Malayalam

Internal to Malayalam, Menon and Pancheva's syntactic variationist analysis runs into problems which a semantic variationist analysis does not encounter. The problems include at least two kinds of missed generalization.

The first concerns the motivation for the functional morphology appealed to by the analysis. While it is debatable whether the nominal categorizing morphology *–am* should be treated syntactically, as it is in Menon and Pancheva's analysis, there is no doubt that *–am* is a nominalizer, i.e. a suffix that makes a noun out of a bound root. In the case of the hypothesized *v* head that categorizes Class 1 roots, however, there is reason for serious doubt, since it is never realized phonologically, and Menon and Pancheva offer no other empirical arguments for its existence.[6] The arguments, rather, are purely theory-internal - the semantically uniform treatment of property concept lexemes, and the broader Distributed Morphology assumptions which entail that all word-formation is syntactic, cannot be maintained without it.[7] A more plausible analysis, we claim, would treat the Class 1 roots not as bound roots at all, but simply lexically as (individual-characterizing) verbs. Such an analysis makes moot the question why there is no evidence for a phonologically null functional head.[8]

[6] See Menon & Pancheva (To appear), however, for a possible argument from color terms.

[7] A Distributed Morphology adherent might argue that there is independent motivation elsewhere for the view that word formation is always syntactic, and that positing a *v* head is justified on those grounds. See Baker (2003, Chapter 5) for a sound rebuttal of this view.

[8] Both analyses still need to account for the fact that class 1 roots/verbs are restricted in distribution and only occur with the relativizer *–a*. One might conjecture that this is because the relevant verbs are native Dravidian verbs that have become obsolete, and 'fossilized' in *–a* forms. But such a view clearly needs to be argued for, and we are not in a position to do so here.

The second problem is that the proposed syntactic view does not in fact make the two classes of roots natural classes. Specifically, nothing in the analysis laid out by Menon and Pancheva blocks any root from occurring with any head, whether *n* or *v*. To state the concern differently, the analysis, as stated, actually predicts that all roots should appear in both canonical and possessive property concept sentences. This is because both functional heads take property denoting bound roots as arguments, and both Class 1 and Class 2 roots are roots of precisely that type on this analysis. In fact, however, the roots are restricted in distribution—Class 1 roots only appear in (overtly) canonical predicating property concept sentences, while Class 2 roots only appear in (overtly) possessive property concept sentences.

A proponent of Distributed Morphology might argue that this deficiency could be overcome by appealing to something like the notion of allosemy elaborated recently in Wood (2015) in the context of DM. The assumption would then be that all roots combine with all heads, but that the combination of Class 2 roots with *n* is not assigned any denotation at LF, and similarly for the combination of Class 1 roots with *v*. This line of argument seems to us to simply recreate the problem elsewhere, as nothing explains why it is that, systematically, whenever the combination of a root with *v* is blocked at LF, the combination of that root with *n* is not, and vice versa. Furthermore, the fact remains that the combination of class 1 roots with e.g. the nominalizer *–am* do not exist in Malayalam. It seems to us that the proponent of Distribute Morphology would have to elaborate a theory that allows discrimination between head-root selection and LF-blocking, with applicable diagnostics to tell the two apart. Finally, there is at least one other reason to believe that the two classes of roots in Malayalam are in fact natural classes, namely the fact that Class 1 roots are native Dravidian roots, whereas Class 2 ones are generally borrowings.

While we believe that these two Malayalam-internal considerations already call the syntactic variationist view into question, they are of relatively minor concern compared to the problems such a view faces when generalized crosslinguistically.

6.2 Crosslinguistic problems

As discussed above, a basic assumption of Menon and Pancheva's syntactic analysis is that property concept lexemes are universally precategorial, property-denoting roots. Variation in the form of property concept sentences across languages is a consequence of differences across languages in (i) the inventory of categorizing functional heads, e.g., whether the particular categorizing heads a language has have a possessive semantics or not, and (ii) whether syntactic material introducing possessive semantics is phonologically realized or not. This theory gives rise to three problematic crosslinguistic predictions.

First, since verbalizing morphology introduces the possessive relation in property concept sentences that are morphosyntactically predicational on the surface, it follows that there should be a general link between categorizing morphology and possessive semantics. For example, Menon and Pancheva's hypothesis about adjectives in languages like English, where property concept sentences are canonically predicating with an adjectivally categorized property concept root, is precisely that the adjectivizer head carries possessive semantics. We should therefore expect to see a crosslinguistically robust coincidence of cat-

egorizing morphology with possessive morphology. Such coincidence arguably occurs for nominalizers in Ulwa. In that language, it is plausibly the case that the morpheme *–ka* that occurs on all property concept roots is at once a nominalizer and a possessive morpheme, as argued in Koontz-Garboden & Francez (2010) and Francez & Koontz-Garboden (2015). So far as we are aware, however, the Ulwa pattern is very much the exception rather than the norm, Huave being the only other language we are aware of in which this pattern of possessive morphosyntax is used, and only marginally at that (see Kim & Koontz-Garboden 2013 for details). The rarity of such a coincidence argues against an association of nominalizers with possessive morphology. For other categorizers, the situation is much worse, and no coincidence is ever observed with possessive morphology. For example, so far as we are aware, no language has adjectivizing morphology on property concept roots that is also possessive, despite the fact that many languages that have adjectives have overt possessive morphemes. Nor are we aware of any language that displays an observable coincidence of verbalizing morphology and possessive morphology. This casts serious doubt on the syntactic view, since, while any syntactic head might well be phonologically null in one language, it is highly implausible for it to be universally null.

Second, if property concept lexemes universally denote properties, and if all categorization is syntactic, then we expect to see categorizing morphology systematically diverge between categorizers of property concept roots and categorizers of other roots in the same syntactic categories. This is on the reasonable assumption that not *all* roots denote properties, i.e. that the roots of many verbs and nouns (like *eat* or *dog*) are not property denoting and do not call for a semantics of possession in combination with arguments. For example, we would expect a crosslinguistically recurring distinction between property concept verbs and other verbs, as well as between property concept nouns and other nouns, in terms of the categorizing morphology used in word formation with them. At the very least, we expect this in languages in which categorizing morphology is overt. While seriously corroborating or disproving this prediction requires a systematic crosslinguistic investigation, which we have not carried out, we are skeptical that it is borne out.

A third false prediction concerns the syntactic categories of property concept words that overt possessive morphosyntax is found with. Absent additional development of the theory, Menon and Pancheva's view has it that categorizers are found in both possessive and non-possessive guises. So, for example, *n* in Malayalam, realized by *–am*, lacks possessive semantics, while in Ulwa, the *n* realized by *–ka* does introduce possessive semantics. While this may be unproblematic in the domain of nominalizers, we are skeptical, and certainly not aware of any evidence, that there is a *v* that combines with property concept roots and does not carry possessive semantics. This would mean that there are verbal property concept words that, in order to form property concept sentences, require combination with external possessive morphosyntax. Worse, as we discuss in Francez & Koontz-Garboden (In prep), it is clear that there are no languages in which adjectival property concept words occur in possessive sentences to express translational equivalents of property concept sentences. What this means in the context of the syntactic view of variation is that the adjectivizing head *always* has possessive semantics. Perhaps there is some explanation, in the context of the syntactic view, for why this might be the case, but it is unclear to us what this explanation might be. What is clear, however, is that the theory as currently stated falsely

predicts that overt possessive morphosyntax should, across languages, be found with all syntactic categories of property concept lexemes.

The conclusion of this discussion is that the semantics of predicative property concept sentences cannot be reduced to that of compositionally constructed, possessive property concept sentences. Property concept lexemes do not have a universal lexical semantics. A theory that maintains that they do leads to overgeneralization in the context of Malayalam, and, more broadly, to a series of crosslinguistic predictions that are not (or, in some cases, at least do not seem to be) borne out.

7. A semantic variationist analysis of Malayalam

If the syntactic approach is undesireable, as we have argued, crosslinguistically and for Malayalam, then we need an alternative analysis of the facts. We believe that the right analysis, both of the crosslinguistic variation and of language internal variation as observed in Malayalam, is the one sketched out above, and laid out in detail in Francez and Koontz-Garboden (2015), namely that variation in the form of property concept sentences is tied to variation in the lexical semantics of property concept lexemes.

Menon and Pancheva's Class 1 and Class 2 differ in exactly the way the semantic variationist approach predicts. The former are individual characterizing, while the latter are substance-characterizing. Morphosyntactically, we assume, by contrast with Menon and Pancheva, that only Class 2 lexemes are bound roots, and must be categorized before being used as words. This is the job of the nominal morphology *–am*. The fact that *–am* appears with nouns of all different kinds, as shown above in §3.2, suggests to us that it is semantically inert, and plays only a morphosyntactic role in word formation, forming nouns from bound roots. Semantically, it is inert, and returns the meaning of the root it combines with. Unlike Menon and Pancheva's analysis, then, the semantic variation analysis does not require multiple denotations for *–am*, and derives the meanings of *–am* nouns simply from their roots. Since property concept roots are substance-characterizing, so too are Class 2 nouns derived from them.[9] Class 1 lexemes, we argue by contrast, are simply lexical verbs (albeit defective ones, see fn. 6) with an individual-characterizing denotation. Given this kind of approach, we predict that property concept sentences formed with Class 2 lexemes will be possessive, while those formed with Class 1 will be canonical-predicating, as is the case.

This analysis is advantageous because it accounts for Malayalam using the same assumptions that account for other languages. It does not invoke syntactic structure for which there doesn't seem to be structural evidence. Further, it does not overgenerate. Since Class 1 and Class 2 lexemes differ in their semantics as well as in their syntactic categories, the prediction is that there should be no crossover in classes. We should not find Class 1 lexemes in possessive property concept sentences or Class 2 lexemes in canonical predicating

[9]Menon and Pancheva have *–am* introduce a degree semantics with their property-denoting roots. We do not need a degree semantics, since as discussed in Francez and Koontz-Garboden (2015), on this analysis gradability already comes from the pre-order on substances. Menon and Pancheva need the degree-based analysis that they propose because they adopt a property-theoretic approach to the semantics of possessive-predicating property concept lexemes.

property concept sentences. This prediction is borne out, as discussed above, and contrasts with the predictions of the syntactic variationist analysis. Most importantly, the semantic variationist analysis makes the morphosyntactic variation in the form of property concept sentences systematic. According to this view, we find possession if and only if the property concept lexeme has a substance denotation. This contrasts with the arbitrary nature of the variation on the syntactic variationist analysis proposed by Menon and Pancheva, in which the the variation is a matter of language specific morphological accident, with possessive sentences arising when there happen to be phonologically null possessive categorizers.

8. Concluding remarks

To conclude, Malayalam presents the same kind of pattern as has been observed within and across other languages. Some property concept sentences are possessive, others are predicative. The question is what is responsible for this pattern—variation in the semantics of property concept lexemes, or variation in the inventory and realization of functional heads. We have argued that the distribution of possessive and predicative form is a consequence of variation in lexical semantics. Possessive sentences arise with substance-characterizing property concept lexemes, and canonical predicative ones arise with individual-characterizing ones.

In contrast, on the syntactic view, at least as articulated in Menon and Pancheva's analysis of Malayalam, the distribution of forms is an accident of morphophonology. Property concept lexemes are universally property-denoting and the semantics of property concept sentences is always possessive, but this is not always reflected in surface form. Specifically, possessive *v* in Malayalam happens to be null, while *n* is overt, but happens not to introduce the semantics of possession.

We have argued that the semantic variationist view is advantageous for several reasons. First, it aligns the Malayalam facts with a general crosslinguistic pattern. Second, it does not invoke syntactic structure for which there doesn't seem to be structural evidence, and does not overgenerate, making apparent natural classes natural. Finally, it avoids a range of problematic crosslingusitic predictions. We conclude that this is a clear case where lexical semantics explains variation in morphosyntactic form. Indirectly, we also view these results as evidence that the lexicon is an important component of grammar, in which various grammatically relevant generalizations must be stated.

References

Asher, Ronald E, & TC Kumari. 1997. Malayalam. *New York, NY: Routledge* .

Baker, Mark C. 2003. *Lexical categories: Verbs, nouns, and adjectives*. Cambridge: Cambridge University Press.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6:339–405.

Chierchia, Gennaro, & Raymond. Turner. 1988. Semantics and property theory. *Linguistics and Philosophy* 11:261–302.

Dixon, R.M.W. 1982. *Where have all the adjectives gone?: And other essays in semantics and syntax*. The Hague: Mouton.

Francez, Itamar, & Andrew Koontz-Garboden. 2015. Semantic variation and the grammar of property concepts. *Language* 93:533–563.

Francez, Itamar, & Andrew Koontz-Garboden. 2016. A note on possession and mereology in Ulwa property concept constructions. *Natural Language and Linguistic Theory* 34.1:93-106.

Francez, Itamar, & Andrew Koontz-Garboden. In prep. *Lexicalization and morphosyntactic variation: Adjectival and nominal property concepts*. Oxford: Oxford University Press.

Keenan, Edward. 1973. Logic and language. *Daedalus* 102:185–194.

Kim, Yuni, & Andrew Koontz-Garboden. 2013. The lexical categories of Huave property concept words. University of Manchester, ms.

Koontz-Garboden, Andrew, & Itamar Francez. 2010. Possessed properties in Ulwa. *Natural Language Semantics* 18:197–240.

Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In *Meaning, use, and interpretation of language*, ed. R. Bäuerle, Chr. Schwarze, & A. von Stechow, 302–323. Berlin: de Gruyter.

Matthewson, Lisa. 2001. Quantification and the nature of crosslinguistic variation. *Natural Language Semantics* 145–189.

Menon, Mythili, & Roumyana Pancheva. 2014. The grammatical life of property concept roots in Malayalam. *Proceedings of Sinn und Bedeutung 18*. Pp. 289–302.

Menon, Mythili, & Roumyana Pancheva. To appear. Decomposing color expressions in malayalam. In *Proceedings of FASAL V*, ed. S. Sundaresan & R. Balusu.

Moltmann, Friederike. 2009. Degree structure as trope structure: A trope-based analysis of positive and comparative adjectives. *Linguistics and Philosophy* 32:51–94.

Talmy, Leonard. 1985. Lexicalization patterns. In *Language typology and syntactic description*, ed. Timothy Shopen, volume 3, 57–149. Cambridge: Cambridge University Press.

Thompson, Sandra A. 1989. A discourse approach to the cross-linguistic category'adjective'. In *Linguistic categorization*, ed. Roberta Corrigan, Fred Eckman, & Michael Noonan, 245–265. Amsterdam: John Benjamins.

Wood, Jim. 2015. *Icelandic morphosyntax and argument structure*. Dordrecht: Springer.

Itamar Francez
ifrancez@uchicago.edu

Andrew Koontz-Garboden
andrewkg@manchester.ac.uk

Expression of adjectival meaning in Kannada[1]

Sindhu Herur

The English and Foreign Languages University (EFLU), Hyderabad, India.

1. Introduction

This paper explores the expression of adjectival meaning in Kannada. Adjectival meaning in Kannada is expressed through nouns, an idea established in Amritavalli & Jayaseelan (2003). Employing the semantic notion of *Property Concept* (PC) (Dixon 1982), to evaluate relevant data, the study proposes, in addition to nouns, a small, closed-class of functional adjectives in the language. The predication structures required by PCs[2] in Kannada work to further strengthen the analysis.

A *property concept term* or a *property concept lexeme*, in the sense of Francez and Koontz-Garboden (2015), is a morphologically simple lexical item which maybe lexicalised as an adjective in languages that have a lexical category of adjectives. In English, a language that has a lexical category of adjectives, the property of *height* is expressed by the adjective *tall*. This is the sense in which the term *property concept* was stated in Dixon (1982). However, in a language like Kannada wherein nouns are used to express *property concepts,* I apply this term to nouns that express a property such as *height, weight, anger* etc. Such terms in Kannada are hence PC nouns. In this paper, as stated earlier, I put forward the proposal for a small, closed-class of adjectives which I shall refer to as PC adjectives. Thus a PC lexeme maybe lexicalised as an adjective, a noun or a verb across languages to express a property.

[1] I would like to thank my supervisor, Prof .R. Amritavalli and my advisory committee head Dr.Rahul Balusu. All errors remain mine.
Kannada in this paper refers to the standard Kannada of Bengaluru-Mysuru.
[2] In this paper, I use the notion of Property Concept (PC) in its broad sense to refer to lexical items, be it a noun or an adjective, which express adjectival meaning in Kannada.

Both these aspects, PCs and the realisation of these PCs in attributive and predicative contexts, lend to a broad consensus in understanding the expression of adjectival meaning in Dravidian in recent work — Balusu (2015) for Telugu and Menon (2013) and Menon & Pancheva (2014) for Malayalam.

We know that adjectives are analysed as expressions which denote relations between objects or individuals and abstract measures or degrees according to the *scalar analysis* in Kennedy (1997). The question then arises as to how PC nouns in Kannada in particular and Dravidian in general, expressing properties such as *anger, happiness, sadness, hunger*, etc. achieve truth conditions? In other words, how are PC nouns related to the individual they are attributive of? In this context, the semantics of possession of these *properties* comes into question as elaborated in Francez and Koontz-Garboden (2010; 2015) wherein certain languages like Ulwa are described to use genitive case to utilise PC nouns attributively and predicatively. In Kannada we shall see that PC nouns employ both the genitive and the dative case to encode possession in attributive and predicative contexts respectively.

Interestingly, the small, closed-class of adjectives in Kannada are the same in Malayalam and Telugu. By this, I mean that the same group of PCs exhibiting the same characteristics are found in Kannada, Malayalam, Telugu and Tamil. However, for Telugu, Balusu (2015) does not make a mention of this group and Menon and Pancheva (2014) have chosen to treat them as roots which take the help of functional architecture to be realised as PCs. My analysis in this paper does not take a pre-lexical perspective as in Menon and Pancheva (2014) but a lexical one and I treat them as a closed class of functional words which exhibit all the characteristics of an adjective.

In addition, Kannada, unlike its sister languages, has an overt instantiation of PC nouns like *udda* 'height' that take dative case, as in *udda-kke* 'height-DAT', in predication. I analyse *udda-kke* and its group members as adjectives. This supports Amritavalli and Jayaseelan's (2003) claim (which we shall delve into further along) that universally nouns incorporate into dative case to be realised as adjectives. The Kannada data provides overt attestation for such a claim.

The outline of the paper is as follows. Section 1 was an introduction to this paper, placing this study in relevance to the current literature on adjectival expression in Dravidian languages. Section 2 is the main body of this paper. Section 2 begins with an account of the background work done to arrive at the descriptive facts of the data and is divided into two subsections, sub-section 2.1 and 2.2, dealing with PCs in Kannada in the attributive position and the predicative position respectively. Sub-section 2.2 dealing with PCs in the predicative position has two further sub-sections. Sub-section 2.2.1 discusses PC adjectives in the predicative position and sub-section 2.2.2 discusses PC nouns in the predicative position. A careful consideration of PC nouns in the predicative position calls for a re-formulation in the two simple groups of PCs in Kannada motivated thus far, namely, the functional closed-class of adjectives and nouns. Section 3 concludes the paper.

2. Property Concepts in Kannada

The background work to this paper is as follows. Working with a data set of fifty PCs in Kannada, tests of attribution and case marking provided the initial classification of PCs into PC adjectives and PC nouns in Kannada. I shall elaborate on these tests shortly. *Oḷḷeya* 'good' exemplifies the group of PC adjectives and henceforth in this paper I refer to this group as the *oḷḷeya* group. The predication structures of PC nouns reveal a further divide based on whether these nouns occur in only dative subject constructions or in dative as well as nominative subject constructions. A majority of the PC nouns behave like 'true-blooded' nouns and occur only in dative subject constructions. *Koopa* 'anger' represents this group of PC nouns and hence we shall refer to this

group as the *koopa* group. A small group of PC nouns however, occur predicatively in nominative as well as dative subject constructions. In nominative subject constructions, these PCs take dative case and are realised as adjectives. *Udda* 'height' represents this group of PC nouns and we shall call this group, the *udda-kke* 'height-DAT' group. I have summed up here, a three-way classification of PCs that the subsequent sub-sections will elaborate. The *oLLeya* group of PC adjectives and the *koopa* group of PC nouns indicating two ends of the spectrum of PCs in Kannada and the *udda(kke)* group of PCs displaying the properties of both these groups. The dative case hence, plays a crucial role in PC noun predication in Kannada; in the sense of whether it is the subject that is case-marked or the PC noun which is case-marked.

2.1 Property Concepts in the attributive position

My first step was to identify if Kannada has adjectives at all. I employ a two-fold test in order to identify PC adjectives in Kannada. The first part of the test is to ascertain whether they can occur in the attributive position of NPs/DPs. The second part of the test is that the PCs should not take case. A handful of PC adjectives, were identified and are listed below in (1).

(1) Members of the ***oḷḷeya* group** of PC adjectives in Kannada- *oḷḷeya* 'good', *keṭṭa* 'bad', *cikka* 'small', *doḍḍa* 'big', *hosa* 'new', *haḷeya* 'old', *yeḷeya* 'tender' and *baḍa* 'poor'

However as the reader may observe in the list in (1) above, it appears difficult to determine a semantic basis for this group of functional lexemes. Below in (2) is the data for one of these adjectives, *oḷḷeya*, being representative of its class, for the two mentioned tests. In (2a) we can see that *oḷḷeya* occurs attributively in the prenominal position and in (2b) we can see that *oḷḷeya* cannot take case.

(2) a. oḷḷeya huḍuga / oḷḷeya huḍugi / oḷḷeya huḍugaru / oḷḷeya huḍugiyaru / oḷḷeya maatu-gaḷu

good boy / good girl / good boys / good girls / good word-PL

'good boy / good girl / good boys / good girls / good words '

b. *oḷḷeya-da / *oḷḷeya-vannu / *oḷḷeya-kke / *oḷḷeya-dalli / *oḷḷeya-dinda

good-GEN / good-ACC / good-DAT / good-LOC / good-ABL

Comparing this to a PC noun in Kannada like *koopa* '*anger*', we see that the opposite holds true. *Koopa* cannot occur attributively without the help of genitive case (as seen in 3a-b); and in (3c) I illustrate the fact that *koopa* 'anger' can combine with various cases in Kannada as any ordinary noun would.

(3) a. *koopa maatu-ga[-u

anger word-PL-NOM

'*anger words' (intended: 'angry words')

b. koopa-**da** maatu-ga[-u

anger-GEN word-PL-NOM

'words of anger' (lit. anger's words)

c. koopa-da / koopa-vannu / koopa-kke / koopa-dalli / koopa-dinda

anger-GEN / anger-ACC / anger-DAT / anger-LOC / anger-ABL

We have seen that *koopa*, a noun, can modify another noun attributively with the help of genitive case. We can liken the use of genitive case in the example above in (3b) to the genitive PP *of* in English. In a language like English too, a noun cannot directly modify another noun without the use of *of* (with the exception of compound nouns). This genitive PP signifies possessive semantics as illustrated by these examples- *the land of the rising sun (ref. the rising sun's land), the home of the deceased (ref. the deceased's home)* and *the music of the soul (ref. the soul's music)*.

Other PC nouns like *koopa* are listed in (4) and the reader may note that the list includes Sanskrit borrowings as well as native Dravidian nouns.

(4) Some members of the ***koopa* group** representing PC nouns in Kannada- *koopa* 'anger', *santosha* 'happiness', *sukha* 'contentment', *dukha* 'sadness', *bhaara* 'heaviness', *bhaya* 'fear', *hagura* 'lightness', *sulabha* 'ease', *kaʃta* 'difficulty', *teḷuvu* 'thinness', *kobbu* 'arrogance' and *beesara* 'sadness/sulk'

The reader may feel that both the *oḷḷeya* group of PC adjectives and the *koopa* group of PC nouns seem to end with *–a*. Phonologically, Kannada has an open syllable coda rule. Any lexeme, be it a functional or a lexical one, can never end with a consonant but have to always end in a vowel. As seen in the PC nouns listed in (4) above, lexemes end in *–a, -i, -e* or *–u*. It would therefore not be prudent to do a morphological analysis of all PCs in Kannada based on the *–a* in their coda position. Having said this, all PC adjectives in Kannada, without an exception, end in *–a* (as seen in 1). The genitive case marker in Kannada is also an *–a*. We have seen that PC adjectives resist genitive case marking (as seen in 2b) - *oḷḷeya* 'good' cannot become *oḷḷeya-da* 'good-GEN'. However, we shall discuss a little later on that *–a* is a very versatile suffix in Kannada and is a case in syncretism. The *–a* is a marker of genitive case, a relative clause and a yes-no question.

The *koopa* group of PC nouns do not resist genitive case. Hence, the *–a* in *koop-a* cannot be genitive case but is the effect of the phonological rule of Kannada of a lexeme ending in an open syllable. *Koopa* and other PC nouns ending in *–a* take genitive case with a phonological consonant epenthesis rule being applied- *koopa* 'anger' becomes *koopa-d-a* 'anger-GEN'. There is a clear difference in behaviour between the two groups of PCs described so far. Thus a straight-forward *–a* ending morphological test of PCs in Kannada will lead us nowhere.

With these tests of attribution and case-marking, we have an initial classification of PCs in Kannada into PC adjectives and PC nouns. PC adjectives are represented by *oḷḷeya* 'good' and PC nouns are represented by *koopa* 'anger'. The former does not combine with various case morphemes in Kannada while the latter does. Furthermore, we have noted that the former group, i.e., PC adjectives, can occur in their 'bare' form in attributive positions of NPs while the latter group, i.e., the PC nouns, require genitive case to attributively modify a noun.

2.2 Property Concepts in the predicative position

PC adjectives in Kannada occur predicatively in verbless copular clauses of the kind illustrated below in (5). We shall discuss this below in 2.2.1.

(5) avaḷ-u oḷḷeya-avaḷu.

she-NOM good-she

'She is a good person.'

PC nouns on the other hand cannot occur in such verbless copular clauses but require the copula to be overt. As illustrated below and mentioned earlier, the members of the *koopa* group of PC nouns occur only in dative subject constructions as seen in (6a-b). The *udda* group of PC nouns can occur in both nominative subject and dative subject constructions as seen in (7a-b). This latter point has been noted along with the corresponding data I have presented (7a-b), in Amritavalli and Jayaseelan (2003; ex. 21 and 22).

(6a) avan-**ige** **koopa** id-e. / (6b) *avan-**u** **koopa-kke** idd-aane.

he-DAT anger be-3.N.SG / he-NOM anger-DAT be-3.M.SG

'He has anger.' / 'intended: He is angry.'

(7a) avan-**u** **udda-kke** idd-aane. / (7b) avan-**ige** **udda** id-e.

he-NOM height-DAT be-3.M.SG / he-DAT height be-3.N.SG

'He is tall.' / 'He has (the) height.'

In Kannada PC nouns, as we shall see shortly, the classification is between PC nouns which can occur predicatively only in dative subject constructions vis a vis PC nouns that can occur in both the nominative subject construction (with the PC noun itself marked for dative case) and the dative subject construction. We shall also see that these two groups of PC nouns have a semantic basis for classification. Thus, with PC nouns in Kannada, the dative case is always present predicatively, either on the subject or on the PC noun.

2.2.1 Predication of Property Concept adjectives

Beginning with PC adjectives in Kannada, I have stated that they occur predicatively in verbless copular clauses. Let us examine such a clause. A sentence like '*John is a doctor*' in Kannada has only an NP NP structure as illustrated in example (8). Such constructions are noted in Amritavalli (2000) as 'verbless clauses' of the form NP NP.

(8) John doctoru.

John doctor

'John is (a/the) doctor.'

Oḷḷeya 'good' occurs predicatively in such a clause structure with a suffixal morpheme encoding what appears to be 'agreement' for number and gender. The invariant value for person is 3rd person. As the data demonstrates, these suffixal morphemes are homophonous with the pronominals in Kannada- *avanu* 'he', *avaLu* 'she' and *avaru* 'they'.

(9) avanu/huḍuga **oḷḷeya-avanu.** / avaḷu/huḍugi **oḷḷeya-avaḷu.** / avaru **oḷḷeya-avaru.**

he/boy good-he / she/girl good-she / they good-they

'He/The boy is a good person/boy.' / 'She/The girl is a good person/girl.' / 'They are good people.'

The adjective *oḷḷeya* 'good' cannot occur predicatively in its 'bare' form.

(10) *avanu/huḍuga oḷḷeya.

he/boy good

'Intended: He/ The boy is good.'

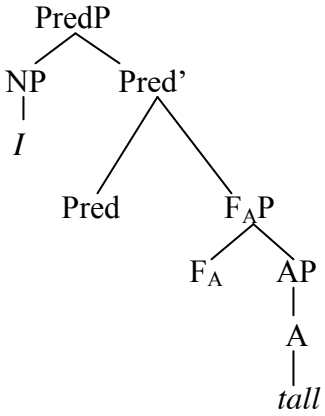
The data in (9) and (10) above, leads us to two ways of analysing the suffixal morpheme on the *oḷḷeya* group of functional adjectives in their predicative position. The *avanu, avaḷu and avaru* morphemes that combine with the adjective can be analysed as below, either as adjectival agreement (as in 11a) or a pronominal (as in 11b). At the phrasal level, the two analyses hence predict different complements, either an AP or an NP complement.

(11a) [[$_{NP}$ avanu] [$_{AP}$ [$_{AgrP}$ avanu] [$_{AP}$ oḷḷeya]]] → [NP AP] clause structure

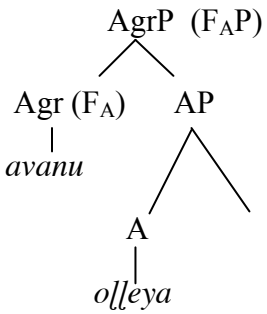
(11b) [[$_{NP}$ avanu] [$_{NP}$ [$_{FP}$ oḷḷeya] [$_{NP}$ avanu]]] → [NP NP] clause structure

According to the first analysis as illustrated in (11a), the data in (9) would have an NP AP clause structure, wherein these morphemes would be analysed as some sort of adjectival agreement in the predicative position (though they do not show agreement in the attributive position). Adopting Baker's (2008) schema as illustrated in (12) for predicate adjectives, we could represent the *adjectival agreement approach* as in (13) [3] - the AgrP headed by the *avanu* morpheme would be a functional layer in the domain of the A.

(12) [structure for 'I am tall'; Baker 2008; pg 37; ex 58b]



(13)



[3] Baker (2008) when discussing Kannada data from Sridhar (1990) suggests treating phrases like *oLLeya-avanu* as predicate nominal constructions rather than as predicate adjectival constructions with agreement. Thus, (13) is not a representation Baker uses to explain the Kannada data. However, for the purpose of this discussion, I adopt Baker's (2008) schematic representation of a predicative adjective.

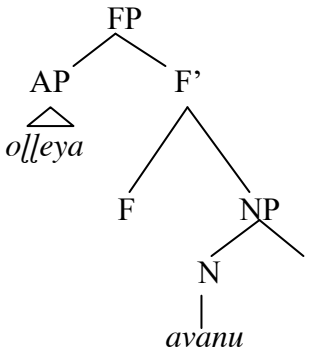
When we consider Baker's (2008) detailed cross-linguistic analysis of the distribution and behaviour of adjectives in the attributive and the predicative positions, *oḷḷeya* in *avanu oḷḷeya-avanu* 'He is a good person (lit. *he good-he*)' can be analysed as an adjective that attributively shows no agreement but predicatively agrees for number and gender. In such a case, the subject NP has interpretable features of person, number and gender while the *avanu* in '*oḷḷeya-avanu*' has uninterpretable features that check for number and gender only (consistent with Baker's 2008 theory of adjective agreement). However, Baker (2008: 61-62) argues that adjectives that don't agree attributively cannot possibly show agreement predicatively. He brings up the Kannada data (ref. Sridhar 1990:249-50) as a possible counterexample to his proposal, but goes on to refute such a possibility. Baker then proposes the solution that *oḷḷeya-avanu* is not a predicate adjective construction but a predicate nominal construction (this is also Sridhar's view). Then the "agreement" is not by Agree (Chomsky 2000). Baker (2008:10-11) would explain it as semantic agreement: the subject NP and the object NP are inserted at random and then their semantic features are matched. If there is a match the utterance is realised. Hence, the features on both the nouns are interpretable. This is similar to the *accidental co-reference* of pronominal coreference Lasnik (1976). Sentence (14) is an example of semantic agreement.

(14) He is an actor/ *He is an actress.

If we adopt Baker's (2008) explanation of this clause structure in Kannada, we run into some problems. The semantic agreement solution has not been elaborated upon in his work. Even if we take the semantic agreement as a possible solution, there is sufficient proof that only the subject pronominal is referential and the suffixal pronominal can never be referential. Thus a featuring matching explanation between the two NPs will not be possible. On the other hand, treating the predicate as an AP which shows agreement, does not seem to hold water. Firstly, it would be problematic to propose a separate clause structure for a small, closed-class of functional lexemes. Secondly, as Baker (2008) notes, adjectives that do not show agreement in the attributive position cannot 'agree' predicatively.

Let us turn now to the second analysis as illustrated in (11b) which would predict an NP NP clause structure wherein *oḷḷeya* 'good' would be treated as a 'nominalised adjective'. Adopting Cinque's (2010) framework, we could represent it as follows. The AP *oḷḷeya* is in the [Spec, FP] position where FP represents a functional projection. The F head chooses an NP as its complement headed by the pronominal *avanu*.

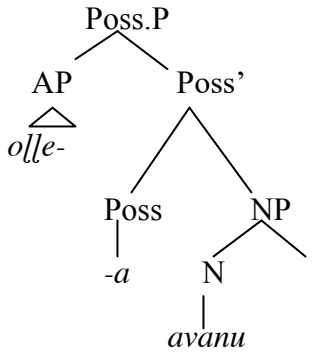
(15)



Under this approach then *oḷḷeya* maybe realised as *oḷḷe-* which is a root, which attaches to *-a*, the genitive case marker in Kannada. The genitive case is the head of the functional layer which aids the root *oḷḷe-* to enter the

derivation process. As explained in (17) below, *oḷḷe+ya* can now modify its nominal complement. At the moment, I am not certain whether I would have to treat the root as a nominal root or a bare root and I keep this question aside for now. We can now re-analyse the representation in (15) above to capture this idea. We have to thus motivate a Possessive Phrase (Poss.P), in the re-representation of (15) in (16) below, whose head is the genitive/possessive case marker *–a* and whose specifier position hosts the root *oLLe-*. The specifier-head combine and the resultant Poss.P selects and modifies its NP complement.

(16)



(17) *oḷḷe + (y[4]) a = oḷḷeya* (lit. meaning 'of goodness', 'goodness'')

oḷḷeya-avanu would then literally mean 'he of goodness', 'goodness's one'

This idea captured in (16-17) above can be seen through, as all the members of the *oḷḷeya* group end with *–a – keṭṭa* 'bad', *hosa* 'new', *haḷeya* 'old', *yeḷeya* 'tender', *cikka* 'small' and *doḍḍa* 'big'. However, there is one problem. The suffix *–a* is also the relative clause marker in Kannada. Hence, another option would be to treat *oḷḷeya-avanu* as a relative clause in terms of 'one who is good'. The suffixal *–a* in Kannada thus is a case in syncretism. Its varied uses include being a genitive case marker, a relative clause marker and a yes-no question particle as mentioned earlier in this paper as well. Either of the two analyses (the genitive case explanation or the relative clause explanation) posit *oḷḷeya* in the domain of the NP and hence syncretism does not pose any real problem to the analysis.

A piece of evidence in support of an NP NP structure is that, the predicate NP can take various case markings.

(18a) oḷḷeya-**avan**-a / oḷḷeya-**avan**-annu / oḷḷeya-**avan**-ige / oḷḷeya-**avan**-alli / oḷḷeya-**avan**-inda

good-he-GEN / good-he-ACC / good-he-DAT / good-he-LOC / good-he-ABL

When the data in (18a) is realised in a sentence, the pronominal in the examples below, *avanu* 'he' in (18b) and *avaru* in (18c), is always non-referential.

[4] consonant epenthesis rule in Kannada phonology.

(18b) avaḷ-u obba **oḷḷeya-avan-annu** maduve-yaagalu bayasi-daḷu.
she-NOM one good-he-ACC marriage-happen desire-3.F.SG
'She desired to marry a good man.'

(18c) avan-u **oḷḷeya-avar-ige** haṇa koṭṭu sahaaya maaḍ-utt-id-danu.
he-NOM good-they-DAT money give help do-IMPER.-be-PST-3.M.SG
'He used to help by giving money to good people.'

A desirable outcome of analysing the *–a* as a genitive case marker would be that, there is a uniform paradigm for PCs occurring attributively and modifying a noun in Kannada. In such an analysis, all PCs in Kannada, whether PC nouns or PC adjectives, would take the help of genitive case in the attributive position, to modify a noun.

2.2.2 Predication of Property Concept nouns

Coming to PC nouns, a sub-classification is motivated on the basis of whether they can occur predicatively bearing dative case in nominative-subject constructions or not. For the convenience of reading, I have repeated the data in (6-7) in (19, 24 and 25). Below is the predication structure of PC noun *udda* 'height' which becomes *udda-kke* 'height-DAT' predicatively. This is the *uddakke* group.

(19) raama **udda-kke** idd-aane.[5]
Rama(nom.) height-DAT be-3.M.SG
'Rama is tall.'

. *Udda-kke* in (19) is analysed as an adjective following the analysis put forward in Amritavalli and Jayaseelan (2003) that a noun like *udda* incorporates into case to result in a syntactic category of adjectives. We may hence represent *udda-kke* as in (20) below.

[5] This construction is attested in my variety of Kannada and may not be in other varieties of the language. Further more (as far as my knowledge goes) this construction is not attested in Tamil, Malayalam, Telugu or Tulu. It is interesting to note however that this pair (7a-b) exists in Marathi. I would like to thank Ashwini Deo for bringing this to my notice (p.c).
a. Raam unchi-laa aahe. / b. Raam-laa unchi aahe.
Ram-NOM height-DAT be-3.sg / Raam-DAT height be-3.sg
'Ram is tall.' / 'Ram has (the) height.'

(20) *udda* $_{n}$ + *-kke* $_{dat}$ = *uddakke* $_{adj}$

To quote Amritavalli and Jayaseelan (2003) on the nature of the debate on derived forms such as *udda-kke* (2003; pg 70), 'Whether the derived forms are categorially compositional, or categorially different from the components, has been open to debate. There are only a few indisputable underived adjectives, such as *oḷḷeya* 'good'.' The question then becomes, what characterises a lexical category vis a vis a semantic type or a syntactic category? To answer this question would be beyond the scope of this paper and I leave this discussion to further research. The study (Amritavalli and Jayaseelan 2003) adopts the framework of Lexical Relational Structure (LRS) from Hale and Keyser (1993) to explain the role of case, the dative case specifically, to propose that all three of the representations below in (21) can be derived from the same underlying thematic structure. The thematic structure of (21a) is representative of the dative experiencer construction we have noted for the *koopa* group of PC nouns in predication. The *to* represents the dative case on the subject DP. The thematic structure of (21b) is representative of the canonical English adjectival predication and is also representative of the *udda-kke* construction we just encountered in (19). So while English incorporates the dative case into the noun to result in a separate lexical category of adjectives, Kannada provides overt evidence for such an incorporation account with *udda-kke* 'PC noun-DAT'. Kannada does not have the thematic structure of (21c) as the language does not have the verb *have*. The dative case in English incorporates with the *be* verb to result in the verb *have*. So while Kannada has (21a-b) in its inventory of thematic structures, it does not have (21c). English on the other hand attests (21b-c) but not (21a) because it has lost 'strong' case features.

(21) a. to-DP be NP e.g. raaman-ige koopa ide. [Amritavalli and Jayaseelan 2013; ex 20]

b. DP(nom.) be AdjP e.g. I am happy.

c. DP(nom.) have NP e.g. I have happiness.

Thus the occurrence of the dative experiencer construction in a language is in fact a diagnostic to suggest that such a language would not have a lexical category of adjectives. This discussion has provided a syntactic account of *udda-kke* 'height-DAT'.

We continue now to a brief yet interesting semantic account of *udda-kke* 'height-DAT'. Following Amritavalli (2013), we may analyse *udda-kke* as a resultative with the dative case encoding the path. Amritavalli (2013) discusses the example below as an accomplishment resultative in Kannada wherein the resultee, *roTTi* 'bread', is a transitive object and the noun *dappa* 'thickness' here is dative case-marked and this fulfils the role of an adjective.

(22) avanu roṭṭi-yannu **dappa-kke** laṭṭisida(nu). [Amritavalli 2013 ; ex 38]

he-NOM bread-ACC thick(ness)-DAT rolled out

'He rolled the bread out thick.'

The example below again indicates a *path,* with *udda-kke,* along which the trees grew.

(23) mara-ga[-u **udda-kke** be[e-d-avu.

tree-PL-NOM height-DAT grow-PST-3.N.PL

'The trees grew tall. (Lit. The trees grew to a height.)'

The similar construction in English encodes the semantics of a resultative with a dative *to- She honed her skills* ***to perfection*** [Amritavalli 2013; ex 40]. Thus, the *udda-kke* group of PCs are a group of PC nouns which have inherent gradability and hence can incorporate with dative case which helps denote a degree or extant to the property through a path projection from the Ramchand (2008) framework. The *udda* group of PCs as listed below in (26) denote tangible, physical properties and hence I propose, come with inherent gradability unlike the *koopa* group.

Interestingly this sub-class of nouns can also occur in a dative subject construction with the meaning of *X has the property Y for something* as in (24) below.

(24) raaman-ige **udda** id-e.

Raama-DAT height be-3N

'Raama has height.'

To distinguish these two types of constructions in (19 and 24), we can understand it better when we draw a parallel to their respective counterparts in English, *Raama is tall* and *Raama has height*. Just as in English, in Kannada too, *Raama has height* is the more marked of the two constructions. However, with the right context, the construction in (24) is fully acceptable just as it would be in English; for example, in a context like *Raama has the height (to join the basketball team)*. This aspect holds true for all the members of the *uddakke* group as listed below in (26).

The PC nouns that cannot occur in nominative-subject constructions such as *koopa*, occur predicatively, only in dative-subject constructions, as seen in (25) below.

(25) raaman-ige **koopa** id-e. / *raama **koopa-kke** idd-aane.

Rama-DAT anger be-3.N.SG / Rama(nom.) anger-DAT be-3.M.SG

'Rama is angry.' (Lit. Rama has anger) / 'Intended: Rama is angry.'

Despite not being able to find a semantic basis for the *oḷḷeya* group of PC adjectives, there is a semantic basis for the *koopa* and the *uddakke* groups. The *koopa* group of nouns (as listed in 4) encode psycho-somatic properties. The *udda(kke)* group of PCs encode physically tangible or measurable properties and are listed below in (26).

(26) Members of the ***udda(kke)*** **group** of PCs in Kannada- *udda* 'height', *saŋŋa* 'thinness', *dappa* 'fatness/thickness', *yetra* 'height', *kuḷḷa* 'shortness', *agala* 'width', *nuŋŋa* 'smoothness', *bisi* 'hotness', *taŋŋa* 'coldness', *mett* 'softness', *gaʈʈi* 'hardness' etc.

3. Conclusion

The paper began by stating that Kannada employs nouns to express adjectival meaning but also employs a small, closed-class of functional adjectives. By adopting the notion of Property Concept expressions from Dixon (1982), we referred to these two groups as PC nouns and PC adjectives. Tests of attribution and case-marking on a data set of fifty PCs in Kannada helped sort the PCs into the two groups, exemplified by *koopa* 'anger' for PC nouns and *oḷḷeya* 'good' for PC adjectives. We saw that in the attributive position, the *oḷḷeya* group of PC adjectives could modify an NP in their 'bare' form. PC nouns could not occur in their 'bare' form attributively and required genitive case to modify an NP. The predicative position revealed interesting aspects about PCs in Kannada. The *oḷḷeya* group of PC adjectives occurred predicatively in 'verbless' copular clauses. Such clauses are well-attested in Kannada and have an NP NP skeletal structure. We had two choices before us in the treatment of the *avanu oḷḷeya-avanu* 'he good-he' clause- to either treat it as an NP AP or an NP NP clause structure. We chose to treat it as an NP NP clause structure wherein *oḷḷeya* in the predicative NP was a nominalised adjective. Such a nominalisation process was explained by decomposing *oḷḷeya* as *oḷḷe* and *–a*, wherein the suffixal *–a* maybe a genitive case marker (possessive marker) or a relative clause head as the same marker is used for both purposes in Kannada. Either ways, this did not affect the analysis of the predicate ultimately being the maximal projection of a nominal head. Coming to PC nouns in the predicative position, we realise that PC nouns were not a uniform, homogenous category exemplified by *koopa* 'anger'. *Koopa* occurred predicatively only in the dative experiencer construction. A group of nouns exemplified by *udda* 'height' occurred predicatively, marked for dative case as *udda-kke* 'height-DAT' and hence analysed as PC adjectives according to the Amritavalli and Jayaseelan (2003) account. This was a syntactic account of the *udda-kke* construction and analysed *udda-kke* as a syntactically derived adjectival form. I also attempted to provide a brief account of this construction in terms of a semantic understanding of *udda-kke*. Here, the dative case in *udda-kke* denoted a PathP according to Amritavalli (2013) in which the Ramchand (2008) framework was employed to explain achievement resultatives in Kannada. Thus a PC noun like *udda* 'height' has inherent gradability or measurability and hence combines with dative case which encodes the extant or measure of *udda*. Such an analysis of the *udda-kke* group of PCs is based on very clear and transparent semantics wherein all the group members are examples of tangible, physically measurable properties such as height, weight, width, length, temperature and texture. The *koopa* group of PC nouns also have a semantic basis of being PC nouns which

encode psycho-somatic properties. The *oḷḷeya* group of PC adjectives are only about eight in number and seem to be a closed-class without a semantic basis to them. They seem to 'frozen' forms of some kind, roots probably as analysed in Menon (2013) and Balusu (2015), 'frozen' in the evolution of Dravidian as a language family as its counterparts are found across Tamil, Malayalam and Telugu.

References

Amritavalli, R. 2000. *Kannada Clause Structure*. In R. Singh et al. (eds.), The Yearbook of South Asian Language and Linguistics, pp. 11-30. New Delhi: Sage Publications.

Amritavalli, R. 2013. *Result Phrases and Dative Experiencers in Kannada*. 19th International Congress of Linguists (ICL) Papers. Edited, published and distributed by Département de Linguistique de l'Université de Genève, Rue de Candolle 2, CH-1205 Genève, Switzerland.

Amritavalli, R. and Jayaseelan, K.A. 2003. *The genesis of syntactic categories and parametric variation*. Proceedings of the 4th GLOW in Asia, pp. 19- 41. Seoul: Hancook.

Baker, Mark C. 2008. *The Syntax of Agreement and Concord*. Cambridge University Press.

Balusu, Rahul . 2015. *Comparison, predication and lexical semantics of PC nouns in Telugu*.

Cinque, G. 2010. *The Syntax of Adjectives- A comparative study*. The MIT Press.

Dixon, R.M.W., 1982. *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin:Mouton

Francez, Itamar and Koontz-Garboden, Andrew. 2015. *Semantic variation and the grammar of Property Concepts*. Language volume: 91, issue-3, pp.533-563.

Kennedy, Christopher D. 1997. Projecting the adjective: The syntax and semantics of gradability and comparison. Ph.D. Dissertation. University of California, Santa Cruz002E

Menon, Mythili. 2013. *The Apparent Lack of Adjectival Category in Malayalam and other Related Languages*. In Nobu Goto, Koichi Otaki, Atsushi Sato, Kensuke Takita eds. Proceedings of Glow-in-Asia IX 2012: The Main Session. Mie University, Japan. Pp. 157-171.

Menon, Mythili and Roumyana Pancheva. 2014. *The grammatical life of property concept roots in Malayalam*. In U. Etxeberria, A. Fălăuș, A. Irurtzun, B. Leferman, Proceedings of Sinn und Bedeutung 2013. pp. 289-302.

Sridhar, S. N. 1990. *Kannada*. London: Routledge.

Sindhu Herur Subramanya

sindhu.herur@gmail.com

Sluicing the verbal double: a view through Meiteilon[*]

Lalit Rajkumar

Jawaharlal Nehru University

I. Introduction

The aim of this study is to explain how the two contrasting issues of sound and meaning mismatch of Ellipsis and Doubling work when both strategies are used simultaneously, using Meiteilon data. It is a well-known fact that doubling is a case of over-pronunciation in the sense that the phenomenon occurs when 'one or more morphosyntactic features of a constituent (i.e., a morpheme, a word or, a phrase) are expressed in two or, more times within a sentence, seemingly without contributing to the semantic interpretation of that sentence' (Barbiers 2008, 2013). That is, an extra pronunciation of a copy (x^i) of an item (x) where the pronunciation of x^i does not add up to the meaning already conveyed by x.

1. **(əy**-gi) puk-p^{h}ə-bi **i**-mɑ

   1P–Gen stomach-good-Nzr (Fem.) **1P**+Gen-mother

   'my benevolent mother'

On the other side, ellipsis is an issue of under-pronunciation which involves 'the omission of elements that are inferable from the context' (Aelbrecht 2010).

2. tombə-nə **həynəw** əmunbə čɑ-gəni ədugə

   Tombə-Subj **mango** ripe eat-will and

   tombi-di _____ əsəŋbə čɑ-gəni

   Tombi-Top[1] green eat-will

[*]I sincerely thank the participants of FASAL-5 for their insightful comments and suggestions. This paper owes a lot to Ayesha Kidwai, Tanmoy Bhattacharya, and Sjef Barbiers and also to the anonymous reviewer for their guidance in giving some shape to it. I am solely responsible for the errors and mistakes found in the paper.

'Tomba will eat ripe mangoes and Tombi will eat green/unripe _____'

However, the paper concentrates more on the formation of verb-doubling constructions in Meiteilon. The reason for concentrating on this phenomenon is that, it will provide a better platform for its comparison with/co-occurrence in a sluiced environment. Through this investigation, the paper further intends to illustrate that sluicing appears to solve the over-pronunciation problem of verb doubling.

II. Verb doubling in Meiteilon

Verb doubling in predicate-cleft constructions is a well-researched topic (see Koopman 1984, 2000; Abels 2001; Kandybowicz 2007, 2008; Trinh 2011 and others) but a common consensus has not been reached yet on its formation. The idea so far conceived about it is that verb-doubling is resulted by verb-topicalization thereby giving a contrastive interpretation. Nevertheless, the unit of topicalization is always an issue with verb-topicalization resulting to doubling. Studies on different languages like Vata (Koopman 1984, 2000), Spanish (Vicente, 2007), Hebrew and Vietnamese (Trinh, 2011) show that it is only the verb which is fronted; while on the other side, languages like Russian (Abels, 2001) and Hebrew (Landau, 2006) can have the internal arguments along with the fronted verb. In Meiteilon, the verb doubling in predicate cleft construction (PCC) looks even more peculiar as it appears that the whole vP is topicalized. The higher copy is marked by either a Topic or Focus marker attached to its non-Finite form[2] whereas the lower copy is the one that bears the verbal inflections (Achom et al, 2013; Rajkumar, 2014a). An example is given below

3.	əy	sɑ	**čɑ**-bə-di	*(**čɑ**)-re[3]	(obligatory doubling)
	I	meat	eat-Nzr-Top	eat-Perf	

[1] Topic marker *–ti* becomes *–di* after a voiced sound.

[2] The higher copy of the verb being non-finite in doubling constructions is not an uncommon occurrence as it has been found to be the case in languages like Russian (Abels, 2001),Spanish and Hungarian (Vicente, 2007), Hebrew and Vietnamese (Trinh, 2011) also.

[3] As it is the case with doubling, an intervener can be inserted in between the doubled items.

əy	sɑ	**čɑ**-bə-di	toynə	**čɑ**-re
I	meat	eat-Nzr-Top	often	eat-Perf

'As for me eating meat, it is often done'

'As for me eating meat, it is eaten' [literally: 'I have started eating meat']

The oddity in the translation is brought by the manner in which the verb is doubled in PCC. That is, as already mentioned above, it appears that in the Meiteilon verb doubling construction it is not just the verb or, the verb along with the internal argument which is fronted for topicalization. It, instead, looks to be the case that the external argument i.e, the Subject is also an important element for verb doubling in the language. An example each of the different types of verb-fronting in different languages are as follows:

4. **Jugar**, Juan suele jugar al futbol los domingos
   play.INF Juan HAB play.INF at soccer the sundays

   'As for playing, Juan usually plays soccer on Sundays.'

Spanish (Vicente 2007)

5. Dumat' čto Xomskij genij on dumaet no
   think that Chomsky genius he thinks but

   čitat' **ego** **knigi** ne čitaet
   read his book not reads
   'He does think Chomsky is a genius, but he doesn't read his books'

Russian (Abels, 2001)

6. **əy** **sɑ** **čɑ**-bə-di *(čɑ)-re

   I meat eat-Nzr-Top eat-Perf

   'As for me eating meat, it is done' [literally: 'I have already eaten meat']

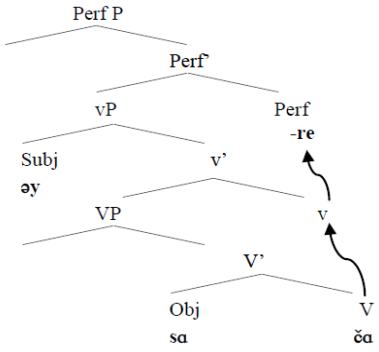
Meiteilon

Before one jumps to the complexities of verb-doubling constructions, let us first go through the derivation of the verb-undoubled version of (6)

7. əy sɑ čɑ-re

   I meat eat-Perf

   'I have eaten meat'



Verb doubling of the type in (6) requires topicalization. In Meiteilon, this is achieved for arguments by marking them with the Topic particle *–ti*. So, let us now check whether topicalization is a sufficient step for doubling in case of the arguments.

8. **Subject Topicalization**

a. əy-*di* sɑ čɑ-re …

I –Top meat eat-Perf

'I have eaten meat (…you/s/he might have done something else)'

b. *əy-di **əy** sɑ čɑ-re

I –Top I meat eat-Perf

9. **Object Topicalization**

a. əy sɑ-*di* čɑ-re …

I meat-Top eat-Perf

'I have eaten meat (…but, not other things)'

b. *əy sɑ-di **sɑ** čɑ-re

I meat-Top meat eat-Perf

Thus, the above examples (8a,b and 9a,b) show that topicalization in itself is not a sufficient condition for doubling. However, for verb doubling to occur, topicalization or, focalization[4] of

[2] Due to the space limitations, the paper will not discuss the other focalized cases of verb doubling.

the verb is a necessary procedure for the double occurrence[5]. Let us now see how simple V-fronting or, VP-fronting is not enough for topicalization of the verb in Meiteilon.

Verb-fronting

10. *[*ča*-bə]-di əy sɑ *ča*-re

    eat-Nzr-Top I meat eat-Perf

VP-fronting

11. *[sɑ *ča*-bə]-di əy *ča*-re

    meat eat-Nzr-Top I eat-Perf

vP-fronting

12. [əy sɑ ***ča***-bə]-*di* ***ča***-re
    I meat eat-Nzr-Top eat-Perf
    'As for me eating meat, it is done'

It is now clear that verb doubling is licensed by overt vP movement for topicalization[6]. As example numbers (10) and (11) above show that neither just V-topicalization (10) nor VP-topicalization (11) is sufficient to license the verbal double. Rather, as in (12), the vP must topicalize. It is worth noting that Meiteilon has do-support which occurs only in such verb-

[5] Verb-Topicalization/Focalization, however, does not always result to verb doubling as it can be a case of predicate-cleft construction without verb doubling where the lower copy (from head-movement) is not pronounced. The lower copy's function of supporting the verbal inflection is either done by a do-support (Rajkumar, 2014b) or, by a modal.

əy sɑ čɑ-bə-di yɑ-y

I meat eat-Nzr-Top Possibility-Ind

'As for me eating meat, it is possible' [Literally: 'I may eat meat']

[6] In order to show that vP forms a constituent a proform test using a demonstrative pronoun 'that' is performed on a similar verb-doubling construction. (However, the language lacks simple sentence-initial adverbs but they occur either as reduplicated elements like *itʰa tʰaǰədənə* 'unbelievably' or, as complex constructions with verb 'say/speak' eg. *seŋnə hayrəbədə* 'honestly speaking'. So, I would not be using examples with such abverbial constructions.)

a. **tombə** **tʰəbək-tu** **təw-bə**-di təw-re ədubu loy-bə-di loy-d-ri
Tomba work-that do-Nzr-Top do-Perf but finish-Nzr-Top finish-Neg-Prog
'As for Tomba doing the work, it is done but it is not yet completed'

a'. mədu-di təw-re loy-bə-di loy-d-ri
that-Top do-Perf finish-Nzr-Top finish-Neg-Prog
'That is done, but it is not yet completed'

fronting environments without giving any change in the meaning of (12) when the lower copy of the verb is replaced by a 'meaningless-DO' (Rajkumar, 2014b).

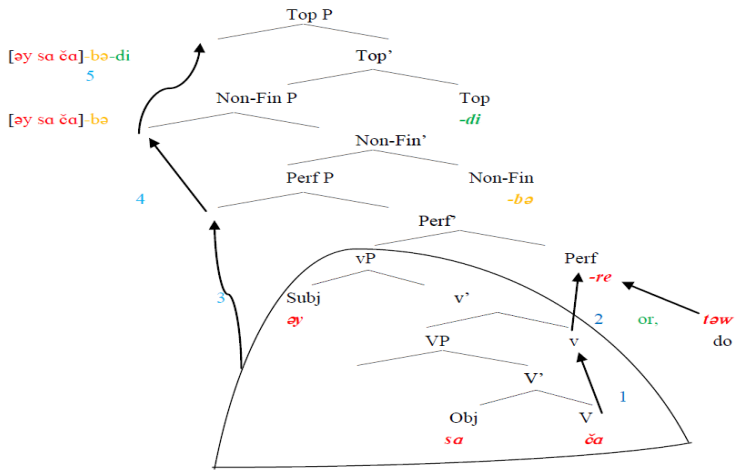
13. [əy sɑ ***ča***-bə]-*di* təw-re

I meat eat-Nzr-Top DO-Perf

'As for me eating meat, it is done'

The derivation of (12) and (13) is combined as (14) in the tree structure below:

14.



Taking into account the undoubled occurrence in example (7), it is evident that the head movements of V-to-v (movement (1)) and then to Perf (movement (2)) take place. The crucial point that I would like to make here is that, at this particular point the verb (V-v attached to the Perf-head) plays a semantically active role[7]. Following the assumption that head movement is a syntactic operation and not a PF-component, it allows us to recover the V-v complex copy which is not yet deleted after movement (2). As the construction requires verbal fronting for

[7] This line of argument follows the arguments put forward by the likes of Matushansky (2006); Bhatt and Keine (2014) and others who have shown that head-movement should not be pushed off as a PF-operation. Instead, head movement does play a role in the syntax and semantics of certain derivations like the German Verb Incorporation (VI) proposal by Bhatt and Keine (2006) to account for verb clusters in the language. Similarly, the numerous deictic particles of Meiteilon are incorporated into the verb stem through head movements which are semantically non-vacuous.

tombə ləmpɑk-lom-tə čen-sil-lu-re

Tomba playground-towards-Loc run-inwards-D.Deic-Perf

'Tomba went, ran towards and on the playground'

topicalization and recalling the need to front the whole vP, it is raised to Spec, PerfP while searching for the nearest Spec (movement (3)) which accidentally results in a Spec-Head configuration of vP with its earlier raised head i.e, V-v. It does not look like a typical Spec-Head agreement configuration as the relationship is between a vP with its complex head[8]. Movement (4) deletes the base copy of the vP at Spec, PerfP[9]. Moreover, the vP needs to raise further for topicalization but before that the clause must become a non-finite in order to get topicalized. So, movement (4) is an obligatory step for verb-topicalization in Meiteilon. Interestingly, after the completion of movement (4), the lower V-v copy from head movement becomes semantically redundant. But, in the presence of another verbal element like a modal in its place[10], verb-doubling does not occur as it is in the case of undoubled predicate-cleft constructions[11].

15. əy sɑ čɑ-bə tɑ-re

I meat eat-Nzr OBL-Perf

'I have to eat meat'

From the above examples, it is not very clear at this juncture whether such an instance is a case of a doubling chain formation or, something else. However, it is indeed very clear that the verbal copy, which resulted from head movement, has become semantically vacuous after movement (5) for topicalization. Thus, the verbal copy at Perf may remain either as it is or, may be replaced

[8] This co-occurrence of doubling with a potential configuration for agreement should be researched in a deeper way as Barbiers (2008) has claimed that agreement is a sub-case of syntactic doubling. Moreover, except for movement to Spec, PerfP (which is too local as pointed out by the reviewer), the vP constituent movements are linearized in a Spec, Head configuration which is needed to get the required markings of the Head onto the moved constituent when it lands on its Spec.

[9] The reviewer has suggested for a lower position of the non-finite marker *-bə* located somewhere below the Perfective projection which was also my initial thoughts but it is not the case as it is also shown in Bhattacharya and Thangjam (2004) as a C^0-projection. Moreover, extending the head movements of the verb to the Topic head via the Perf and Non-Fin heads sequentially give rise to an If… then clause/conditional construction in the language.

[10] But never a full verb like 'sleep'

*əy sɑ čɑ-bə tum-me

I meat eat-Nzr sleep-Perf

[11] But, it looks more likely the case that because of the presence of a modal head, the lexical verb head cannot cross it following Head Movement Constraint (Travis 1984). Hence the verb in this case is not doubled without the much needed head movement and the vP simply undergoes fronting. Another, interesting fact is that the modals can also be doubled using the same mechanisms as verb-doubling but it is not discussed here due to its divergence from the objective of this paper.

by a meaningless *do* since it is semantically redundant (Rajkumar, 2014b). It should also be noted that the vP at Spec, Non-FinP is also deleted after (5).

Let us now check the strength of our hypothesis on verb doubling in PCC in Meiteilon by adding a NegP into the structure.

II.I Adding a NegP (illustration of 16a,b,c is given in 17)

16.

a. əy sɑ **čɑ**-bə-di **čɑ**-d-re

I meat eat-Nzr-Top eat-Neg-Perf

'As for me eating meat, it is not done' [or, 'I have stopped eating meat']

b. əy sɑ **čɑ**-bə-di **təw**-d-re

I meat eat-Nzr-Top do-Neg-Perf

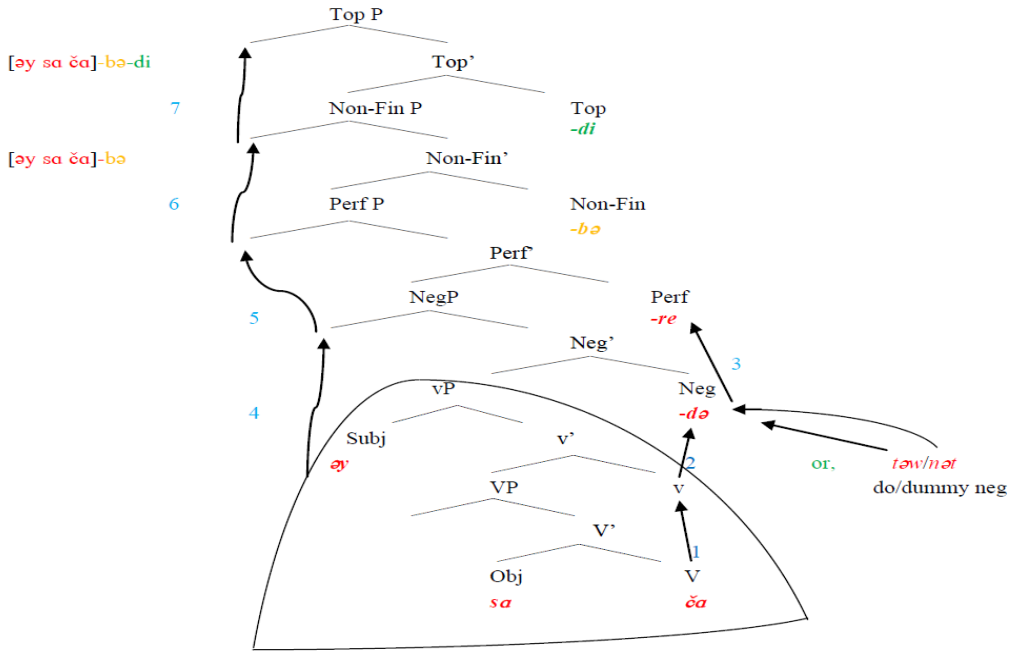
'As for me eating meat, it is not done' [or, 'I have stopped eating meat']

c. əy sɑ **čɑ**-bə-di **nət**-t-re

I meat eat-Nzr-Top Dummy Neg -Neg-Perf

'As for me eating meat, it is not done' [or, 'I have stopped eating meat']

17.



As examples (16a,b,c) are still cases of vP-topicalization, there is not much change in the structure except for the fact that addition of a semantically active Neg introduces another dummy insertion in the form of *nət* which is glossed as a *dummy Neg*. The presence of a dummy version of the negation is possible only when it attaches to a semantic negation because it is just like a negative version of a do-support. Thus, the possibility of these dummy insertions namely, *təw* 'do' and *nət* 'dummy Neg' in place of the lower V after vP-topicalization, without any change in the meaning, confirms the earlier formulation that the lower copy of the verb becomes semantically redundant after the vP has been topicalized.

We can now move forward with a more ambitious step by increasing the size of the unit of topicalization from just a vP[12].

II.II NegP-Topicalization construction

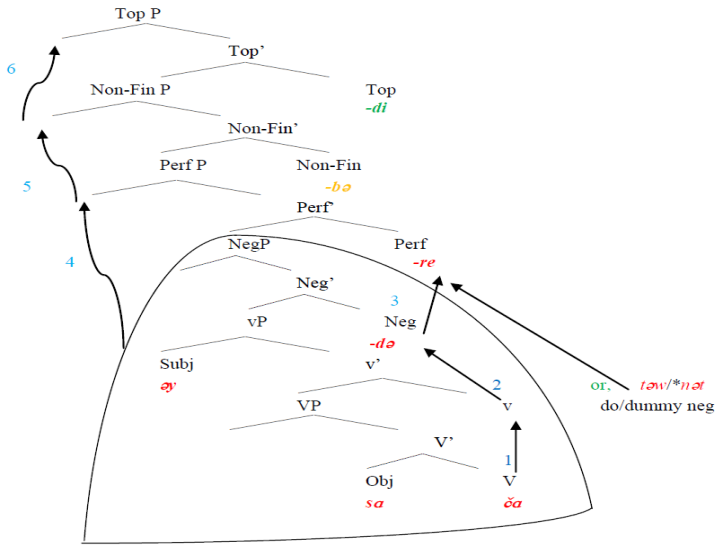
18. əy sɑ **čɑ-də**-bə-di **čɑ-d/təw/*nət**-re

[12] It is to be noted that, whatsoever morpheme which can come in between the verb stem and the nominalizer *-bə* can become the unit of topicalization in Meiteilon verb constructions- for example, the causative morpheme *-hən*, the evidential marker *-ləm*, the deictic particles *–sin*, *-t^{h}ok* etc.

I meat eat-Neg-Nzr-Top eat-Neg/do/Dummy Neg-Perf

'As for not eating meat, it is not done' [or, 'I have stopped eating meat']

19.



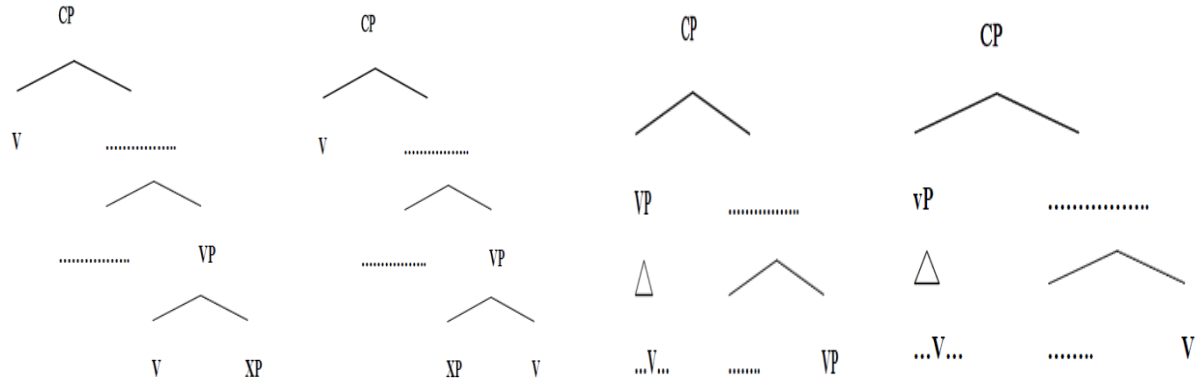
As the unit that we have chosen for topicalization is a NegP and not a vP, it is not a simple case of verb doubling through vP-topicalization but rather, a case of negative doubling which contains verb doubling inside the structure. The movement procedures are almost the same as verb doubling. The V-v-Neg complex, merged as a unit through head movements 1-2-3, has escaped the ellipsis site (which happens after movement (4)). Hence, with similar fronting phenomenon as that of a vP, the NegP undergoes deletions of its lower copies. The different and most interesting part of this NegP-topicalization is that *do*-support is able to substitute the lower V+Neg after movement (6) as both the V and the Neg are meaningless in this position thereby, conforming to the earlier postulation. Hence, the dummy Neg element *nət* cannot be inserted to the Perf head as it has to be always licensed by a semantically active Neg which is not present anymore after the NegP topicalization.

Since the analysis for verb doubling still holds good even with the addition of a NegP both in cases of vP-topicalization and NegP topicalization. Trinh (2011) has proposed three types of predicate-clefting structures, namely- Type-1, Type-2, and Type-3 in order to account for the presence of doubling and non-doubling structures of verb. Due to space limitations, only the diagrammatic representations of Trinh's predicate-clefting types are shown in (20a,b,c). In

addition to these three types of predicate-clefting structures, I propose a fourth type referred as Type-4, structurally illustrated in (20d).

20.

a.Type-1 **b.Type-2** **c.Type-3** **d.Type-4 (Meiteilon)**



III. Sluicing the doubled verb

Let us look at the construction which involves the interaction of verb-doubling and sluicing. Following from the earlier section (II), we have seen that topicalization is a motivation required by the verb to undergo doubling. We will now see what happens to the verb-doubling construction (in the *reconstructed* ellipsis-site) when the motivation i.e, topicalization is snatched away by the remnant in the sluice.

21. əy-nə ŋərɑŋ kəri-no əmə **ča**-bə-**di** **ča**-k^{h}-re
I –Subj yesterday something eat-Nzr-Top eat-?-Perf

ədubu əy kəri-no (hɑy-bə)*(-**di**) niŋsiŋ-d-re
but I what-Q say-Nzr-Top remember-Neg-Perf

[*əy-nə ŋəraŋ* ***ča**-k^{h}i-bə* *pot* *ədu*]---**RECONSTRUCTED ELLIPSIS SITE**
I –Subj yesterday eat-?-Nzr thing that

'As for me eating something, I did yesterday but I don't remember what' [the thing that I ate yesterday]

The above example (20) shows that sluicing requires obligatory topicalization of the remnant. And, it is already shown earlier in section (II) that topicalization plays a major role in verbal doubling construction of Meiteilon. It is then very evident that even if the antecedent has verb doubling in the structure, the ellipsis site will most probably not have it as the remnant is obligatorily topicalized. Therefore, there is no verb doubling at the ellipsis-site, but rather just head movement of the verb. Since, the sluice as one constituent can have only a single topicalized unit which is the remnant here[13], the verb in its ellipsis-site cannot have it anymore.

IV. Conclusion

Hence, through the investigation of the sound-meaning mismatch problems through Meiteilon data, four proposals have been made in this paper. Firstly, the paper has illustrated that in a verb doubling construction, the higher copy of the verb is in its non-finite form. Secondly, it also demonstrates that NegP-topicalization introduces the possibility of Negative concord-like doubling in the language that lacks subject-verb agreement. Thirdly, the paper proposes for a Type-4 variety of predicate-clefting construction in addition to the earlier three types of predicate-clefting with evidence from Meiteilon. Lastly, it argues that sluicing the verb doubling construction seems to have repaired the sound-meaning mismatch of the later at the ellipsis-site. Apart from the four proposals, this paper also intends to highlight a further issue which is to check for the correlation between topicalization (and/or focalization) and the two sound-meaning mismatch problems (i.e., doubling and ellipsis); an issue which I aim to undertake in my future work.

Abbreviations

1P- First Person; **D.Deic-** Distal Deictic; **Fem-** Feminine; **Gen-** Genetive case; **HAB**- Habitual; **Ind-** Indicative; **INF**- Infinitive; **Loc-** Locative; **Neg-** Negation; **Non-Fin**- Non-Finite; **Nzr**- Nominalizer; **OBL**- Obligation; **Perf**- Perfective; **Top**- Topic

[13] This is a very weak postulation as the paper has not discussed with instances of topicalized remnant with un-doubled verbs in the antecedent. A preliminary investigation shows that the remnant still needs a topic marker. So, a much more detailed study is required on the interaction of doubling and ellipsis which cannot be covered in the limited space of this paper.

Reference

Abels, Klaus. 2001. The Predicate Cleft Construction in Russian. In *Formal Approaches to Slavic Linguistics 9*, eds. Steven Franks and Michael Yadroff, 1-19. Bloomington, IN: Michigan Slavic Publications.

Achom, P., Ashem, R., Hidam, G.S., Khaidem, A., Oinam, Ng., Rajkumar, L. 2013. Syntactic Doubling in Meiteilon presented in *Sjef Barbier's Workshop at LISSIM-7*, Sidhbari.

Aelbrecht, L. 2010. *The syntactic licensing of ellipsis*. Amsterdam: John Benjamins.

Barbiers, S. 2008. Microvariation in Syntactic Doubling: An Introduction. *Syntax and Semantics 36*, 1-34.

Barbiers, S. 2013. *Introduction to doubling and to Dutch*. Sidhbari, Himachal Pradesh: Classes taught in LISSIM-7.

Bhatt, R. and Keine, S. 2014. "Two roots in one phase: An environment for semantically contentful head movement". Paper presented at *WCCFL 32*, University of Southern California.

Kandybowicz, J. 2007. On Fusion and Multiple Copy Spell-out: The Case of Verbal Repetition. In Norbert Corver and Jairo Nunes (Eds.), *The Copy Theory of Movement* [Linguistik Aktuell/Linguistics Today 107], 119-150. Amsterdam: John Benjamins.

Kandybowicz, J. 2008. *The grammar of repetition: Nupe grammar at the syntax-phonology interface*. John Benjamins Publishing Co.

Koopman, H. 1984. *The syntax of verbs: From verb movement rules in the Kru languages to universal grammar*. Foris.

Koopman, H. 2000. *The Syntax of Specifiers and Heads: Collected Essays of Hilda j. Koopman*.

Routledge.

Landau, I. 2006. Chain resolution in Hebrew V(P)-fronting. *Syntax 9*: 32-66.

Matushansky, O. 2006. Head Movement in Linguistic Theory. *Linguistic Inquiry 37.1*: 69-109.

Merchant, M. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford: Oxford University Press.

Rajkumar, L. 2014a. "Verbal doubling in Meiteilon". Paper presented at the *30th South Asian Languages Analysis Roundtable* (*SALA Roundtable-30*), 6th to 8th February, 2014 held at University of Hyderabad.

Rajkumar, L. 2014b. "DO-support in Meiteilon: questioning the language specific-ness". Paper presented at the *36th International Conference of Linguistic Society of India* (*ICOLSI-36*), 1st to 4th December, 2014 held at University of Kerala.

Travis, L. 1984. *Parameters and effects of word order variation*. Doctoral Dissertation, MIT.

Trinh, T. 2011. *Edges and Linearization*. Doctoral Dissertation, MIT.

Vicente, L. 2007. *The Syntax of Heads and Phrases*. *A Study of Verb (Phrase) Fronting*. Ph.D. Dissertation, Leiden University. LOT Dissertations 154, Utrecht.

Lalit Rajkumar

litrajkumar@yahoo.com

Sluicing in Dravidian: Tracing the source [1]

Rahul Balusu

The English and Foreign Languages University, Hyderabad

1. Introduction

Conventionally, Dravidian languages are described as being *wh*-in-situ (Subba Rao 2012). Jayaseelan (2001 *et seq.*) proposes that there is (partial) *wh*-movement to IP internal Focus positions, taking the fact that the *wh*-subject surfaces not clause-initially but between the objects and the final verb, as the primary evidence, among other diagnostics. In a revival of the *wh*-in-situ idea, Mathew (2014) argues that the finite verb in Dravidian is in C, with the *wh*-material in-situ, while the other arguments move to Topic positions, creating the illusion of *wh*-movement to a preverbal position. This paper seeks to establish which of these proposals is on the right track for Dravidian, using as the divining rod, characteristics of sluicing, a phenomenon that has not been discussed in any of the previous literature on Dravidian.

Dravidian has sluicing, as shown in (1), with a matrix sluice in (1a), and an embedded sluice in (1b) —all the data in this paper are from Telugu, but the same essential facts obtain in the other major Dravidian languages. Cross-Dravidian differences will surely crop up at various points as the finer details of the terrain of sluicing are covered, and will be explored as the project progresses. I am using here the terminology of Vicente (2014), where the term 'sluicing' applies to any construction with a *wh*-remnant.

(1) a. raamu eed-oo konnaaDu. —avunaa?! eemiTi?
Ramu what-DISJ bought. Really what
'Ramu bought something. —Really?! What?'

b. raamu eed-oo konnaaDu, kaani eemiT-oo naa-ku tel-iyadu.
Ramu what-DISJ bought, but what-DISJ I-DAT know-not
'Ramu bought something, but I don't know what.'

If Dravidian is *wh*-in-situ (and *wh*-movement to Spec CP, followed by deletion of IP, is

[1] I would like to thank an anonymous reviewer, and audiences at FASAL 5 and Grasping ELLIPSIS for valuable comments and suggestions. Any errors are my own.

unavailable), there are two possibilities for the source. The first, the source for sluicing (or in another terminology, pseudo-sluicing[2] (Merchant 1998)) is copular –either a simple copular structure, with null subject and null copula (as in Chinese; Adams (2004), Wei (2004)), or a reduced cleft structure[3] (as in Japanese; Kizu 1997). The second, is the view of sluicing proposed for *wh*-in-situ languages like Farsi (Toosarvandani 2008) and Turkish (Ince 2012), that there is exceptional *wh*-movement to Spec CP in sluicing, followed by deletion of IP, and that it is not the weak *wh*-features that result in the overt *wh*-movement, but what causes and permits the overt raising of *wh*-expressions to the CP-domain is the checking of focus rather than *wh*-features, that come into play in the information structure mechanism needed for sluicing.

On the face of it, the IP-internal *wh*-movement structure for Dravidian (Jayaseelan 2001, *et seq*.) runs into trouble as a possible source for the sluice, because the *wh*-remnant would be an intermediate position that is pronounced, while phrases on either side of it are elided, and they don't form a constituent, as shown in (2). But another structure[4] that is formed in this derivation, prior to the subject moving out of the vP could be a viable source for the sluice, as shown in (3), because the elided material is one contiguous constituent separate from the *wh*-remnant.

(2) ..., kaani [$_{IP}$ ~~raamu$_j$~~ I [$_{FocusP}$ eed$_i$] [$_{vP}$ ~~t$_j$ konnaaDu t$_i$~~]] -oo] naa-ku tel-iyadu
..., but what bought -DISJ I-DAT know-not

(3) ..., kaani [$_{IP}$ I [$_{FocusP}$ eed$_i$] [$_{vP}$ ~~raamu konnaaD t$_i$~~]] -oo] naa-ku tel-iyadu
..., but what bought -DISJ I-DAT know-not

This account will not have any problem accounting for the case-matching effects that we will see later on, because the *wh*-remnant comes from a case-marking position. It will also not have a problem with multiple-sluicing (that is possible in Telugu), because multiple *wh*-elements can occur in focus positions in the preverbal field in the IP (Jayaseelan 2001).

Mathew (2014) claims that in the verb-final construction in Malayalam, the verb undergoes V-to-C movement, while the *wh*-phrase is in-situ, and it has no freedom of movement[5]. All the other elements in the clause mandatorily move out of the vP to higher Topic positions, making the *wh*-word appear in the immediately preverbal position, creating the illusion that the *wh*-word has moved to the left of V. [6] The sluice source in such an analysis

[2]If sluicing is defined as IP-deletion, leaving behind a CP remnant, where the moved *wh*-element is pronounced in SpecCP, then the term cannot apply to *wh*-in-situ languages. Instead such constructions are called Sluicing Like Constructions (SLC), as in Manetta (2013), Paul & Potsdam (2012), Gribanova (2013).

[3]This is also termed a Truncated Cleft. Even some languages exhibiting overt *wh*-movement have been claimed to exhibit cleft strategies to form sluicing structures (Vicente 2008, van Craenenbroeck 2010).

[4]A structure with a null subject in Spec IP is also a possible source, but that would mean the subject can sometimes be overt. But it never shows up in the sluices, and therefore rules this possibility out.

[5]She proposes another structure for clefts, the *aanu* construction, that demands obligatory movement of certain *wh*- phrases to the pre-auxiliary position, which is much like the exceptional *wh*-movement to SpecCP structure we have already enumerated. Hence we won't consider it separately.

[6]This analysis is contra the analysis in Jayaseelan (2001), that the *wh*-phrases undergo obligatory movement to a Focus phrase at the left periphery of the vP in Dravidian, and that the other elements in the complement of V then moves to the left of the Focus position, to higher Topic positions.

would be a structure like the one given in (4). At no point in the derivation would the subject and the verb form a constituent to the exclusion of the *wh*-remnant. Therefore, deletion would have to involve a discontinuous string, or there would have to be two deletion operations targeting two different constituents. This problem seems unsurmountable for this approach, if the source of the sluice has to come from a normal *wh*-structure derivation.

(4) ..., kaani $[_{CP}[_{TopP}$~~raamu~~ $_{J}]$ $[_{vP}t_i$ eedi ~~konnaaDu~~] -oo] naa-ku tel-iyadu
..., but what bought -DISJ I-DAT know-not

As pro-drop and null copula are both features of Dravidian, as shown in (7)-(9), a copular structure, like in (6), could be a possible source for the sentence in (1b). But a copular structure is ruled out because the *wh*-remnant gets variously case-marked, obligatorily matching the case on the correlate[7], as shown in (11). The subjects in copular structures mostly bear nominative case, and cannot bear accusative case as shown in (10).

(6) raamu eed-oo konnaaDu, kaani [*pro* eemiT-oo ϕ] naa-ku tel-iyadu.
Ramu what-DISJ bought, but what-DISJ I-DAT know-not
'Ramu bought something, but I don't know what.'

(7) vacc-eeDu
came-3MS
'(He) came.'

(8) idi pustakamu
this book
'this is a book'

(9) naa-ku rendu carlu
I-DAT two cars
'I have two cars'

(10) *nannu presidentu
I-ACC president
'Me, President'

(11) a. raamu evari-n-oo koTTeeDu, evari-n-oo/*evar-oo naa-ku telusu.
Ramu who-ACC-DISJ hit, who-ACC-DISJ/who-DISJ I-DAT know
'Ramu hit someone, I know who(m).'

b. raamu evari-k-oo pustakam icceeDu, evari-k-oo/*evar-oo naa-ku telusu.
Ramu who-DAT-DISJ book gave, who-DAT-DISJ/who-DISJ I-DAT know
'Ramu gave the book to someone, I know who (to).'

However, when a demonstrative is present along with the *wh*-word, a copular structure is clearly the source, as shown in (12).

(12) raamu eed-oo konnaaDu, kaani [adi eemiT-oo] naa-ku tel-iyadu.
Ramu what-DISJ bought, but that what-DISJ I-DAT know-not
'Ramu bought something, but I don't know what that (is).'

[7] Telugu obeys the 'Form Identity Generalisation I' of Merchant (2001):

(5) *Form Identity Generalisation I: Case-Matching*
The sluiced *wh*-phrase must bear the case that its correlate bears.

At this point, after our initial exploration of sluicing in Dravidian, we are left with three viable options for the source structure in sluicing constructions –a cleft source (with ellipsis of the CP cleft clause that the *wh*-pivot comes out of); a source with exceptional *wh*-movement to SpecCP (induced by focus features) and deletion of IP; and, an IP-internal move-and-delete source (with movement of *wh*-remnant to IP-internal FocusP and deletion of the vP with the subject in Spec vP). In one source the CP is deleted (reduced cleft), in the second source, the IP is deleted (exceptional full *wh*-movement), and in the third, the vP is deleted (IP-internal *wh*-movement). We can test for the height and extent of the deletion in Dravidian sluicing by testing with various elements that occur at various heights in the clausal tree, like adverbs, negation, auxilliaries, etc., to see if they survive elision in the sluice or not. If they do, elision is below that height, and if they don't, elision encompasses that height. Thus, such carefully constructed data will help us choose between these three possible sources for Dravidian sluicing.

2. Tracing the source: Is it CP, IP, or, vP deletion?

2.1 Testing with material positioned between IP and vP

Here we will use some of the tests developed by Manetta (2013) for diagnosing the sluicing structure in Hindi-Urdu.

2.1.1 Sentential negation in sluicing structures

In Telugu, when the correlate has negation, negation has to be interpreted inside the site of the ellipsis, as shown in (13).

(13) a. raamu eed-oo kon-a-leedu. eemiT-oo naaku teliy-adu
Ramu what-DISJ buy-NEG. what-DISJ I-DAT know-NEG
'Ramu did not buy something. I don't know what.'
b. = Ramu did not buy something. I don't know what Ramu did not buy.
c. ≠ Ramu did not buy something. I don't know what Ramu bought.

Also, negation cannot appear outside the ellipsis site, as shown in (14).

(14) raamu eed-oo kon-a-leedu. *eemiTi kaad-oo naaku teliy-adu
Ramu what-DISJ buy-NEG. what not-DISJ I-DAT know-NEG
'Ramu did not buy something. I don't know what not.'

It is standardly assumed that negation is projected below the IP, and the verb picks up the negation suffix along the way to I. If sluicing in Dravidian is elision of the vP, with the *wh*-remnant in the IP-internal, vP left-adjacent FocusP, then negation should not be present in the interpretation in (14), and it should be possible for negation to appear outside the site of ellipsis in (14), along with the *wh*-remnant. But neither of them is the case here.

2.1.2 Speaker/Subject oriented adverbs

Along the same lines as the tests with negation, testing with adverbs adjoined above the vP, as shown in (15) and (16), again reveals that the elision has to include material above the vP.

(15) a. telivi-gaa evar-oo daakunnaaru. evar-oo naaku telusu
intelligence-gaa who-DISJ hid. who-DISJ I-DAT know.
'Cleverly, someone hid. I know who'
b. = Cleverly, someone hid. I know who clevery hid.
c. ≠ Cleverly, someone hid. I know who hid.

(16) telivi-gaa evar-oo daakunnaaru. *telivi-gaa evar-oo naaku telusu
intelligence-gaa who-DISJ hid. intelligence-gaa who-DISJ I-DAT know.
'Cleverly, someone hid. I know who cleverly.'

2.1.3 Does Dravidian have vP ellipsis in general?

If the elision of vP-sized constituents is generally not possible in Dravidian, it will make the vP level elision in sluicing a less likely possibility, as that would make it a special and exceptional mechanism, that is not otherwise available in the language. By examing V/v-stranding vP ellipsis, and possibly light-verb ellipsis (Toosarvandani 2009), we can figure out how widely available vP ellipsis is as a strategy in Dravidian. This exploration is left for future research, following the lead of Takahashi (2013), and Simpson *et al.* (2013).

At the end of this sub-section, we can conclude that the height of the elision in the Dravidian sluice is above the level of the vP, based on the two tests we deployed.

2.2 Is the sluice derived by IP deletion after high focus movement?

There are languages where *wh*-fronting triggered by high focus movement, mediated by a [Foc]-feature, followed by deletion of the IP, feeds sluicing, as is proposed for *wh*-in-situ languages like Farsi (Toosarvandani 2009), & Turkish (Ince 2012), and *wh*-movement languages like Hungarian, Czech & Romanian (van Craenenbroeck & Liptak 2013). Could the Dravidian sluice structure also involve exceptional focus movement of the interrogative phrase to a left peripheral position in the C domain, followed by deletion of the rest of the clause, the IP?

Toosarvandani (2009) proposes that sluicing in Persian is fed by movement to a high focus projection (above IP). There is evidence that this position is independently active in Persian for contrastive focus (Karimi 2003), as shown in (17)

(17) giti midune ke pesTE sohraab xaride. FARSI
Giti know.3SG that pistachio Sohrab bought.3SG
'Giti knows that Sohrab bought PISTACHIOS.'

While Dravidian features focus movement, it is not left-peripheral but to a preverbal position inside the IP. Previous work (Jayaseelan 1999, 2001) suggests that the unmarked position for both interrogative and non-interrogative focus in Dravidian is low, immediately preceding the clause-final verb, as shown in (18)-(20).

(18) raamu pustakam ravi-ki icceeDu
Ramu book Ravi-DAT give-PST
'Ramu gave the book to RAVI.'

(19) pustakam ravi-ki raamu icceeDu
book Ravi-DAT Ramu give-PST
'RAMU gave the book to Ravi.'

(20) giti-ki telusu sohrab pistacio-lu konnaaDu ani
Giti-DAT know Sohrab pistachios bought.3SG that
'Giti knows that Sohrab bought PISTACHIOS.'

So there is no evidence that a high focus position is independently active in Telugu, and a sluicing structure derived by high focus movement of the *wh*-element to the CP, followed by IP deletion is unlikely.

2.2.1 Comparison with focus fronting in Persian

If sluicing involves deletion of IP, then we expect that the complementizer should be able to appear in a sluice. This expectation is borne out in Persian (Toosarvandani, to appear). Sluicing in Persian can leave behind not just a *wh*-phrase remnant, but also a complementizer and a topicalized phrase.

But this is not possible in Dravidian. Sluicing can leave behind neither the complementizer nor a topicalized phrase along with the *wh*-remnant, as shown in (21)-(22)

(21) raamu eed-oo konnaaDu, *kaani eemiT-oo ani naa-ku tel-iyadu.
Ramu what-DISJ bought, but what-DISJ that I-DAT know-not
'Ramu bought something, *but I don't know that what.'

(22) raamu pustakaalu bommalu evari-koo icceeDu. *naa-ku telusu pustakaalu evari-koo.
Ramu books toys who-DISJ gave. I-DAT know books who-DISJ
'Ramu gave books and toys to someone. *I know the books to whom'

At the end of this sub-section, we have successfully excluded an IP-deletion account for Dravidian sluicing.

2.3 Could it be a case of stripping?

Before going further, we should check to make sure that the construction we are examining in Telugu is, in fact, a type of sluicing and not stripping (Hankamer 1979, Merchant 2005), also called bare argument ellipsis, e.g. *Suzanne plays cello, and Michael too*, where

everything in the second conjunct goes missing except for the single constituent *Michael*, the non-*wh*-phrase which is focus-moved.

There are two properties of stripping that distinguish it from sluicing (and the other ellipsis constructions like verb phrase ellipsis and noun phrase ellipsis) (Lobeck 1995). First, stripping is ungrammatical in embedded contexts (23), while sluicing is fine in this environment (24)

(23) *Suzanne plays cello, and I think that Michael too. STRIPPING

(24) Suzanne plays something, but I don't think she ever told me what. SLUICING

The sluicing construction in Telugu, too, can be embedded, as shown in (25)b, whereas the stripping structure cannot, as shown in (25)a.

(25) a. raamu oka pustakam konnaaDu. ravi kuuDaa *(konnaaDu) *pro* anukunnaanu
Ramu a book bought. Ravi too (bought) thought
'Ramu bought a book. *I thought Ravi too.'

b. ii vigraham 3rd century-loo ceyya-baDDadi. evari-dwaaraa-n-oo naaku
this statue 3rd century-IN made-PSV. who–BY-DISJ I-DAT
telusu ani anukunnaanu
know that thought
'This statue was made in the 3rd century. I thought that I know by who'

Second, stripping cannot occur before its antecedent, as illustrated in (26). This contrasts with sluicing which, as shown in (27), can precede its antecedent (as long as it does not command it).

(26) *Michael too, and Suzanne plays cello. STRIPPING

(27) I don't know what, but I'm sure Suzanne plays something. SLUICING

In Telugu, a sluice is also able to precede its antecedent, as shown in (28).

(28) eemiT-oo teliyadu kaani, raamu eed-oo vastu konnaaDu ani naaku telusu
what-DISJ know-not but, Ramu what-DISJ thing bought that I-DAT know
'I don't know what, but I know that Ramu bought something.'

Hankamer (2011) criticizes the analysis of Turkish sluicing in Ince (2012), pointing out that the seemingly embedded context is actually a root question which has undergone stripping, followed by the separate assertion *I don't know*. But this can't be a possible analysis in Telugu, because the *wh*-remnant can intervene between the matrix subject and predicate as shown in (29), thus ruling out an intonational aside.

(29) raamu eed-oo konnaaDu, kaani naaku eemiT-oo tel-iyadu.
Ramu what-DISJ bought, but I-DAT what-DISJ know-not
'Ramu bought something, but I don't know what.'

Therefore, we can conclude that what we see in Telugu is not an instance of stripping, because the sluicing construction can tolerate some distance and complexity in structure between the antecedent and the ellipsis site, whereas stripping has to be extremely local.

2.4 The source is not a deep cleft / pseudocleft

The source can't be a pseudocleft either, because the pivot of the pseudocleft always bears nominative case, it cannot bear any other case, as shown in (30). The sluice in Telugu shows obligatory case-matching, and can bear non-nominative case, as we saw in (11).

(30) a. *neenu koTT-in-a-vaaDu raamu-ni
I hit-PST-REL-3MS Ramu-ACC
'Who I hit is Ramu'

b. *neenu pustakamu icc-in-a-vaaDu raamu-ki
I book give-PST-REL-3MS Ramu-DAT
'Who I gave the book to is Ramu'

The pivot of the pseuodcleft can't be non-subject arguments. Objects/adjuncts cannot be pivots, as shown in (31)-(32). The *wh*-remnant of a sluice can be an object/adjunct.

(31) *neenu koTT-in-a-vaaDu raamu
I hit-PST-REL-3MS Ramu
'Who I hit is Ramu'

(32) *neenu bomma icc-in-a-vaaDu raamu
I toy give-PST-REL-3MS Ramu
'Who I gave the toy to is Ramu'

The verb in the Dravidian pseudocleft agrees with the pivot DP, as shown in (33).

(33) nannu koTT-in-a-vaaDu/vaallu raamu/pillalu
I-ACC hit-PST-REL-3MS/3P Ramu/kids
'Who hit me is Ramu/kids'

The differences between clefts and pseudoclefts in general are listed in (34), and they are true of Dravidian as well, as we see from the above data for pseudoclefts, and the properties of clefts in Telugu that will be presented in the next subsection. So we can draw a clear line between the two constructions —clefts and pseudoclefts in Dravidian. This is necessary because the two are sometimes quite similar on the surface, and we don't want to mix up the properties of clefts and pseudoclefts in our analysis of sluicing.

(34)

Pseudocleft / Deep cleft	*Cleft / Shallow cleft*
Verb agrees with the pivot DP	Verb bears default agreement
No case matching on the pivot	Case matching on the pivot
Pivot can only be an argument	Pivot is an argument or adjunct

2.5 The source is a cleft: CP deletion at work

Obligatory case-matching on the remnant[8], and a remnant that can be argument or adjunct, still leaves a reduced cleft structure as a possible source (also considering that cleft ques-

[8]This also eliminates a base-generation model of sluicing as in Chung et al. (1995), where the sluiced *wh*-phrase is directly merged in Spec CP, because it cannot account for case marking that would have to be assigned in a case-marking position or merged in a theta position with lexical case. For case-matching, the *wh*-phrase must be assigned case in a clause-internal position and then moved before deleting the rest of that clause.

tions are such a prevalent strategy in this language family) for sluicing, because the pivots of clefts can be variously case-marked in Dravidian, as shown in (36)[9]. In clefts in Telugu, an overt expletive is absent, and the copula is also null.

(35) raamu koTT-in-di ravi-ni
Ramu hit-PST-CLM Ravi-ACC
'It is Ravi that Ramu hit.'

(36) raamu pustakam is-tun-di ravi-ki
Ramu book give-CONT-CLM Ravi-DAT
'It is Ravi that Ramu is giving the book to.'

Reduced-cleft sources of sluicing are possibly formed by ellipsis of a CP constituent in underlying cleft constructions, as shown in (37). This source is clear in a sentence like (38), where the cleft marked verb is overt.

(37) ..., kaani [*pro*[$_{TopicP}$[$_{CP}$ ~~ramu t$_i$ konnadi~~]] [$_{FocusP}$ eemiT$_i$]-oo] naa-ku tel-iyadu
..., but what-DISJ I-DAT know-not
'Ramu bought something, but I don't know what (it is that Ramu bought).'

(38) a. raamu eed-oo konnaaDu, kaani konn-adi eemiT-oo naa-ku tel-iyadu.
Ramu what-DISJ bought, but bought-CLM what-DISJ I-DAT know-not
'Ramu bought something, but I don't know what (it is that Ramu) bought .'

b. ..., kaani [*pro*[$_{TopicP}$[$_{CP}$ ~~*pro*~~ t$_i$ konnadi]] [$_{FocusP}$ eemiT$_i$]-oo] naa-ku tel-iyadu
..., but what-DISJ I-DAT know-not

There are *wh*-in-situ languages, like Turkish, which also show case-matching on the *wh*-remnant, where a reduced-cleft source for sluicing can be discounted based on multiple sluicing. In Turkish, while multiple *wh*-remnants are possible in a sluiced structure, multiple *wh*-pivots are not possible in a cleft structure (Ince 2012), ruling out ellipsis in a cleft structure as the source. A number of languages, including English, do not allow multiple pivots in clefts. So multiple sluicing could be a good test to rule out a cleft source —if a language allows multiple sluicing, but not multiple pivots in clefts, then clefts can't be the source for the sluice.

Telugu allows multiple *wh*-remnants in the sluice, as illustrated in (39), with multiple arguments –(39)a, and, an argument plus an adjunct –(39)b.

(39) a. raamu evar-ik-oo eed-oo icceeDu, evar-ik-oo eed-oo naa-ku telusu.
Ramu who-DAT-DISJ what-DISJ gave, who-DAT-DISJ what-DISJ I-DAT know
'Ramu gave someone something, I know whom what.'

b. raamu ekkaD-oo eed-oo konnaaDu, ekkaD-oo eed-oo naa-ku telusu.
Ramu where-DISJ what-DISJ bought, where-DISJ what-DISJ I-DAT know
'Ramu bought something somewhere, I know what where.'

Does Telugu also allow multiple pivots in clefts? Controlling for non-interference of comma intonation, while keeping the focus intonation on the multiple cleft pivots is tricky, but it does seem like multiple pivots are allowed in clefts, as shown in (40).

[9]Here CLM = Cleft marker (3rd Person non-masculine agreement).

(40) a. raamu icc-in-di ravi-ki pustakam
Ramu give-PST-CLM Ravi-DAT book
'It is a book to Ravi that Ramu gave.'

b. raamu icc-in-di evari-ki eemiTi?
Ramu give-PST-CLM who-DAT what
'What is it to whom that Ramu gave?'

The properties of the sluicing remnants and cleft pivots match (more data will be illustrated in §3). In sluicing in Telugu, the *wh*-remnant obligatorily case-matches the correlate. In addition, not only argument, but also adjunct *wh*-phrases can be sluiced. Postposition pied-piping is obligatory in sluicing structures –Telugu obeys the 'Form Identity Generalisation II' of Merchant 2001, as shown in (41), and multiple sluicing is permitted.

(41) a. raamu pustakamu deeni pakka-noo daaceeDu. kaani deeni pakka-n-oo
Ramu book which next-DISJ hid. But which next-DISJ,
naaku teliyadu
I-DAT know-not
'Ramu hid the book next to something, but I don't know next to which.'

b. raamu ninna raattiraa pustakamu cadiveeDu. kaani eppuDu vara-k-oo
Ramu yesterday night book read. But when till-DISJ,
naaku teliyadu
I-DAT know-not
'Ramu read the book last last night. But I don't know till when.'

In cleft structures in Telugu, the pivot can bear various cases. Not only argument, but also adjunct pivots are possible. Postposition pied-piping is obligatory with the pivot, and multiple pivots are permitted. So clefting in Telugu has all the properties that make it a good candidate for the sluice source –the pivot can get various cases, the pivots can be multiple, and the elided material in the cleft clause forms a single constituent.

Cross-linguistically, apparently *wh*-in-situ, SOV languages have shown two patterns of sluicing, as given in (42). In the Japanese[10] type of *wh*-in-situ language, the characteristics of sluicing match with the characteristics of clefting, making a cleft-source for sluicing very likely. In the Turkish type of *wh*-in-situ language, the characteristics of sluicing are very unlike the characteristics of clefting, making the cleft-source unviable[11]. As shown in (42), Dravidian patterns with the Japanese type of languages, with the important exception of contrastive clefting/sluicing, which is discussed next.

[10]The data is more complex than is presented here for Japanese, which pertains to a certain subset of the Japanese sluicing data.

[11]Ince (2012) proposes that in Turkish sluicing structures, what causes and permits the overt raising of *wh*-expressions to the CP-domain is the checking of focus rather than *wh*-features, because in Turkish, *wh*-features are weak, as the language is *wh*-in-situ, and so it cannot be *wh*-features that result in the overt *wh*-movement of sluicing.

(42)

	Japanese		*Telugu*		*Turkish*	
	Cleft	Sluice	Cleft	Sluice	Cleft	Sluice
Case matching	Y	Y	Y	Y	N	Y
Multiple *wh*	Y	Y	Y	Y	N	Y
Adjunct *wh*	Y	Y	Y	Y	N	Y
Pied-piped post-positions	Y	Y	Y	Y	N	Y
Contrastive *else* modification			Y	N		

2.6 A puzzle: no contrast sluicing

In a kind of sluicing construction called contrast sluicing (Merchant 2001), the correlate is definite and the remnant asks for alternatives to the correlate. An example from English is given in (43). In Dravidian, contrast sluicing is not possible, as shown in (44). The vP is obligatory in the contrastive phrase in these constructions, as shown in (45).

(43) She met RINGO, but I don't know who else. Merchant (2001, 2008)

(44) a. raamu podduna oka iDli tinnaaDu. *kaani inkaa eemiT-oo naa-ku tel-iyadu
Ramu morning one idly ate. but still what-DISJ I-DAT know-not
'Intended: Ramu ate an idly in the morning, but I don't know what else. '

b. raamu siita-ki oka pustakam icceeDu, *kaani inkaa emiT-oo naa-ku
Ramu Sita-DAT one book gave, but still what-DISJ I-DAT
tel-iyadu
know-not
'Intended: Ramu gave Sita a book, but I don't know what else. '

(45) a. ... kaani raamu podduna tinnadi inkaa eemiT-oo naa-ku tel-iyadu
but Ramu morning ate-CLM still what-DISJ I-DAT know-not
'... but I don't know what else it is that Ramu ate in the morning. '

b. ... kaani raamu siita-ki iccindi inkaa emiT-oo naa-ku tel-iyadu
but Ramu Sita-DAT gave-CLM, still what-DISJ I-DAT know-not
'...I don't know what else it is that Ramu gave to Sita. '

Why is contrast sluicing not possible in Telugu, when regular and multiple sluicing are? Could the explanation be a cleft source? After all, clefts don't allow contrasting in English, as shown in (46).

(46) John ate an apple. *But I don't know what else it is that he ate.

So could a cleft source explain why a contrast sluice is not possible in Telugu? As it turns out, contrastive pivots are possible in clefts in Telugu, as shown in (60).

(47) raamu apple tinnaaDu. vaaDu tin-in-di inkaa eemiT-oo naaku teliyadu
Ramu apple ate. He eat-PST-CLM still what-DISJ I-DAT know-not
'Ramu ate an apple. I don't know what else it is that he ate.'

The puzzle therefore still remains unsolved. To be in a position to solve it, we first need to determine the structure of the Dravidian cleft that is the source for the sluice. This will be taken up in the next section.

At the end of this section we are forced to conclude that the only viable option for the source of sluicing in Dravidian is a reduced or truncated cleft. In the next section we get down to analyzing the exact structure of the copular cleft that feeds sluicing.

3. The structure of the sluicing cleft in Dravidian

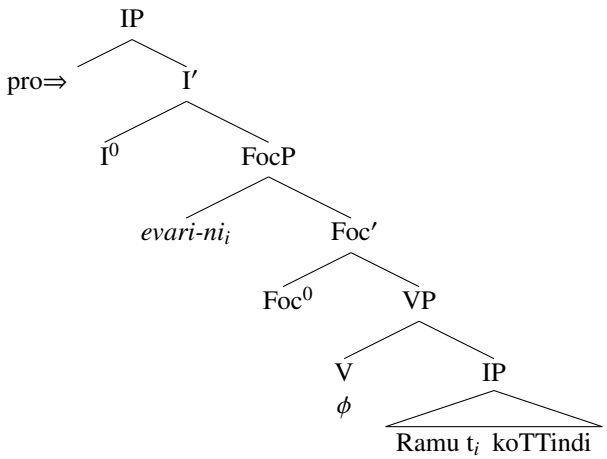
The source of the sluicing structure in Dravidian is the cleft, more precisely, the *wh*-cleft. A normal *wh*-cleft construction in Telugu is illustrated in (48)-(49). The neutral position for the cleft focus is at the end of the sentence, after the cleft-marked verb.

(48) raamu koTT-in-di evari-ni
Ramu hit-PST-CLM who-ACC
'Who is it that Ramu hit?'

(49) evari-ni raamu koTT-in-di
who-ACC Ramu hit-PST-CLM
'Who is it that Ramu hit?'

I follow Jayaseelan (1999, 2001) in analyzing cleft constructions in Dravidian as movement to focus positions[12] in IP, as shown in (50) —The verb 'be' takes a clausal complement; and a focused phrase from within this clausal complement is moved into the focus position above the VP headed by 'be'. The copula is null in Telugu; and since Dravidian has pro-drop, the subject position can be filled by *pro*.

(50)



As Jayaseelan & Amritavalli (2005) point out, normally, the movements to IP-internal topic and focus positions are clause internal, not long-distance [13]. There is no independent evidence for successive-cyclic movement of *wh*-phrases in any Dravidian language. A *wh*-phrase in an embedded clause cannot move to the focus position of the matrix clause, as illustrated with Telugu data in (51).

[12] A non-movement analysis of clefts is proposed for Malayalam clefts in Mohanan (1982).

[13] But long-distance topicalization, as they note, is always possible to clause-initial position in Dravidian. So in the default order in Telugu, shown in (48), the entire presuppositional cleft clause is long-distance topicalized to the left of the cleft pivot, which is in the focus position.

(51) a. *ravi [raamu t_i cuuseeDu ani] evari-ni$_i$ annaaDu ?
Ravi Ramu saw COMP who-ACC said
b. *ravi evari-ni$_i$ annaaDu [raamu t_i cuuseeDu ani] ?
Ravi who-ACC said Mary saw COMP
Intended: 'Who did Ravi say that Ramu saw?'

This presents a problem for the cleft analysis because the *wh*-phrase does move out of the embedded clause to the pivot position in the matrix clause. There is also an asymmetry of floating in short-distance clefts (clefts with pivots from matrix clauses) but lack of floating in long-distance clefts (clefts with pivots coming out of embedded clauses[14]).

They solve the two problems by analyzing the operations underlying short-distance and long-distance clefting as different, and showing that the operation that results in the long-distance cleft allows long-distance extraction but not floating, and the operation that results in the short-distance cleft allows floating but not long-distance extraction.

We will look at the properties and structure of long-distance and short-distance clefts next.

3.1 Short and long-distance clefts: IP clefts & CP clefts

Long-distance clefts are clefts whose pivots come out of embedded clauses, as shown in (52)a - (53)a. The Long-distance cleft does not allow floating, as shown in (52)b,c - (53)b,c.

(52) a. ravi naa-ku raamu koTTeeDu ani ceppin-(a)-di **ramesh-ni**
Ravi I-DAT Ramu hit that said-REL-CLM Ramesh-ACC
'It is Ramesh that Ravi told me that Ramu hit.'
b. * V **rameshni** naaku M koTTeeDu ani ceppindi
c. * V naaku **rameshni** M koTTeeDu ani ceppindi

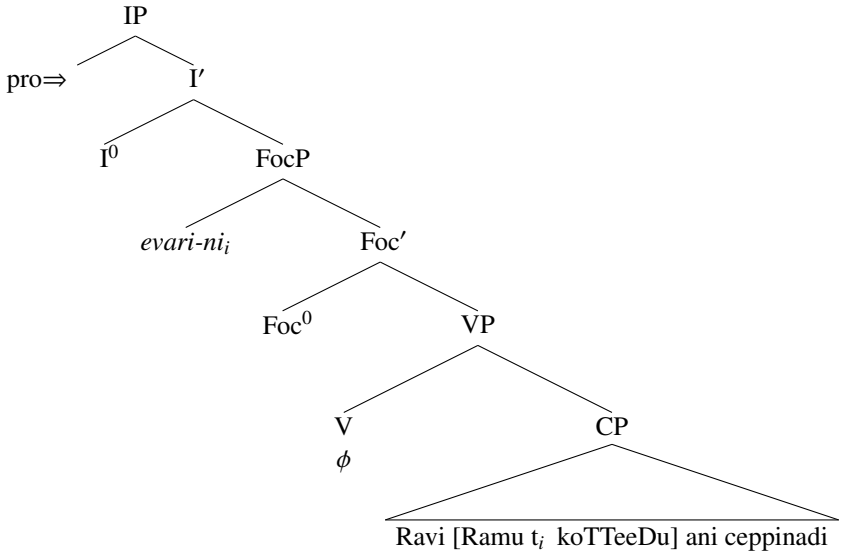
(53) a. ravi raamu koTTeeDu ani ceppin-(a)-di **evari-ni**
Ravi Ramu hit that said-REL-CLM who-ACC
'Who is it that Ravi said that Ramu hit?'
b. * V **evari-ni** M koTTeeDu ani ceppin-(a)-di
c. * V M **evari-ni** koTTeeDu ani ceppin-(a)-di

Jayaseelan & Amritavalli (2005) propose that the pivot moves out of the CP cleft clause in long-distance clefts by relativization[15], a well known long-distance movement, into the focus position of the (null) copular clause. Relativization uses an 'escape hatch' in the C-system to extract the phrase out of the relativized clause, as shown in (54). So like relativization, this clefting operation is also long-distance, is also island sensitive, and also does not allow floating.

[14]Crucially, even floating from the matrix part of the cleft clause is not possible in a long-distance cleft, as they mention, thus ruling out a 'finiteness' based solution, a possibility, considering that the cleft clause is nonfinite.

[15]They reanalyze the cleft marking, not as default agreement, but as a form of the relativizer.

(54) Jayaseelan and Amritavalli (2005) structure for the Dravidian long-distance cleft



On the other hand, in the short-distance cleft, with the pivot extracted from a matrix clause, the cleft focus can freely float into the cleft clause, as shown in (55)-(56). I again adopt the proposal in Jayaseelan (1999, 2001) that the effect of floating is created by the movement to topic positions in the copular clause, above the focus phrase, of the elements that appear to the left of the cleft focus.

(55) a. R. idi iccin-(a)-di **evari-ki**
R. this gave-REL-CLM who-DAT
'Who is it that R. gave this to?'
b. R. idi **evari-ki** iccin-(a)-di
c. R. **evari-ki** idi iccin-(a)-di
d. **evari-ki** R. idi iccin-(a)-di

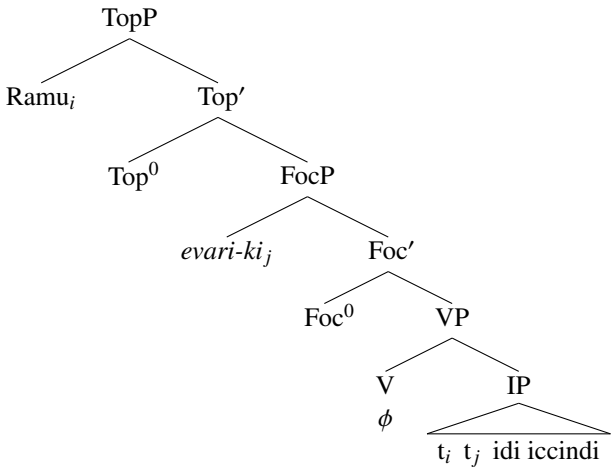
(56) a. R. idi iccin-(a)-di **Pavan-ki**
R. this gave-REL-CLM Pavan-DAT
'It is Pavan that R. gave this to?'
b. R. idi **Pavan-ki** iccin-(a)-di
c. R. **Pavan-ki** idi iccin-(a)-di
d. **Pavan-ki** R. idi iccin-(a)-di

According to Jayaseelan & Amritavalli (2005), this clefting operation does not use the relativization operation at all[16], it does not move the pivot phrase through any phase-edge location in the C-system, but instead, moves both the pivot phrase and any phrases that get topicalized directly from within the transparent cleft clause into the matrix clause, without any landing site in the C-domain (essentially it is extraction out of a nonfinite clause, which shows independent transparency to such operations.)

So in the short-distance cleft, the cleft clause does not project the C-domain at all, it is merely an IP. Its transparency for extraction in terms of the phase theory does not count as a violation because the cleft clause does not count as a phase, and therefore a matrix topic/focus probe can extract a phrase from inside this cleft, which is accessible to the next phase, as shown in (57). In long-distance clefts, however, as we saw above, the cleft clause is a CP, a phase boundary, and it does not allow for topic extraction.

[16]They provide some morphological evidence, through an indirect route, by bringing up another clefting structure in Malayalam which clearly does not have a relative marker. Their prediction then would be that this cleft structure will be 'transparent' to extraction, will allow floating, but will not allow long-distance clefting (extraction of the cleft pivot from within an embedded clause). This is borne out.

(57)



In conclusion, the long-distance cleft has a CP-layer, and employs a relativization mechanism to extract the pivot out of the opaque cleft clause. In contrast, the short-distance cleft has only an IP-layer, and does not need a relativization mechanism to extract the pivot, as it is transparent to the next phase. Therefore, floating is also possible in this non-phasal cleft.

We have so far established that the Dravidian sluice is a clefting structure, and we also saw that there are two kinds of clefts in Dravidian. But which of the two clefting strategies is the source of the Telugu sluice? We will find the answer to this next.

3.2 The Telugu sluice is a long-distance cleft

We are armed with one very clear test to differentiate short-distance and long-distance clefts –short-distance clefts allow floating, long-distance clefts prohibit floating. Applying this test to the sluicing cleft structure will tell us which of the two clefts it is. The Telugu sluice does not allow floating, as shown in (58). Topicalization of non-wh remnants is not possible, which it should be, if sluicing in Telugu is based on short-distance clefts, which allow topic extraction, and this happens before deletion of the cleft clause. Hence we conclude that it must be a long-distance cleft.

(58) *raamu evari-k-oo idi icceeDu, kaani raamu evari-k-oo naa-ku tel-iyadu.
Ramu who-DAT-DISJ this gave, but Ramu who-DAT-DISJ I-DAT know-not
'Intended: Ramu gave this to someone, but I don't know Ramu who to.'

3.3 No contrastive focus in sluices and long-distance clefts

We are finally in a position to solve the contrast sluicing puzzle that we raised in §2.6. We saw that contrast sluicing is not possible in Telugu, as given again in (59). However, contrastive pivots are possible in clefts in Telugu, as shown again in (60).

(59) raamu siita-ki bommalu icceeDu. *inkaa evari-k-oo naa-ku tel-iyadu
Ramu Sita-DAT toys gave. still who-DAT-DISJ I-DAT know-not
'Intended: Ramu gave toys to Sita. I don't know to who else.'

(60) raamu oka battaai tinnaaDu. vaaDu tinn-(a)-di inkaa eemiT-oo naa-ku tel-iyadu
Ramu one orange ate. He ate-REL-CLM still what-DISJ I-DAT know-not
'Ramu ate an orange. I don't know what else it is that he ate.'

Now that we know that the sluice is a long-distance cleft, we need to check for contrastive focus not in the short-distance cleft as above, but in the long-distance cleft, the source for the sluice. The long-distance cleft prohibits contrastive pivots, as shown in (61) - (63).

(61) *ravi naa-ku raamu koTTeeDu ani ceppin-(a)-di inkaa ramesh-ni
Ravi I-DAT Ramu hit that said-REL-CLM still Ramesh-ACC
'Intended: It is also Ramesh that Ravi told me that Ramu hit.'

(62) *ravi naa-ku raamu koTTeeDu ani ceppin-(a)-di ramesh-ni kuuDaa
Ravi I-DAT Ramu hit that said-REL-CLM Ramesh-ACC also
'Intended: It is also Ramesh that Ravi told me that Ramu hit.'

(63) *ravi raamu koTTeeDu ani ceppin-(a)-di inkaa evari-ni
Ravi Ramu hit that said-REL-CLM still who-ACC
'Intended: Who else is it that Ravi said that Ramu hit?'

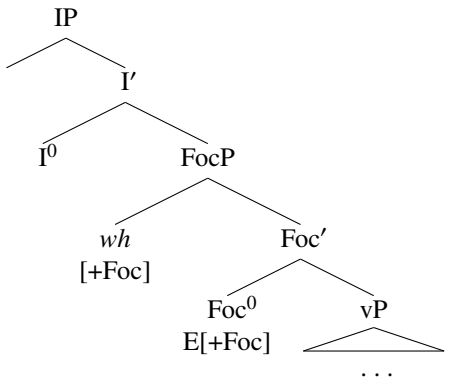
So the solution to the puzzle is very clear. Sluicing in Telugu does not allow contrastive *wh*-phrases because the source, long-distance clefts, do not allow contrastive pivots. But why don't long-distance clefts allow contrastive pivots, whereas short-distance clefts do? I leave this question open for future research.

3.4 The mechanics of sluicing in Dravidian

Merchant's (2001, 2006) technical implementation of sluicing involves a formal feature (called [E]), which has syntactic, semantic and phonological effects, that determine ellipsis. Merchant implements the syntactic restriction of sluicing to *wh*-phrases in a language like English, by giving [E] an uninterpretable [*wh*]-feature.

What is the specification of the [E]-feature in Telugu? Since we know now that the source of sluicing is the cleft, and the *wh*-remnant is a cleft pivot, the [E]-feature in Dravidian has to contain a [Foc]-feature, not unlike the specification in Hungarian (van Craenenbroeck & Liptak 2013), as shown in (64).

(64) The specification of the [E]-feature in Dravidian



3.5 The *wh*-/sluicing correlation

van Craenenbroeck & Liptak (2013) propose that the syntax of sluicing should track that of *wh*-movement in all languages, and formalize it as given in (65).

(65) THE WH/SLUICING-CORRELATION: The syntactic features that the [E]-feature has to check in a language L are identical to the strong features a *wh*-phrase has to check in a regular constituent question in L.

Going by this hypothesis, if the content of the [E] feature in Dravidian is [+Foc], then *wh*-phrases in Dravidian also have to check a [+Foc] feature in regular constituent questions. This indirectly lends support to the proposal in Jayaseelan (1999, 2001) that in Dravidian there is (partial) *wh*-movement to IP-internal Focus positions, and that Dravidian is not *wh*-in-situ.

van Craenenbroeck & Liptak (2013) also note that "The restriction to *wh*-phrases is no longer a reliable diagnostic for sluicing: the syntax of overt *wh*-movement in a language determines the syntactic properties of [E], which in turn determines what kind of remnants can occur in sluicing." They note that "if a language has overt movement of *wh*-phrases to Spec of FocP, it should also allow focus sluicing.", with a non-*wh*-phrase as remnant. Indeed, Telugu allows non-*wh* remnants. They can be referential NPs, as shown in (66), or adverbs, as shown in (67), or PPs, as shown in (68).

(66) raamu evari-n-oo koTTeeDu, neenu ravi-ni anukunnaanu
Ramu who-ACC–DISJ hit-PST, I Ravi-ACC thought
'Ramu hit someone, I thought Ravi.'

(67) raamu America velleeDu, neenu ninna ani anukunnaanu
Ramu America go-PST, I yesterday that thought
'Ramu went to America, I thought that yesterday.'

(68) raamu evari-too-noo velleeDu, naaku ravi-too ani teliyadu
Ramu who-with-DISJ go-PST, I-DAT Ravi-with that know-not
'Ramu went with someone, I didn't know that with Ravi.'

The indicative complementizer can also be one of the non-wh remnants as seen in (68), (67), and in (66b-c). These cannot be cases of stripping, because stripping cannot occur in embedded clauses.

But if the structure allows non-*wh*-elements to be remnants, does it mean that we can no longer call it sluicing? van Craenenbroek & Liptak (2013) point out that the restriction to *wh*-questions is actually not a reliable diagnostic of sluicing and they warn against being too construction or language specific: "A revealing example in this respect is the line of reasoning initiated by Jayaseelan (1990), who tries to reduce pseudogapping to VP-ellipsis. To the extent that this analysis is on the right track, it suggests that whatever properties set apart pseudogapping from VP-ellipsis (e.g. sensitivity to the Backwards Anaphora Constraint) is not a distinctive trait of VP-ellipsis and hence should not be used in the identification of this construction."

4. Conclusion

While Jayaseelan (1999, 2001) and Jayaseelan & Amritavalli (2005) have linked IP-internal focus constructions and cleft constructions in Dravidian and found evidence for two kinds of clefts in the language family –short and long-distance clefts, this paper establishes the link between sluicing and clefts in Dravidian.

All the properties of sluicing in Telugu can be assimilated to and according to this analysis fall out of the properties of long-distance clefts. The IP-internal Topic/Focus analysis also receives indirect support via the wh/sluicing correlation.

Given that the source of the Dravidian sluice is a cleft, and the antecedent is not a cleft, the identity requirement for sluicing cannot exactly be syntactic isomorphism. The implication this has for the identity requirement in ellipsis is similar to the implications of Potsdam (2007)'s pseudocleft analysis of Malagasy sluicing.

One question that remains unanswered is why the sluicing source is always a long-distance cleft and never a short-distance cleft in Telugu. I speculate that this has to do with the remnant requiring to land in the C-space to check features, before moving further. Hence the CP-domain always has to be projected in the cleft, making it a long-distance cleft.

Another interesting puzzle raised by a cleft source for sluicing in Telugu is Sprouting –a sub-type of sluicing, in which the remnant *wh*-phrase has no overt correlate in the antecedent (Chung *et al*. 1995). Sprouting is possible in Telugu, as shown in (69). But if there is no overt correlate for the cleft in the antecedent, how is the presuppositional cleft clause generated?

(69) a. raamu pustakam raaseeDu, kaani deeni-gurinc-oo naa-ku teliyadu
Ramu book write-PST, but what-about-DISJ I-DAT know-not
'Ramu wrote a book. But I don't know what about.'

b. raamu pustakam konnaaDu, kaani evari-koosam-oo naa-ku teliyadu
Ramu book buy-PST, but who-for-DISJ I-DAT know-not
'Ramu bought a book. But I don't know who for.'

Finally, it will be fruitful to extend this project to a comparison of Japanese & Dravidian vis-a-vis sluicing and clefting because clefts, in-situ focus, and sluicing/stripping in Japanese have also been proposed to share the same underlying structure by Hiraiwa & Ishihara (2012), but with a completely different syntactic structure than the one explored in this paper –they propose a monoclausal structure with focus movement to the CP domain for clefts and sluicing in Japanese, whereas we have pursued a biclausal structure with IP-internal focus movement for clefts and sluicing in Dravidian. It would be interesting to see how much of the Japanese data is amenable to an IP-internal Topic/Focus structure.

References

Adams, Perng Wang. 2004. The structure of sluicing in mandarin chinese. *University of Pennsylvania Working Papers in Linguistics* 10:2.

Chung, Sandra, William A. Ladusaw, & James McCloskey. 1995. Sluicing and logical form. *Natural language semantics* 3:239–282.

Gribanova, Vera. 2013. Copular clauses, clefts, and putative sluicing in uzbek. *Language* 89:830–882.

Hankamer, Jorge. 1979. *Constraints on deletion in syntax*. New York: Garland.

Hankamer, Jorge. 2011. Turkish pseudo-sluicing. In *West Coast Conference on Formal Linguistics (WCCFL)*.

Hiraiwa, Ken, & Shinichiro Ishihara. 2012. Syntactic metamorphosis: Clefts, sluicing, and in-situ focus in japanese. *Syntax* 15:142–180.

Ince, Atakan. 2012. Sluicing in turkish. *Sluicing: cross-linguistic explorations. Amsterdam: John Benjamins* .

Jayaseelan, KA, & R Amritavalli. 2005. Scrambling in the cleft construction in dravidian. *The free word order phenomenon: Its syntactic sources and diversity* 69:137.

Jayaseelan, Karattuparambil A. 1990. Incomplete vp deletion and gapping. *Linguistic Analysis* 20:64–81.

Jayaseelan, Karattuparambil A. 1999. A focus phrase above vp. In *Proceedings of the Nanzan GLOW*, volume 1, 195–212. Nanzan University. Nagoya.

Jayaseelan, Karattuparambil A. 2001. Questions and question-word incorporating quantifiers in malayalam. *Syntax* 4:63–93.

Karimi, Simin. 2003. Focus movement and uninterpretable features. *Formal approaches to function in grammar: In honor of Eloise Jelinek* 297–306.

Kizu, Mika. 1997. Sluicing in wh-in-situ languages. In *Proceedings of the Thirty-Third Regional Meeting of the Chicago Linguistic Society; Papers from the Main Session*, 231–244. Chicago Linguistic Society.

Lobeck, Anne C. 1995. *Ellipsis: Functional heads, licensing, and identification*. Oxford University Press Oxford.

Manetta, Emily. 2013. Copy theory in wh-in-situ languages: Sluicing in hindi-urdu. *Journal of South Asian Linguistics* 6.

Mathew, Rosmin. 2014. The syntactic effect of head movement: Wh and verb movement in malayalam. Doctoral dissertation, UiT The Arctic University of Norway.

Merchant, Jason. 1998. Pseudosluicing': elliptical clefts in japanese and english. *ZAS working papers in linguistics* 10:88–112.

Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands, and the theory of ellipsis*. Oxford University Press Oxford.

Merchant, Jason. 2005. Fragments and ellipsis. *Linguistics and philosophy* 27:661–738.

Merchant, Jason. 2006. Sluicing. *The Blackwell Companion to Syntax* 271–291.

Merchant, Jason. 2008. Variable island repair under ellipsis. *Topics in ellipsis* 132–153.

Mohanan, Karvanuur P. 1982. Grammatical relations and clause structure in malayalam. *The mental representation of grammatical relations* 504:589.

Paul, Ileana, & Eric Potsdam. 2012. Sluicing without wh-movement in malagasy. *Sluicing: Cross-linguistic perspectives, eds. Andrew Simpson and Jason Merchant* 164–182.

Potsdam, Eric. 2007. Malagasy sluicing and its consequences for the identity requirement on ellipsis. *Natural Language & Linguistic Theory* 25:577–613.

Simpson, Andrew, Arunima Choudhury, & Mythili Menon. 2013. Argument ellipsis and the licensing of covert nominals in bangla, hindi and malayalam. *Lingua* 134:103–128.

Subbārāo, Kārumūri V. 2012. *South asian languages: A syntactic typology*. Cambridge University Press.

Takahashi, Daiko. 2013. Argument ellipsis in Japanese and Malayalam. *Nanzan Linguistics* 9:173–192.

Toosarvandani, Maziar. 2008. Wh-movement and the syntax of sluicing. *Journal of Linguistics* 44:677–722.

Toosarvandani, Maziar. 2009. Ellipsis in farsi complex predicates. *Syntax* 12:60–92.

Toosarvandani, Maziar. to appear. Persian. In *The oxford handbook of ellipsis*, ed. Jeroen van Craenenbroeck & Tanja Temmerman.

Van Craenenbroeck, Jeroen. 2010. Invisible last resort: A note on clefts as the underlying source for sluicing. *Lingua* 120:1714–1726.

Van Craenenbroeck, Jeroen, & Anikó Lipták. 2013. What sluicing can do, what it can't, and in which language. *Diagnosing syntax* 502–536.

Vicente, Luis. 2008. Syntactic isomorphism and non-isomorphism under ellipsis. Ms., University of California, Santa Cruz.

Vicente, Luis. 2014. Sluicing and its subtypes. Ms., Universität Potsdam.

Wei, Ting-Chi. 2004. Predication and sluicing in mandarin chinese. *PhD diss., National Kaohsiung Normal University* .

Rahul Balusu
kodiguddu@gmail.com

Anticausatives in Sinhala: A View To The Middle*

John Beavers — The University of Texas at Austin
Cala Zubair — State University of New York at Buffalo

1. Introduction

"Middle voice verbs" (hence "middles") form several well-defined classes, including inherent reflexives (Kemmer 1993), where the sole argument is interpreted as an agent acting upon himself/herself, so-called "middle constructions" (Kemmer's 1993, 147-149 "Facilitative Middles"; Condoravdi 1989, Fagan 1992, Ackema & Schoorlemmer 1994, 2005, *inter alia*), where the the sole argument is a patient acted upon by an implicit agent on a generic or habitual reading, anticausatives (Kemmer's 1993: 142-147 "Spontaneous Middles"; Chierchia 2004, Koontz-Garboden 2009, Beavers & Zubair 2013), where the sole argument is a patient not necessarily acted upon by any other entity, and passives (Kemmer's 1993: 147-149 "Passive Middles"; Siewierska 1984, 162-185, Maldonado Soto 1992, 233-258), where the sole argument is a patient acted upon by an unexpressed agent and the reading is more episodic. These are illustrated for Bahasa Indonesian in (1), where each verb bears the *ber-* middle prefix (see Kemmer 1993, Kardana 2011, Beavers & Udayana 2016), save anticausatives, which bear *ter-*, in contrast with active *meN-* or unmarked forms:[1]

(1) a. *Ali ber-dandan.*
Ali MV-dress
'Ali dressed (himself).' (inherent reflexive of transitive *(men-)dandan*)

b. *Mobil itu ber-jual dengan mudah.*
car that MV-sell with easy
'The car sells easily.' (middle construction of transitive *(men-)jual*)

*We would like to thank Ashwini Deo, Itamar Francez, Andrew Koontz-Garboden, and I Nyoman Udayana, as well as the audience at FASAL 2015, for their feedback.

[1]The Indonesian data represents the Balinese dialect spoken in Bali and the Minagkabaunese dialect spoken in West Sumatra. Our Sinhala speakers speak Kandyan and Colombo dialects. The following abbreviations are used throughout the paper: 1=first person, 3=third person, ACC=accusative, AV=agent voice, CAUS=causative, DAT=dative, DEF=definite, INDF=indefinite, INF=infinitive, INST=instrumental, INV=involitive, EMPH=emphatic, MV=middle voice, NEG=negation, NPST=non-past tense, OV=object voice, PASS=passive voice, PL=plural, POST=postpositional case, PRT=participle, PST=past tense, REFL=reflexive, REL=relativizer, SG=singular, VOL=volitive.

c. *Pintu itu ter-buka.*
door that MV-open
'The door opened.' (anticausative of transitive *(mem-)buka*)

d. *Mobil itu ber-jual kemarin.*
car that MV-sell yesterday
'The car sold yesterday.' (passive middle of *(men-)jual*)

A fundamental question is what unifies all of these middles together. One common approach has focused on the syntactic unity of middles as involving detransitivization (Grimshaw 1982, Keyser & Roeper 1984, Ackema & Schoorlemmer 1994, Doron 2003, Reinhart & Siloni 2005, Alexiadou et al. 2006, Alexiadou 2010, Alexiadou & Doron 2012).[2] For example, Embick (2004) defines predicates like those in (1) as not projecting a base external argument, and thus the object must raise to subject position, a type of unaccusative syntax. However, there is debate on what the exact syntax of middles is, where some have suggested that inherent reflexive and middle constructions are unergative (see Keyser & Roeper 1984, Stroik 1992, Embick 2004, Ackema & Schoorlemmer 2005, Reinhart & Siloni 2005, Alexiadou & Schäfer 2014 for discussion). Others have suggested that middles are instead a notional category, i.e. a specific reading of independently attested constructions (Condoravdi 1989, Lekakou 2002, Fábregas & Putnam 2014), or a family of constructions (Reinhart 2002, Alexiadou & Doron 2012).

However, nearly all such approaches either ignore the semantics or take it to be heterogeneous (e.g. Alexiadou & Doron 2012). Kemmer offers a unified theory of middle semantics, proposing that the core semantics is "low distinguishability of participants", e.g. the verb's agent and patient are not distinguished clearly from one another, giving rise to a type of reflexive reading, which she calls "intrinsic to the lexical semantics of middle verbs" (p. 94). However, low distinguishability of participants does not easily extend to middle constructions and passive middles, which implicate distinct agents and patients, nor to anticausatives, which lack obvious lexical entailment of multiple thematic roles for their subjects. Kemmer thus generalizes low distinguishability of participants to "low elaboration of events" — separate subevents in the verb's meaning are not differentiated, where the subevent associated with the agent is conflated with that of the patient (inherent reflexives), left unspecified (middle constructions and passive middles), or not present (anticausatives). However, Kemmer is not clear on what types of low elaboration constitute the semantics of middles, nor how middles differ from non-specific expressions (e.g. indefinite pronouns). The question then is whether there is a true semantic or syntactic unity to all middles.

Colloquial Sinhala presents an extreme challenge in this regard: unlike relatively well-behaved Indonesian, *none* of the middles illustrated above are formally identical. As discussed by Beavers & Zubair (2013), anticausatives in Sinhala are overtly coded not by valence changing morphology of any sort, but by a morphological contrast in the verb stem that indicates volitive vs. involitive mood, otherwise indicating roughly volitional vs. non-

[2] In at least some middles, such as Kemmer's cognition middles, the base form takes three arguments and the middle form takes two, suggesting that the operation is more generally reduction of valence by one argument. We set these more general cases aside and focus on detransitivization here, though in principle a simple generalization of the analysis discussed here can extend to higher valences as well.

volitional action. In particular, in (2a,b) the same transitive verb can occur transitively in either the volitive or involitive form respectively, but the involitive (and not the volitive) also has an intransitive form with a nominative subject corresponding to an anticausative, as in (3) (Beavers & Zubair 2013, 3, (2)-(3)).[3]

(2) a. *Aruni Nimal-wə giluwa.*
Aruni Nimal-ACC drown.VOL.PST
'Aruni intentionally drowned Nimal.' (volitive transitive)

b. *Aruni atiŋ Nimal-wə giluna.*
Aruni POST Nimal-ACC drown.INV.PST
'Aruni accidentally drowned Nimal.' (involitive transitive)

(3) *Nimal giluna/*giluwa.*
Nimal drown.INV.PST/drown.VOL.PST
'Nimal drowned.' (involitive intransitive qua anticausative)

Furthermore, the form in (3) can take an accusative rather than nominative subject on a passive reading, thus realizing another type of middle, albeit indicated by subject case:[4]

(4) *Nimal-wə giluna/*giluwa.*
Nimal-ACC drown.INV.PST/drown.VOL.PST
'Nimal was drowned (by someone).' (passive middle; Beavers & Zubair 2013, 3, (4))

This might suggest that the involitive is the middle form in Sinhala. However, this does not extend directly to other middle types. Inherent reflexives are realized primarily by a combination of a participial verb form plus a volitivity-neutral light verb *gannə* 'take':

(5) *Mamə naa gatta.*
1SG bathe.PRT take.PST
'I bathed/had a shower.' (inherent reflexive; Chandralal 2010, 138, (62))

Finally, middle constructions are found in both volitive and involitive forms, as well as in the *gannə* light verb construction:

(6) *Meeka kaar-eka pahasuven vikunenəwa/vikunanəwa/viku gannəwa.*
These cars-INDF easily sell.INV.NPST/sell.VOL.NPST/sell.PRT take.NPST
'These cars sell easily.' (middle construction; see also Gair 1970, 70-71, 76)

The formal diversity of middle types (plus the semantic heterogeneity) argues against even a family of constructions analysis, since there is little family resemblance across subtypes.

[3]The morphological distinction between volitive and involitive verbs has to do with a combination of the place of articulation of the vowels of the verbal root plus the choice of thematic vowel (conditioned also by tense). Furthermore, while volitive verbs typically take nominative subjects, involitives assign a range of quirky cases to their subjects, contingent largely on verb class and semantics. The details are irrelevant here save where noted; see Beavers & Zubair (2010) for discussion.

[4]Sinhala is a *pro*-drop language and as such (4) has a reading as with an unexpressed subject and with the accusative DP as the object. However, as Beavers & Zubair (2013, 27-30) discuss, a passive middle reading is also possible and there is grammatical evidence that the accusative DP is the subject in this case. Here and below all such examples are intended only on this reading.

Nonetheless, we suggest that there is a unity to middles in Sinhala, building on the analysis of Sinhala anticausatives of Beavers & Zubair (2013) and its extension to Indonesian middles in Beavers & Udayana (2016). Following Beavers & Udayana, we suggest that middles represent a mismatch between the syntactic and semantic properties of the middle form: a semantically dyadic predicate (i.e. describing a relation between two individuals) is realized as syntactically monadic (taking just a surface subject argument) through syntactic valence reduction. While one of the two semantic arguments can be directly realized as the sole syntactic argument, the unrealized participant must be interpreted through some other means: either coreferential with the expressed argument, yielding a reflexive reading, or with disjoint reference, receiving an existential interpretation. However, while this operation can account for all Indonesian middles, in Sinhala there is a significant interaction between argument suppression and volitive mood that rules out certain middles being formed by this operation. In particular, as argued by Beavers & Zubair (2013), Sinhala volitive mood stems have grammaticalized a notion of agentivity of their subjects that forces verbs that undergo middle formation into the involitive mood since their sole arguments cannot be agents in the appropriate sense. But involitive mood is semantically incompatible with some middle interpretations, and in exactly these cases alternative forms in the language instead express the relevant meanings. Thus while there is little overt resemblance across Sinhala middles, there is a principled explanation for why the overt diversity exists.

Before we continue, a brief comment is in order on the nature of volitive and involitive mood in Sinhala that will be relevant below. Semantically, volitive verbs typically indicate volitional action, whereas involitive verbs indicate non-volitional action. However, these are only default readings. Volitionality *per se* is sometimes cancelable with volitives verbs:

(7) *Laməya piŋgaanəyə kæḍuwa, eet hitəla nemeyi.*
child plate broke.VOL.PST but intention without
'The child broke the plate unintentionally.' (Inman 1993, 98, (39))

However, as Beavers & Zubair (2013) discuss, if volitionality does not obtain there is a requirement that the subject have acted in some way, i.e. in (7) it cannot be that the child broke the vase through accidental neglect. The only non-action reading must be volitional non-action, e.g. in the (8) the causing action of not watering must have been deliberate:

(8) *Joon mal watərə nokərə nisaa, mal vinaashə-kəraa.*
John flower.PL water do.NEG because flower.PL destroy-do.VOL.PST
'Because John deliberately didn't water the flowers, he destroyed them.'

Thus the volitive requires action, volition, or both of its subject, a disjunction Beavers & Zubair (2013, 14) call "agentivity".

Similarly, involitives do not always require non-volitionality of their subjects. One case of a volitional reading is when the involitive is used to express ironic denial for interlocutors engaging in playful taunting, as in (9) where speaker B is saying something clearly false in response to what s/he perceives to be a stupid question by hearer A (Zubair 2008, Beavers & Zubair 2010), but is describing an action that is clearly volitional.

(9) *((B shows A pictures of Nuwara Eliya; A asks if B went there; B responds.))*
Ehe giye? Nææ, machang. Mamə atiŋ par-e hinganna-gen ewaa
there go.VOL.PST NEG dude 1SG POST street-LOC beggar-INSTR 3PL
horəkən keruna.
steal do.INV.PST
'Go there? No, dude. I stole [the pictures] from a beggar on the street.'

In sum, volitives require agent subjects, while involitives allow non-agents (see Inman 1993 and Beavers & Zubair 2010 for further discussion of the meaning of (non-)volitionality). In the following we first review basic properties of the various middles, and then outline Beavers & Zubair's (2013) analysis of anticausatives and volitives, which serves as a background for our analysis of other middles.

2. A Overview of Middle Constructions

The middles illustrated in (1) have various semantic and grammatical properties that distinguish them, and these tend to be relatively similar across languages. We illustrate some such properties with Indonesian. Indonesian has a distinction between two types of active voice — agent voice *meN-* and unmarked object voice forms — and passive *di-* forms:

(10) a. *Tono men-dandan Ali.*
Tono AV-dress Ali
'Tono dressed Ali.' (agent voice)

b. *Ali Tono dandan.*
Ali Tono OV.dress
'Tono dressed Ali.' (object voice)

c. *Ali di-dandan (oleh Tono).*
Ali PASS-dress by Tono
'Ali was dressed by Tono.' (passive)

That the middles in (1) are distinct from canonical actives in (10a,b) is evident from the fact that they take one core argument rather than two. That they are distinct from canonical passives in (10c) — and from each other — is motivated by their interpretation and modificational properties. Middle constructions and passive middles do not license *dengan sendirinya* 'by itself' modifiers, nor purposive modifiers, but they do entail that there was some external, unexpressed causer in the event, consistent with the subject being a patient but not a causer and there being an unexpressed (and syntactically inert) causer in the event:

(11) a. *#Mobil itu ber-jual dengan sendirinya*
car that MV-sell with REFL
#'The car sells by itself'

b. *[*Wanita itu*]$_i$ *ber-jual* [PRO$_{j/i}$ *untuk men-(t)erima komisi 10%*]
woman that MV-sell to AV-receive commission 10%
*'The woman sold to receive a 10% commission.'

c. *#Mobil itu ber-jua tapi tidak ada orang yang men-jual=nya.*
car that MV-sell but NEG exist man REL AV-sell=3SG
#'The car sold, but nobody sold it.'

The difference between middle constructions and passive middles is the modal vs. episodic interpretation. Inherent reflexives license *dengan sendirinya* and purposives, but do not entail external causation, consistent with the subject being both the causer and patient:

(12) a. *Gadis itu ber-dandan dengan sendirinya.*
girl that MV-dress with REFL
'The girl dressed by herself.'

b. [*Gadis itu*]$_i$ *ber-dandan* [PRO$_i$ *untuk meng-ikuti kontes kecantikan*].
girl that MV-dress to AV-join contest beauty
'The girl dressed (herself) to join the beauty contest.'

c. *Gadis itu ber-dandan tapi tidak ada orang yang men-dandan=nya*
girl that MV-dressed but NEG exist man REL AV-dress=3SG
'She dressed, but nobody dressed her.'

Anticausatives license *dengan sendirinya* 'by itself' (which we discuss further below) but not purposives, and do not entail external causation, thus describing changes-of-state of their subjects but making no commitment that there is any separate causer participant:[5]

(13) a. *Pasukan itu ter-pecah dengan sendirinya*
troop that MV-break with REFL
'The troop broke by itself.'

b. *[*Pintu itu*]$_i$ *ter-buka* [PRO$_{i/j}$ *untuk men-dapatkan hawa segar*].
door that MV-open to AV-get air fresh
'The door opened to allow fresh air.'

c. *Pasukan itu ter-pecah dua tapi tidak ada yang mem-ecah=nya.*
troop that MV-break two but NEG exist REL AV-break=3SG
'The troop broke into two but nobody/nothing broke them.'
(On intended reading)

Di- passives differ from all of these in not taking *dengan sendirinya* 'by itself', but entailing external causation and taking purposives, with the unexpressed causer as the controller:

(14) a. *#Kapal itu di-tambat dengan sendirinya*
boat that PASS-moor with REFL
#'The boat was moored by itself'

b. [*Orang itu*]$_i$ *di-jual* [PRO$_{j/*i}$ *untuk men-erima komisi 10%*].
man that PASS-sell to AV-receive commission 10%
'The man was sold to receive a 10% commission.' (e.g. sold into slavery)

[5] *Ter-* has a separate use marking a type of involitive passive, but we leave this interpretation aside here. Note that Indonesian is generally a causativizing language; and inchoatives are more often unmarked.

c. *#Mobil itu di-jual tapi tidak ada orang yang men-jual=nya.*
car that PASS-sell but NEG exist man REL AV-sell=3SG
#'The car (was) sold, but nobody sold it

Another difference between middles and canonical passives is verb class restrictions. Middle constructions are usually assumed to be restricted to verbs that entail a change-of-state (e.g. the Affectedness Constraint of Anderson 1979, Jaeggli 1986, Tenny 1992, Beavers 2008, *inter alia*) or to verbs with (potentially) agentive subjects (Ackema & Schoorlemmer 1994). Anticausatives on the other hand are typically limited to those caused change-of-state verbs that lack agentive entailments of their subjects, i.e. they take "effector" subjects neutral to agentivity (Guerssel et al. 1985, Haspelmath 1993, Levin & Rappaport Hovav 1995, Van Valin & Wilkins 1996, Reinhart 2000, 2002, Koontz-Garboden 2009). Inherent reflexive middles are usually found with specific subclasses of verbs that describe actions that are canonically or often performed on the self, such as bodily care and grooming verbs (Kemmer 1993, 53-70). Finally, while we are not aware of explicit claims of the limits of passive middles, consistent with Indonesian and the data discussed in Siewierska (1984, 162-185), Maldonado Soto (1992, 233-258), and Kemmer (1993, 147-149) these typically (though perhaps not exclusively) occur with change-of-state verbs. Canonical passives are typically unrestricted lexically. These distinctions justify that each class is grammatically and semantically distinct from the others, and from canonical passives. We now turn to Sinhala, starting with our earlier analysis of anticausatives and volitive mood.

3. The Semantic Nature of Anticausativization in Sinhala

Sinhala anticausatives pattern like those in Indonesian. First, the relevant roots are limited to effector subject verbs as in (15), which lack agentivity entailments for the subjects of their corresponding causative variants, e.g. allowing not just animate causers but also natural forces and instruments, something not true e.g. of *minimarannə* 'murder', as in (16).

(15) *marannə/mærennə* 'kill/die', *waṭannə/wæṭennə* 'drop/fall', *gilannə/gilennə* 'drown', *kaḍannə/kæḍennə*, 'break', *arannə/ærennə*, *lissannə/lissennə* 'slip'

(16) *Gaŋgə pusaa-wə mæruwa/*minimæruwa.*
river cat-ACC kill.VOL.PST/murder.VOL.PST
'The river killed/*murdered the cat.'

They also do not entail an external causer, or allow *ibeemə* 'by itself' or purposives:

(17) a. *Siri giluna, eet kawuruwat/kisivat eyaa-wə*
Siri drown.INV.PST but nobody/nothing 3SG-ACC
gileuwe nææ.
drown.VOL.CAUS.PST.EMPH NEG
'Siri drowned, but nobody/nothing caused him to drown.'

b. *Eewa okkomə ibeemə kæḍenəwa.*
3PL all by REFL break.INV.NPST
'They$_i$ all just break by themselves$_i$.' (Henadeerage 2002, 133, (27)-(28))

c. *[PRO$_i$ *kaarekə harigassannə*] *dæŋ plooṭərekə$_i$ issenəwa.*
car.DEF repair.VOL.INF now carburetor raise.INV.NPST
*'The carburetor$_i$ rises now [PRO$_i$ to repair the car].' (Gair 1990, 35–36)

Thus Sinhala anticausatives share the same essential properties as those in Indonesian.

Koontz-Garboden (2009), examining Spanish anticausatives formed by "reflexive" *se*, argues for a unified analysis of anticausatives and reflexives that explains these facts, whereby both are derived by coidentifying the subject and object of a base transitive verb:

(18) $[\![se]\!] = \lambda\mathscr{R}\lambda x[\mathscr{R}(x,x)]$

Crucially, (18) has different outputs for different verb classes. Verbs like Spanish *asesinar* 'assassinate' take agent subjects while verbs like *romper* 'break' take an effector subject:

(19) a. $[\![asesinar]\!] = \lambda y\lambda x\lambda e\exists v[agent'(x,v) \wedge cause'(v,e) \wedge result'(y,e,\mathbf{dead}')]$
b. $[\![romper]\!] = \lambda y\lambda x\lambda e\exists v[effector'(x,v) \wedge cause'(v,e) \wedge result'(y,e,\mathbf{broken}')]$

With *se* (19a) forms a canonical "agent act on self" reflexive and (19b) an anticausative:

(20) a. *El senador se asesinó.*
the senator REFL assassinated.3SG
'The senator assassinated himself.' $\exists e\exists v[agent'(\mathbf{s},v) \wedge cause'(v,e) \wedge result'(\mathbf{s},e,\mathbf{dead}')]$
b. *El vaso se rompió.*
the cup REFL broke
'The cup broke.' $\exists e\exists v[effector'(\mathbf{c},v) \wedge cause'(v,e) \wedge result'(\mathbf{c},e,\mathbf{broken}')]$

This analysis explains the lack of external causer entailments (since the patient is the causer), the non-agentive causer restriction (since the relevant reading only arises with effector-subject verbs), and why anticausatives take *by itself* type modifiers, which generally only occur with verbs with explicit causer subjects (as per Chierchia 2004, Koontz-Garboden 2009) but not statives or unergatives, since on a reflexive analysis of anticausatives the subject is a causer.[6] Finally, since effectors in context could be interpreted as agents, this analysis predicts that anticausatives could license a purposive modifier on an "agent act on self" reading, something possible in Spanish (Koontz-Garboden 2009, 100, (52a)):

(21) *aquel día ... cuando Phil$_i$ se ahogó* [*para* PRO$_i$ *salvar-le la vida a Jim*]
that day when Phil REFL drowned for save-3SG the life DAT Jim
'And on that day ... when Phil drowned himself to save Jim's life...'

Thus the reflexivization analysis of anticausatives accounts for all of the relevant properties.

But as Beavers & Zubair (2013) point out, this analysis cannot be extended to Sinhala directly. First, unlike Spanish, agent-subject verbs comparable to Spanish *asesinarse* in (20a) have no detransitivzed forms, even on an "agent act on self" reading. Second, Sinhala purposives are categorically ruled out with anticausatives, even when volitionality is not at issue, unlike Spanish. This is illustrated by the fact that even on uses of the involitive that can have a volitional subject — such as ironic denial uses — purposives are unacceptable (data based on Beavers & Zubair 2013, 23, (41)):

[6]The degree to which *by itself* modifiers do occur with non-causative verbs a sufficiently rich context is required to establish that the subject is also a causer; see Beavers & Koontz-Garboden (2013).

(22) *((Mary's mother dies in a car wreck; A asks if it was an accident; B responds.))*

**Naeae, machang.* [PRO$_i$ *minihageŋ gæləwennə*], *eyaa$_i$ mæruna.*
no dude husband.INST escape.VOL.INF 3SG die.INV.PST
'No, dude. She died to escape her husband.'

Third, as noted in §1, anticausatives cannot occur in the volitive mood, even on an "agent act on self" reading. Nothing about a reflexive analysis explains these facts. Intuitively, the problem is that the forms that do not permit anticausatives — volitive stems, 'murder' verbs, and clauses with purposive modifiers — all require their subjects to be agents. Perhaps there is an additional constraint on Sinhala anticausativization that requires the subject to *not* be an agent. However, nothing precludes agentivity semantically, e.g. if it is clearly established that someone agentively acted upon themselves, then a clause headed by an anticausative verb is still necessarily true (Beavers & Zubair 2013, 27, (52)):

(23) *#Joon eyaa-wə-mə giluwa, hæbai eyaa gilune nææ.*
John 3SG-ACC-REFL drown.VOL.PST, but 3SG drown.INV.PST.EMPH NEG
'John drowned himself, but he didn't drown.'

Thus anticausatives are not *semantically* non-agentive. They just reject *grammatical*, *morphological*, or *lexical* contexts explicitly encoding agentivity.

A final problem with a reflexivization analysis is accusative subject anticausatives as in (24a), which occur with the same verbs in (15) but *require* external causation as in (24b), also resisting *ibeemə* 'by itself' modification as in (24a) and purposives as in (24c):

(24) a. *Meeri-wə (#ibeemə) giluna.*
Mary-ACC by REFL drown.INV.PST
'Mary drowned.'

b. *#Eyaa-wə lissuna, eet kawuruwat eyaa-wə lisseuwe nææ.*
3SG-ACC slip.INV.PST but nobody 3SG-ACC push.VOL.CAUS.PST.EMPH NEG
'She fell, but nobody pushed her.'

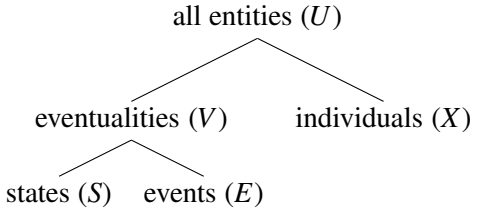
c. *[PRO$_{i/j}$ *Rakshana salli gannə*], *Meeri-wə$_i$ giluna.*
insurance money take.VOL.INF Mary-ACC drown.INV.PST
*'Mary$_i$ drowned [PRO$_{i/j}$ to collect the insurance money].'

This is clearly not reflexive, being more like a passive. What explains these properties?

The insight of Beavers & Zubair (2013) is that in canonical Sinhala anticausatives — nominative subject anticausatives — the patient is indeed interpreted reflexively as the causer, but it is a property or disposition of the patient that caused the change rather than an event it participated in. For example, in *The vase broke* — assuming no external causers or anthropomorphism — the reading is that something about the vase (e.g. a weakness in its structural integrity) lead to its breaking (see also Levin & Rappaport Hovav's 1995, 91-92 internal causation, and Prior et al. 1982 and Copley & Wolff 2014 on dispositions as causers). This differs from agentive causation, where some action by the causer lead to the

change (assuming volitional non-action is eventive, consistent with descriptions of such events licensing progressive aspect as in *John/*the statue is standing still* as per Dowty 1979, perhaps due to having stages à la Landman 1992). Thus non-agentive causation involves a causing state and agentive causation a causing event. Beavers & Zubair formalize this via the following sortal typology (building on Chierchia 2004, 37):

(25)



Agentive, non-agentive, and effector causers reflect the causing event sort as follows:[7]

(26) a. Agentive causer - *causer′* participant of causing *event* (in E).
b. Non-agentive causer - *causer′* participant of causing *state* (in S).
c. Effector - *causer′* participant of causing *eventuality* (in V) or *individual* (in X).

A key property of this analysis is that event vs. state causation is not just a truth conditional contrast but also a formal one, i.e. encoded in the *sort* of the causing eventuality.

This opens up the possibility that agentivity in some languages is grammaticalized, which Beavers & Zubair (2013) suggest is the case in Sinhala. In particular, they propose that causatives that require agent subjects take subjects representing causing events in E, while causatives that take effector subjects take a maximally general individual in U:

(27) a. $[\![\textit{minimara-}]\!] = \lambda y \lambda v \in E \lambda e[\textit{cause}'(v,e) \wedge \textit{result}'(y,e,\textbf{dead}')]$
b. $[\![\textit{kaḍa-}]\!] = \lambda y \lambda x \in U \lambda e[\textit{cause}'(x,e) \wedge \textit{result}'(y,e,\textbf{broken}')]$

Subject DPs denote Generalized Quantifiers (with event variables), supplying their VP argument with a causing eventuality in V with the DP's informal referent as the *causer′*:

(28) $[\![\textit{John}]\!] = \lambda P \lambda e \exists v \in V[\textit{causer}'(\mathbf{j},v) \wedge P(v,e)]$
"John is the causer of eventuality v that caused event e described by P."

Combining this interpretation of *John* with a *minimara-* 'murder' VP resolves the eventuality introduced by *John* to an event in E, thus requiring agentivity, as in (29a). Combining it with a *kada-* VP resolves the *cause′* introduced by the verb to an eventuality in V as in (29b), which could be interpreted in context as reflecting agentive ($v \in E$) or non-agentive ($v \in S$) causation. (The other effects of mood, and tense, are ignored here.)

(29) a. $[\![\textit{John Siri-wə minimæruwa}]\!]$ ('John murdered Siri')
$= \lambda e \exists v \in E[\textit{causer}'(\mathbf{j},v) \wedge \textit{cause}'(v,e) \wedge \textit{result}'(\mathbf{s},e,\textbf{dead}')]$
b. $[\![\textit{John piŋgaanəyə kaḍuwa}]\!]$ ('John broke the plate.')
$= \lambda e \exists v \in V[\textit{causer}'(\mathbf{j},v) \wedge \textit{cause}'(v,e) \wedge \textit{result}'(\mathbf{p},e,\textbf{broken}')]$

[7] Since agentivity is analyzed via causing eventuality sort, the role *causer′* is hence used for all causers.

On the basis of this, the agent-restriction of volitive stems can be stated as follows: they require the subject of the predicate they occur with to be an event, as in (30), which has no effect on 'murder'-type verbs but will have an effect on 'break'-type verbs.

(30) $[\![+\emptyset_{vol}]\!] = \lambda P \lambda x_1 \ldots \lambda x_n \lambda v \in E \lambda e[P(x_1, \ldots x_n, v, e)]$

a. $[\![minimara{+}\emptyset_{vol}]\!] = \lambda y \lambda v \in E \lambda e[cause'(v,e) \wedge result'(y,e,\mathbf{dead}')]$

b. $[\![kaḍa{+}\emptyset_{vol}]\!] = \lambda y \lambda v \in E \lambda e[cause'(v,e) \wedge result'(y,e,\mathbf{broken}')]$

Purposives can be given a similar analysis, applying to VPs and requiring an event subject:

(31) [[PRO *rakshana salli gannə*]] ('PRO to collect the insurance money')
$= \lambda P \lambda v \in U_E \lambda e \exists e'[P(v,e) \wedge collect'(\text{PRO}, \mathbf{r}, e') \wedge in\_order\_that'(e,e')]$

Thus agentivity is partly a formal, grammaticalized property in Sinhala.

This offers an explanation for the properties of Sinhala anticausatives. Beavers & Zubair (2013, 31, (62)) suggest that Sinhala anticausativization is more general than reflexivization, representing an operation which strips a causer from the verb's argument structure but preserves it as part of its truth conditional content, analyzed as saturation by an open variable in X, underlined for expository purposives (see also Kaufmann 2007, Piñón 2012):

(32) **Causer Suppression:** $[\![+\emptyset_{CS}]\!] = \lambda P \lambda y \lambda e[P(y, \underline{x}, e) \wedge \underline{x} \in X]$
Precondition: $\forall x' \forall y' \forall e'[P(y', x', e') \rightarrow cause'(x', e')]$

There are two ways of interpreting the open variable vis-a-vis the expressed argument:[8]

(33) $[\![kaḍa{+}\emptyset_{CS}]\!] = \lambda y \lambda e[cause'(\underline{x}, e) \wedge result'(y,e,\mathbf{broken}') \wedge \underline{x} \in X]$

a. Causer is ***co-referential*** with the patient:
$\lambda y \lambda e[cause'(y,e) \wedge result'(y,e,\mathbf{broken}') \wedge y \in X]$

b. Causer is ***not co-referential*** with the patient (i.e. $\exists$-bound):
$\lambda y \lambda e \exists x[cause'(x,e) \wedge result'(y,e,\mathbf{broken}') \wedge x \in X]$

The two interpretations correspond to nominative and accusative subject anticausatives respectively.[9] Crucially, the resulting verb forms take patient subjects, which are typed as individuals in X. This means Sinhala anticausatives are incompatible with any constructions such as volitive mood or purposives that require the subject to be in E, as well as deriving that 'murder'-type verbs will not permit anticausativization since their subjects are also in E. Conversely, the involitive *does* allow individual causers (since it imposes no constraints on its subjects). This crucially predicts that anticausatives will only allow

[8]The causer is typed as an individual in X, which Beavers & Zubair (2013, 37, (75)) assume is compatible with an analysis of causation as a relation between events by the meaning postulate in (i) that equates it with a causing event in V, thus ensuring neutrality to agentivity.

(i) $\forall x \in X \forall e \in V[[cause'(x,e) \wedge \ldots] \leftrightarrow \exists v \in V[causer'(x,v) \wedge cause'(v,e) \wedge \ldots]]$

An alternative would be that the subjects of effector subject verbs are causing eventualities in V. However, this would preclude the reflexive interpretation in (33a).

[9]Following Beavers & Zubair (2010, 2014) we assume accusative has a use as a semantic case indicating a patient acted on by an external causer (cp. the analysis of accusative of Wunderlich 1997), whereas nominative is checked structurally, e.g. in Spec,TP (Chou & Hettiarachchi to appear).

involitive stems. In sum, volitive mood is grammatically and semantically agentive, while anticausatives are semantically unspecified for agentivity (though still causative) but grammatically resistant to agentivity, predicting that they can occur in pragmatic contexts in which agentivity does or does not obtain but not grammatical contexts. We next show how this analysis can in principle be extended to other middles, focusing on Indonesian before returning to the more complex case of Sinhala.

4. Analyzing Other Middle Types

The key ingredient of Causer Suppression regarding argument structure is that the underlying verbal predicate is relational but the output is grammatically intransitive, with the suppressed argument interpreted in some other way. If anticausatives are a type of middle, then the question arises of whether this analysis is applicable to other middles as well. Beavers & Udayana (2016) propose exactly this for Indonesian middles. On the simplest extension, *ber/ter-* just reflect overt argument suppression as with Sinhala $+\emptyset_{CS}$, with *ter-* restricted to effector subject verbs and *ber-* the elsewhere case:[10]

(34) $[\![\textit{ter/ber-}]\!] = \lambda P \lambda y \lambda e[P(y, \underline{x}, e)]$

With both inherent reflexives as in (1a)/(12) and anticausatives as in (1c)/(13) the reading is reflexive, with the difference being that anticausatives arise with effector subject verbs and inherent reflexives with certain subclasses of agent subject verbs (those that reflect bodily grooming and other event types whose canonical association is "agent act on self"):

(35) a. $[\![\textit{dandan}]\!] = \lambda y \lambda x \lambda e \exists v \in E[\textit{causer}'(x,v) \wedge \textit{cause}'(v,e) \wedge \textit{result}'(y,e,\textbf{dressed}')]$

b. $[\![\textit{ber-dandan}]\!] = \lambda y \lambda e \exists v \in E[\textit{causer}'(y,v) \wedge \textit{cause}'(v,e) \wedge \textit{result}'(y,e,\textbf{dressed}')]$

(36) a. $[\![\textit{pecah}]\!] = \lambda y \lambda x \lambda e \exists v \in V[\textit{causer}'(x,v) \wedge \textit{cause}'(v,e) \wedge \textit{result}'(y,e,\textbf{broken}')]$

b. $[\![\textit{ter-pecah}]\!] = \lambda y \lambda e \exists v \in V[\textit{causer}'(y,v) \wedge \textit{cause}'(v,e) \wedge \textit{result}'(y,e,\textbf{broken}')]$

Conversely, middle constructions and passive middle interpretations as in (11) arise from binding off the suppressed argument, differing in that passive middles have existential quantification over the suppressed argument and an episodic reading, while middle constructions reflect either a generic binding of the suppressed argument (à la Condoravdi 1989) or existential binding with the entire predicate embedded under a covert generic modal G (roughly in the spirit of Lekakou 2002, 2006), the latter illustrated here for (1b,d):

(37) a. $[\![\textit{jual}]\!] = \lambda y \lambda x \lambda e \exists v \in V[\textit{causer}'(x,v) \wedge \textit{cause}'(v,e) \wedge \textit{result}'(y,e,\textbf{sold}')]$

b. $[\![\textit{ber-jual}]\!] = \lambda y \lambda e \exists x \exists v \in V[\textit{causer}'(x,v) \wedge \textit{cause}'(v,e) \wedge \textit{result}'(y,e,\textbf{sold}')]$

i. Middle construction: $\mathrm{G}(\exists e \exists x \exists v \in V[\textit{causer}'(x,v) \wedge \textit{cause}'(v,e) \wedge \textit{result}'(\textbf{car}',e,\textbf{sold}')])$

ii. Passive middle: $\exists e \exists x \exists v \in V[\textit{causer}'(x,v) \wedge \textit{cause}'(v,e) \wedge \textit{result}'(\textbf{car}',e,\textbf{sold}')]$

[10] Given that in Indonesian there appears to be no grammaticalized agentivity, we assume that subjects are individuals related thematically to appropriate causing events (and we ignore conditions on the suppressed argument being a causer, since as Beavers & Udayana (2016) show suppression can apply other arguments as well, though these data are not relevant for present purposes).

Thus all types of middles in (1) are amenable to the same core analysis, with the different subtypes arising from a combination of how the suppressed argument is interpreted, the root class suppression is applied to, and the modal interpretation of the predicate. The question is whether Sinhala middles are amenable to this analysis. As noted above, Sinhala's middles are far more heterogeneous, arguing against a unified analysis as per Indonesian. However, we suggest that the extension to Indonesian middles applies equally well to Sinhala, but there is a significant interaction with (in)volitive mood that predicts where the heterogeneity occurs, suggesting a principled core to middles despite the heterogeneity.

5. Sinhala Inherent Reflexives - Principled Limits on Causer Suppression

We first consider reflexive middles, which include anticausatives and inherent reflexives. As discussed above, of course, anticausatives are derived from the proposed Causer Suppression operation of Sinhala. But what about inherent reflexives? Given that bodily care verbs are generally agentive, the expectation is that Causer Suppression will not apply to them, if agentivity is grammaticalized as a type-theoretic constraint on their subjects as it is with *minimarannə* 'murder'. Surprisingly, however, there is in fact an intransitive involitive form with such verbs which has an agentive, reflexive reading, not entailing external causation, as in (38) (acceptable in a context where the subject is bathing a flailing toddler with water splashing around, and ends up washing himself).

(38) *Nimal sedhuna/næwuna, (eet kawuruwat/kisivat eyaa-wə*
Nimal wash.INV.PST/bathe.INV.PST but nobody/nothing 3.SG-ACC
sedheuwe/næweuwe nææ).
washed.VOL.CAUS.PST.EMPH/bathed.VOL.CAUS.PST.EMPH NEG
'Nimal accidentally washed/bathed, but nobody/nothing washed/bathed him.'

This suggests that Causer Suppression is possible with at least some agent subject verbs. These forms also allow *ibeemə* 'by itself', as expected if they are reflexive causatives (however, purposive modification is out since these are semantically non-volitional; see below):

(39) *Nimal ibeemə sedhuna/næwuna.*
Nimal by REFL wash.INV.PST/bathe.INV.PST
'Nimal accidentally bathed by himself.'

Still further evidence that (38) is derived via Causer Suppression comes from the fact that in addition to the nominative subjects as in (38) these forms also allow accusative subjects, crucially on the passive-type reading wherein there is necessarily an external causer, thus also rejecting *ibeemə*:

(40) *Nimal-wə (#ibeemə) sedhuna/næwuna, (#eet kawuruwat eyaa-wə*
Nimal-ACC by REFL wash.INV.PST/bathe.INV.PST but nobody 3.SG-ACC
sedheuwe/næweuwe nææ).
washed.VOL.CAUS.PST.EMPH/bathed.VOL.CAUS.PST.EMPH NEG
'Nimal got washed/bathed (#by himself), #but nobody washed/bathed him.'

The existence of these forms suggests that Causer Suppression is possible. How could this be, given that 'murder'-type verbs do not allow this?

In fact it is not surprising that at least some agent-subject verbs allow Causer Suppression: nothing prevents an agent-subject verb from taking a subject in U rather than in E, with agentivity ensured by some other means, e.g. as a lexical entailment deriving from the specific result state à la Beavers & Koontz-Garboden (2012) (e.g. part of the content of **bathed′** in (41)), and thereby being amenable to Causer Suppression:

(41) a. $[\![naa\text{-}]\!] = \lambda y \lambda v \in U \lambda e[cause'(v,e) \wedge result'(y,e,\mathbf{bathed'})]$

b. $[\![naa{+}\emptyset_{CS}]\!] = \lambda y \lambda e[cause'(\underline{x},e) \wedge result'(y,e,\mathbf{bathed'})]$

The output in (41b) derives a reflexive reading for (38) and passive reading of (40):

(42) a. $\exists e[cause'(\mathbf{nimal'},e) \wedge result'(\mathbf{nimal'},e,\mathbf{bathed'})]$

b. $\exists x \exists e[cause'(x,e) \wedge result'(\mathbf{nimal'},e,\mathbf{bathed'})]$

That bodily grooming verbs behave like a distinct class among agentive verbs is also not surprising, as this is the case in other languages as well (e.g. in English they are reflexive with no reflexive pronoun, and in languages in which the base form is intransitive rather than transitive the transitive is derived via causativization but on an "antireflexivization" reading; Krejci 2012). Furthermore, the existence of these forms justifies that the Causer Suppression analysis of anticausatives as reflexives is plausible, since these data independently demonstrate that reflexivization is a possible interpretation for this operation.

However, there is a crucial limitation with this understanding of Sinhala inherent reflexives, namely that these middles are involitive. This is as expected given the discussion above (since Causer Suppression produces forms that can only be involitive). But it does mean that the reading is therefore necessarily non-volitional (save for ironic denial readings), something borne out by the fact that they do not permit modifiers indicating volition:

(43) *Nimal (*hitəla) sedhuna/næwuna.*
Nimal deliberately wash.INV.PST/bathe.INV.PST
'Nimal accidentally washed/bathed (*deliberately).'

The question arises of how one would express the presumably more canonical inherent reflexive meaning of volitional self-action. There appear to be two alternative means.

First, as discussed by Chandralal (2010, 136-139), and consistent with our informants, inherent reflexives are most canonically expressed via a volitivity-neutral light verb *gannə* 'take' combined with a participial form of the verbal root (see also Gair 1970, 123).

(44) *Mamə sedhaa/naa gatta.*
1SG washed.PRT/bathe.PRT take.PST
'I washed/bathed.'

This form (unlike the Causer Suppressed form) has all of the canonical properties of inherent reflexives, e.g. a volitional reading is possible, as are purposives and *ibeemə* 'by itself', and no external causation is entailed:

(45) a. *Mamə hitəla sedhaa/naa gatta.*
1SG deliberately washed.PRT/bathe.PRT take.PST
'I deliberately washed/bathed.'

b. *Mamə$_i$* [PRO$_i$ *saadaya-ṭə yannə*] *sedhaa/naa* *gatta.*
1SG party-DAT go.VOL.INF washed.PRT/bathe.PRT take.PST
'I washed/bathed to go to the party.'

c. *Mamə ibeemə* *sedhaa/naa* *gatta.*
1SG by REFL washed.PRT/bathe.PRT take.PST
'I washed/bathed by myself.'

d. *Mamə sedhaa/naa* *gatta,* *eet kawuruwat/kisivat mamə-wə*
1SG washed.PRT/bathe.PRT take.PST, but nobody/nothing 1.SG-ACC
sedheuwe/næweuwe *nææ.*
washed.VOL.CAUS.PST.EMPH/bathed.VOL.CAUS.PST.EMPH NEG
'I washed/bathed, but nobody/nothing washed/bathed me.'

The existence of a separate form for expressing this meaning has an obvious functional motivation: the light verb is not subject to the constraints on Causer Suppression that necessarily generate involitives, and thus permits volitional readings, filling in this gap in the paradigm. Thus while there is disunity in the expression of inherent reflexive middles, it is a principled disunity given the Sinhala-specific constraints limiting the use of the otherwise cross-linguistically "canonical" way of deriving middles.

There is also a second expression for a volitional inherent reflexive, namely a bare *volitive* form of the verb that also has a reflexive reading and canonical properties:

(46) a. *Nimal (hitəla)* *sedhuwa/næwuwa.*
Nimal deliberately washed.VOL.PST/bathe.VOL.PST
'Nimal deliberately washed/bathed.'

b. *Nimal$_i$* [PRO$_i$ *saadaya-ṭə yannə*] *sedhuwa/næwuwa.*
Nimal party-DAT go.VOL.INF washed.VOL.PST/bathe.VOL.PST
'Nimal washed/bathed to go to the party.'

c. *Nimal ibeemə* *sedhuwa/næwuwa.*
Nimal by REFL washed.VOL.PST/bathe.VOL.PST
'Nimal washed/bathed by himself.'

This form is more mysterious, since it looks like the output of Causer Suppression (i.e. a nominative subject intransitive variant of an otherwise transitive verb), something unexpected if Causer Suppression is incompatible with volitive mood due to a clash in the subject type. However, there is an alternative analysis of (46), namely that it involves object *pro*-drop on a reflexive interpretation (e.g. a reflexive pro_{refl}). There are several pieces of evidence that this is the correct analysis of (46). First, the reflexive reading is not strictly necessary; a disjoint reference reading is also possible (i.e. Sinhala permits object *pro*-drop more generally), something not generally true of the *gannə* light verb construction:

(47) *((Lots of stuff is happening to Aruni. Bill fed her, John talked to her, and now...))*
*Nimal næwuwa/*naa* *gatta.*
Nimal bathe.VOL.PST/bath.PRT take.PST
'Nimal bathed her.'

Indeed, as Chandralal (2010, 137-138) explicitly notes, light verb vs. bare forms alternate with a reflexive vs. disjoint *pro*-drop reading with roots that are not inherent reflexives:

(48) *Ranjit hapaa gatta/hæpuwa.*
Ranjit bit.PRT take.PST/bit.VOL.PST
'Ranjit bit himself/bit someone.' (*gatta*=reflexive, bare verb=disjoint)

This suggests that while the light verb (in at least some cases; see §6) has an inherently reflexive interpretation, the seemingly intransitive volitive form can in principle be disjoint or reflexive. However, the reflexive option only arises with inherent reflexive verb classes; the seemingly intransitive volitive in (48) with a non-inherent reflexive ("obviative") verb does not admit this interpretation. This might argue against the existence of pro_{refl} in (46), since if it generally exists it should be possible in (48) as well with *hæpuwa*. However, there is further evidence that there is a pro_{refl} in (46), namely that in these cases it is also possible for the subject to be marked by the postpositional subject-case marker *atiŋ*:

(49) *Nimal atiŋ sedhuna/næwuna.*
Nimal POST wash.INV.PST/bathe.INV.PST
'Nimal washed/bathed.'

Crucially, as discussed by Beavers & Zubair (2010, 87-89), *atiŋ* only ever occurs marking subjects of transitive verbs that take a separate direct object DP, thus motivating that there is a null pro_{refl} in (49) and justifying that this analysis could extend easily to (46) as well.[11]

But if pro_{refl} is exists in Sinhala, why is it only attested with inherent reflexive verbs, and why is it furthermore the default reading for them? We suggest that the available readings are essentially root-conditioned (building on Kemmer 1993, Alexiadou & Doron 2012). The simplest analysis would be to say that while all verbs can select non-reflexive *pro* in object *pro*-drop, it is a special fact about inherent reflexives that they may also select for pro_{refl}. However, an alternative analysis may derive this from more basic principles of markedness. In particular, while both obviative and inherently reflexive verbs with overt objects allow reflexive or non-reflexive readings depending on the choice of object, they describe events for which the default expectation is self-action in the case of inherently reflexive verbs and non-self-action in the case of obviative verbs. We suggest that this is the reason the unmarked interpretation of object *pro*-drop for an obviative verb is non-reflexive and the unmarked reading for an inherent reflexive is reflexive. If so, that a reflexive reading is ruled out for obviative verbs can then be explained by an appeal to markedness — this would be a marked reading, and there are overt marked expressions for this reading in the language, namely the light verb construction, which we suggest therefore blocks pro_{refl} from occurring with these verbs. Conversely, for inherent reflexives the marked reading is the obviative one. But in this case there is no marked obviative expression equivalent to *pro*-drop, and thus obviative *pro*-drop is allowed.[12] Thus default expectations about interpretation for different verb classes plus form-to-meaning markedness principles can derive

[11] Our informants also accepted a *pro*-drop reading of (38), though it was dispreferred. Here we suggest this is a variant of (49) with *atiŋ* dropped (something that showed up occasionally in naturally occurring data with otherwise transitive verbs). The crucial point is that *atiŋ* only otherwise occurs with transitive verbs.

[12] There are overt pronouns and reflexives in Sinhala, though these are not entirely freely interchangeable with *pro*-drop in that they convey a different information structural status of their referents.

the distribution of pro_{refl}, meaning the data in (46) can be independently explained and the proposal that Causer Suppression indeed produces only involitives can be maintained.

In sum, the *gannə* light verb seems to be *the* canonical expression of inherent reflexives allowing volitional readings, with pro_{refl} serving as a secondary strategy, and Causer Suppression arising only in cases where the subject acts non-volitionally. The simplest analysis of *gannə* is that it takes a transitive verb permitting eventuality subjects and outputs an intransitive form with the same subject, but binding off the patient and introducing conditions that ensure that whatever referent is introduced by the Generalized Quantifier subject DP as the $causer'$ of the causing event v is also the patient, thereby deriving a reflexive reading:

(50) $[\![gannə]\!] = \lambda P \lambda v \in V \lambda e \exists y[P(y,v,e) \wedge \forall z[causer'(z,v) \rightarrow z = y]]$

Applied to the participial form of (41a), the resulting form would be (ignoring tense again):

(51) $[\![naa\ gatta]\!] = \lambda v \in V \lambda e \exists y[cause'(v,e) \wedge result'(y,e,\mathbf{bathed'}) \wedge \forall z[causer'(z,v) \rightarrow z = y]]$

That the subject could still in principle be an event in E licenses purposive modification. The denotation for the relevant form in (44) for a causing eventuality in E would be:

(52) $[\![Mamə\ naa\ gatta]\!] = \lambda e \exists v \in E \exists y[causer'(\mathbf{I},v) \wedge cause'(v,e) \wedge result'(y,e,\mathbf{bathed'}) \wedge \forall z[causer'(z,v) \rightarrow z = y]]$

The key point is that while there is disunity among expressions of inherent reflexives, the disunity has a principled explanation: the degree to which Causer Suppression is the Sinhala instantiation of the cross-linguistically attested "normal" middle forming operation posited for Indonesian (and presumably extant in other languages), its use is limited by an interaction with Sinhala volitive mood to only allow non-volitional readings (modulo specialized uses such as ironic denial). The other possible expressions of the middle lack this constraint, and serve to fill in the lacuna, and indeed the *gannə* light verb in particular produces forms whose meanings are truth conditionally equivalent to the output of Causer Suppression were it to apply among volitive verbs, suggesting that there is a core unity to all middles even if the overt expression differs considerably. We now consider non-reflexive middles in Sinhala, looking first at middle constructions and then middle passives.

6. Existential Binding Middles

Starting with middle constructions, we note first that it was difficult to get consistent judgments from our informants since the middle construction reading is hard to get across accurately. That said, three variants arose that seem to serve this functionality. First, some speakers found accusative subject involitives to most naturally allow this reading:[13]

(53) *Vesi-wə* *pahasuven vikunenəwa.*
prostitutes-ACC easily sell.INV.NPST
'Prostitutes sell easily.'

[13] Sinhala is a differential object marking language and as such accusative mainly only occurs on human DPs, hence the need for plausible human subjects even with verbs meaning 'sell'.

These forms disallow *ibeemə* and purposives, but entail external causation:

(54) a. *#Vesi-wə ibeemə pahasuven vikunenəwa.*
prostitutes-ACC by REFL easily sell.INV.NPST
#'Prostitutes sell easily by themselves.'

b. **Vesi-wə$_i$ [PRO$_{i/j}$ tarangaya dinannə] pahasuven vikunenəwa.*
prostitutes-ACC contest win.VOL.INF easily sell.INV.NPST
*'Prostitutes sell easily to win the contest.'

c. *#Vesi-wə pahasuven vikunanuna, eet kawuruwat eewa-wə*
prostitutes-ACC easily sell.INV.PST, but nobody 3.PL-ACC
vikunaneuwe næ.
sell.VOL.CAUS.PST.EMPH NEG
#'Prostitutes sold easily, but nobody sold them.'

This is all as expected — if the appropriate analysis of accusative subject intransitive involitives involves existential binding of the underlying subject argument then such a form with a generic or ability modal interpretation will serve as a middle construction.

However, there are limits to the applicability of this operation in forming middle constructions since it would only occur with verb forms with general subjects, i.e. just those verbs that otherwise form anticausatives and inherent reflexives, since in general Sinhala Causer Suppression only applies to these verbs. Yet as discussed in §2, cross-linguistically middle constructions occur with a much wider range of (mostly) change-of-state verbs. Causer Suppression indeed does not generate middles with other agent-subject verbs such as words meaning 'cut' that would otherwise form acceptable middle constructions in English (evidenced by the acceptable translation):

(55) **Vesi-wə pahasuven kæpenəwa.*
prostitutes-ACC easily cut.INV.NPST
'Prostitutes cut easily.'

This leaves open how (if at all) middle constructions could even be formed with such verbs.

For some speakers we consulted the light verb construction instead served as the canonical middle construction expression (and allowed middle constructions like *kapaa gatta* 'cut took' "got cut", contra (55)), entailing external causation and rejecting purposives and *ibeemə*, consistent with other languages:

(56) a. *Meeka kaar-eka pahasuven viku gannəwa*
this car-INDF easily sell.PRT take.NPST
'This car sells easily.'

b. **[Meeka kaar-eka]$_i$ [PRO$_{i/j}$ tarangaya dinannə] pahasuven viku*
this car-INDF contest win.VOL.INF easily sell.PRT
gannəwa
take.NPST
*'This car sells easily to win the contest.'

c. *#Meeka kaar-eka ibeemə viku gannəwa*
this car-INDF by REFL sell.PRT take.NPST
#'This car sells by itself.' (on intended reading)

d. *#Meeka kaar-eka pahasuven viku gatta, eet kawuruwat eyaa-wə*
this car-INDF easily sell.PRT take.PST but nobody 3.SG-ACC
vikunaneuwe næ.
sell.VOL.CAUS.PST.EMPH NEG
#'This car sold easily, but nobody sold it.'

For these speakers, the analysis of *gannə* given above — wherein it always generates a reflexive reading by binding off the patient, stipulating that the subject is an eventuality (and thus ensuring there is an *causer′* introduced by the subject DP), and equating the patient with the *causer′* — is not appropriate. Rather, there must be a separate use of *gannə* that binds off the causing eventuality and introduces a *causer′* distinct from the patient:

(57) $[\![gannə]\!] = \lambda P \lambda y \lambda e \exists v \in V \exists x [causer'(x, v) \wedge P(y, v, e)]$

Applied to the 'sell' root the interpretation would be like an accusative subject anticausative:

(58) $[\![viku\ gatta]\!] = \lambda y \lambda e \exists v \in V \exists x [causer'(x, v) \wedge cause'(v, e) \wedge result'(y, e, \textbf{sold}')]$

The middle construction itself is the result of a generic modality applied over a clause headed by such a predicate. In this case the paradigmatic contrast between $+\emptyset_{CS}$ and *gannə* is sharpest, essentially reflecting volitivity-sensitive and volitivity-neutral variants of the same operation, albeit achieved through slightly different means (saturation of an argument vs. existential binding of different arguments and conditions on co-reference/disjoint reference). That said, other speakers we consulted *only* got a reflexive reading with *gannə*; examples corresponding to (57) readings are not available at all, and so *gannə* and Causer Suppression are not entirely interchangeable for these speakers. Thus for some speakers middle constructions may be more limited in Sinhala than in other languages.

That said, a further encoding option for middle constructions is discussed by Gair (1970, 70-71, 76), who explicitly notes a class of what he refers to as "Subjectless Active Clauses" and their corresponding involitives ("Subjectless Inactive Clauses") that are glossed as what seems clearly to be middle constructions given their semantics, occurring in both volitive and involitive mood (adapted from Gair 1970, 70, 76):[14]

(59) a. *Mee wat-te wii wawənəwa.*
this estate-DAT unhusked_rice grow.VOL.NPST
'Unhusked rice is grown on this estate.'

b. *Dawəsəkə-ṭə pol siiak witərə kæḍenəwa.*
day-DAT coconut hundred about cut.INV.NPST
'About a hundred coconuts a day are/get picked.'

As Gair discusses, in each case it is possible to insert an overt subject into such constructions with no other grammatical change:

[14]Our informants did not produce these possibilities and we leave it for future research to verify their acceptability for speakers who also use the middle constructions discussed above.

(60) a. *Taatta mee wat-te wii wawənəwa.*
father this estate-DAT unhusked_rice grow.VOL.NPST
'Father grows unhusked rice on this estate.' (adapted from Gair 1970, 70)

b. *Nimal atiŋ dawəsəkə-ṭə pol siiak witərə kæḍenəwa.*
Nimal POST day-DAT coconut hundred about cut.INV.NPST
'Nimal picks about a hundred coconuts a day.'

This might suggest that the constructions in (59) therefore involve no argument structural shift at all, but instead reflect a type of *pro*-drop grammatically (which is in fact Gair's analysis). If data such as this involves *pro*-drop, though, then the relevant null pronoun must be of an appropriate type to generate the middle construction type reading, e.g. pro_{one}. Crucially, this would predict that the volitive examples at least should actually allow purposives modifiers, unlike canonical middle constructions, despite sharing a related semantics, though we have so far been unable to verify this. In sum, it appears that there are a range of candidates for expressing the middle construction, in this case not all necessarily grammatical or semantically fully equivalent, with speaker variation on what is possible. Crucially, the options beyond Causer Suppression are also not subject to conditions ruling out volitive mood stems, thus again filling in that lacuna in the paradigm of "normal" middle formation.

We now briefly discuss passive middles, before turning to some broader commentary on these two types of middles together. The existence of passive middles in Sinhala has in fact already been discussed above in §3 — these would correspond to Causer Suppressed verbs with accusative subjects, which crucially show all of the properties of passive middles in Indonesian, e.g. they reject purposive modifiers (unlike regular passives in Indonesian) and *ibeemə* 'by itself' modifiers, but do entail external causation, and bear episodic readings (data repeated from (24)):

(61) a. *Meeri-wə (#ibeemə) giluna.*
Mary-ACC by REFL drown.INV.PST
'Mary drowned.'

b. *#Eyaa-wə lissuna, eet kawuruwat/kisivat eyaa-wə*
3SG-ACC slip.INV.PST but nobody/nothing 3SG-ACC
lisseuwe nææ.
push.VOL.CAUS.PST.EMPH NEG
'She fell, but nobody/nothing pushed her.'

c. *[$PRO_{i/j}$ *Rakshana salli gannə*], *Meeri-wə$_i$ giluna.*
insurance money take.VOL.INF Mary-ACC drown.INV.PST
*'Mary$_i$ drowned [$PRO_{i/j}$ to collect the insurance money].'

One significant aspect of Sinhala, though, is that these middles are again restricted to the verb classes that allow anticausative or inherent reflexive readings, i.e. those that are amenable to Causer Suppression at all, something that follows from the constraints on which verbs may undergo this operation to begin with. In Indonesian the verbs that allow the passive middle interpretation are those that allow the middle construction reading, typically caused change-of-state verbs regardless of whether agentivity is entailed or not

of their subjects, something that as far as we can tell is similar in other languages such as Spanish. At this point it is again a fair question to ask how other passive readings are derived for other verbs in general. In the case of Indonesian (as discussed above) there is a separate personal passive *di-* form that has significantly more general applicability, and thus in principle there are no particular constraints on which forms may show some type of passive, although the *di-* passives and middle passives are not identical in their grammatical properties (e.g. the former admit *oleh* 'by' PPs expressing the agent and the agent is accessible to purposive modifiers, as discussed above, but not in passive middles). In spoken Colloquial Sinhala there is however no canonical passive equivalent to this. Rather, as discussed by Chandralal (2010, 152-160), the functionality of a passive qua its role in a language like English or Indonesian in deemphasizing the agent but preserving it is instead picked up by a range of other constructions, including topic-comment structures, various sorts of *pro*-drop constructions (both subject and object *pro*-drop), and various uses of the involitive.[15] In this way, the passive middle and middle construction bear much in common in terms of a considerable diversity in encoding: in both cases a wide range of grammatically and semantically disparate constructions in Sinhala convey what in some other languages may be one or two separate constructions. This accords, however, with work specifically focusing on middle constructions that have suggested that these particular middles are truly notional, i.e. just a (generic or ability) reading of some other construction (Condoravdi 1989, Lekakou 2002, Ackema & Schoorlemmer 2005, Fábregas & Putnam 2014). In this case the expectation is that the various constructions that serve as middle constructions should in fact behave differently, inheriting whatever properties the underlying construction has, and the same would presumably be true of the passive middle. That said, the argument suppressing operations that generate other middles ($+\emptyset_{CS}$ and *gannə*) are among those operations that middle constructions and passive middles can be based around, and in general the semantics of other middle construction and passive forms is consistent with the kinds of semantics generated by more explicit middle formation operations. Plus the distribution of different ways of forming these middles follows the lines expected on the analysis of Causer Suppression in Sinhala suggested above. Thus again all types of middles have at least a partially unified analysis.

7. Conclusion

We have suggested that all middles have a fundamental commonality: an inherently dyadic verb has its valence reduced but not its truth conditional content, with two ways of understanding the suppressed argument, either reflexively or with disjoint reference to the expressed argument. In principle this plus verb type constraints will derive the core classes of middles, and in many languages the realization of this operation is consistent across middles. In Sinhala, however, volitives but not involitives require agent subjects in a type-

[15]The involitive has in fact been argued to simply be a passive form in some prior literature (see e.g. Gunasinghe 1985, Gunasinghe & Kess 1989, Kahr 1989, Wijayawardhana et al. 1991), but the semantic effects of involitivity in terms of non-volitionality and the restrictions on which verbs show it in which argument structures strongly argue against this classification, although as noted it nonetheless may in some contexts serve a function similar to a canonical passive.

theoretic (qua grammatical) sense, and this interacts significantly with the Sinhala equivalent of the middle formation operation otherwise attested in languages like Indonesian to rule out certain classes of middles from being formed by it. Other constructions serve to realize the middles that are ruled out due to this clash. The various middles indeed split along predicted lines: effector subject verbs and inherent reflexives form middles in a consistent fashion, but for other verb classes other constructions fill in the gaps.

(62)

Interpretation	**Verb Type**		**Middle Type**
	Effector Subject Verbs	*Other*	
Reflexive	NOM subj+$V_{inv} + \emptyset_{CS}$	N/A	Anticausatives
	NOM subj+$V_{inv} + \emptyset_{CS}$	V+*gannə*/pro_{refl}+$V_{inv/vol}$	Inherent reflexive
∃-*binding*	ACC subj+$V_{inv} + \emptyset_{CS}$	Various	Passive
	ACC subj+$V_{inv} + \emptyset_{CS}$	V+*gannə*//pro_{one}+$V_{inv/vol}$	Middle construction

In Indonesian no such distinction exists, and all verbs are treated identically via one Causer Suppression operation. Thus the core unity is the same across languages, but the way it manifests in different languages due to language internal factors can create the appearance of dissimilarity, albeit dissimilarity that is principled in nature.

References

Ackema, Peter, & Maaike Schoorlemmer. 1994. The middle construction and the syntax-semantics interface. *Lingua* 93:59–90.

Ackema, Peter, & Maaike Schoorlemmer. 2005. Middles. In *The blackwell companion to syntax*, ed. Martin Everaert & Henk van Riemsdijk, volume III, 131–203. Oxford: Blackwell.

Alexiadou, Aartemis, & Edit Doron. 2012. The syntactic construction fo two non-active Voices: Passive and middle. *Journal of Linguistics* 48:1–34.

Alexiadou, Artemis. 2010. On the morpho-syntax of (anti-)causative verbs. In *Syntax, lexical semantics and event structure*, ed. M. Rappaport Hovav, E. Doron, & I. Sichel, 177–203. Oxford: Oxford University Press.

Alexiadou, Artemis, Elena Anagnostopoulou, & Florian Schäfer. 2006. The properties of anti-causatives crosslinguistically. In *Phases of interpretation*, ed. Mara Frascarelli, 187–211. Berlin: Mouton de Gruyter.

Alexiadou, Artemis, & Florian Schäfer. 2014. Towards a non-uniform analysis of naturally reflexive verbs. In *Proceedings of the 31st West Coast Conference on Formal Linguistics*, ed. Robert E. Santana-LaBarge, 1–10. Somerville: Cascadilla Proceedings Project.

Anderson, Mona. 1979. Noun phrase structure. Doctoral dissertation, The University of Connecticut.

Beavers, John. 2008. The true role of affectedness in DP-preposing. Talk given at the 2008 Annual LSA Meeting, Chicago.

Beavers, John, & Andrew Koontz-Garboden. 2012. Manner and result in the roots of verbal meaning. *Linguistic Inquiry* 43:331–369.

Beavers, John, & Andrew Koontz-Garboden. 2013. Complications in diagnosing lexical meaning: A rejoinder to horvath and siloni (2013). *Lingua* 134:210–218.

Beavers, John, & I Nyoman Udayana. 2016. Middle voice in Indonesian. Unpublished ms., The University of Texas at Austin and Udayana University.

Beavers, John, & Cala Zubair. 2010. The interaction of transitivity features in the Sinhala involitive. In *Transitivity: Form, meaning, acquisition, and processing*, ed. Patrick Brandt & Marco Garcia, 69–92. Amsterdam: Benjamins.

Beavers, John, & Cala Zubair. 2013. Anticausatives in Sinhala: Involitivity and causer suppression. *Natural Language and Linguistic Theory* 31:1–46.

Beavers, John, & Cala Zubair. 2014. Default case and the involitive in colloquial sinhala. Talk given at FASAL 4, March 29-30.

Chandralal, Dileep. 2010. *Sinhala*. Amsterdam: John Benjamins.

Chierchia, Gennaro. 2004. A semantics for unaccusatives and its syntactic consequences. In *The unaccusativity puzzle*, ed. Artemis Alexiadou, Elena Anagnostopoulou, & Martin Everaert, 22–59. Oxford: Oxford University Press.

Chou, C.-T. Tim, & Sujeewa Hettiarachchi. To appear. On sinhala subject case marking and A-movement. *Lingua* .

Condoravdi, Cleo. 1989. The middle: Where semantics and morphology meet. In *Mit working papers in linguistics 11*, 16–31. MIT Press.

Copley, Bridget, & Phillip Wolff. 2014. Theories of causation should inform linguistic theory and vice versa. In *Causation in grammatical structures*, ed. Bridget Copley & Fabienne Martin, 11–57. Oxford: Oxford University Press.

Doron, Edit. 2003. Agency and voice: The semantics of the Semitic templates. *Natural Language Semantics* 11:1–67.

Dowty, David. 1979. *Word meaning and Montague Grammar*. Dordrecht: Reidel.

Embick, David. 2004. On the structure of resultative participles in English. *Linguistic Inquiry* 35:355–392.

Fábregas, Antonio, & Michael Putnam. 2014. The emergence of middle voice structures with and without agents. *The Linguistic Review* 33:193–240.

Fagan, Sarah M. B. 1992. *The syntax and semantics of middle constructions*. Cambridge: Cambridge University Press.

Gair, James W. 1970. *Colloquial Sinhalese clause structures*. The Hague: Mouton.

Gair, James W. 1990. Subjects, case, and INFL in Sinhala. In *Experiencer subjects in south asian languages*, ed. Manindra K. Verma & K. P. Mohanan, 13–42. Stanford: CSLI Publications.

Grimshaw, Jane. 1982. On the lexical representation of Romance reflexive clitics. In *The mental representation of grammatical relations*, ed. Joan Bresnan, 87–148. Cambridge, MA: MIT Press.

Guerssel, Mohamed, Kenneth Hale, Mary Laughren, Beth Levin, & Josie White Eagle. 1985. A cross-linguistic study of transitivity alternations. In *CLS 21, Part 2: Papers from the Parasession on Causatives and Agentivity*, ed. William H. Eilfort, Paul D. Kroeber, & Karen L. Peterson, 48–63. Chicago Linguistic Society.

Gunasinghe, Khema Hemamala Himaransi. 1985. Passive voice: A new perspective, some evidence for a re-analysis from sinhala. Doctoral dissertation, University of Victoria.

Gunasinghe, Khema Hemamala Himaransi, & Joseph F. Kess. 1989. Volition, agency and the Sinhala passive. *Indian Linguistics* 50:47–74.

Haspelmath, Martin. 1993. More on the typology of inchoative/causative verb alternations. In *Causation and transitivity*, ed. Bernard Comrie & Maria Polinsky. Amdersdam: John Bejamins.

Henadeerage, Deepthi Kumara. 2002. Topics in Sinhala syntax. Doctoral dissertation, The Australian National University.

Inman, Michael Vincent. 1993. Semantics and pragmatics of Colloquial Sinhala involitive verbs. Doctoral dissertation, Stanford University.

Jaeggli, Osvaldo A. 1986. Passive. *Linguistic Inquiry* 17:587–622.

Kahr, Joan Caspar. 1989. Discourse functions of the sinhala passive. Doctoral dissertation, Harvard University, Cambridge, MA.

Kardana, I Nyoman. 2011. Types of middle voice in Indonesian language (*tipe-tipe diatesis medial dalam bahasa indonesia*). *Jurnal Melayu* 7:83–105.

Kaufmann, Ingrid. 2007. Middle voice. *Lingua* 117:1677–1714.

Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam: John Benjamins.

Keyser, Samuel Jay, & Thomas Roeper. 1984. On the middle and ergative constructions in English. *Linguistic Inquiry* 15:381–416.

Koontz-Garboden, Andrew. 2009. Anticausativization. *Natural Language and Linguistic Theory* 27:77–138.

Krejci, Bonnie Jean. 2012. Causativization as antireflexivization: A study of middle and ingestive verbs. Master's thesis, The University of Texas at Austin, Austin, TX.

Landman, Fred. 1992. The progressive. *Natural Language Semantics* 1:1–32.

Lekakou, Marika. 2002. Middle semantics and its realization in english and greek. *UCL Working Papers in Linguistics* 14:399–416.

Lekakou, Marika. 2006. A comparative view of the requirement for adverbial modification in middles. In *Demoting the agent*, ed. Benjamin Lyngfelt & Torgrim Solstad, 167–196. Amsterdam: John Benjamins.

Levin, Beth, & Malka Rappaport Hovav. 1995. *Unaccusativity: At the syntax-lexical semantics interface*. Cambridge, MA: MIT Press.

Maldonado Soto, Ricardo. 1992. Middle voice; the case of Spanish *se*. Doctoral dissertation, University of California, San Diego, San Diego, CA.

Piñón, Christopher. 2012. The reflexive impersonal construction in Polish. Paper presented at the Workshop on Argument Structure, The University of Debrecen, Debrecen, Hungary, May 25-27.

Prior, Elizabeth W., Robert Pargetter, & Frank Jackson. 1982. Three theses about dispositions. *American Philosophical Quarterly* 19:251–257.

Reinhart, Tanya. 2000. The theta system: Syntactic realization of verbal concepts. In *Ots working papers in linguistics*. Utrecht Institute of Linguistics, University of Utrecht.

Reinhart, Tanya. 2002. The theta system - an overview. *Theoretical Linguistics* 28:229–90.

Reinhart, Tanya, & Tal Siloni. 2005. The lexicon-syntax parameter: Reflexivization and other arity operations. *Linguistic Inquiry* 36:389–436.

Siewierska, Anna. 1984. *The passive: A comparative linguistic analysis*. London: Croom Helm.

Stroik, Thomas. 1992. Middles and movement. *Linguistic Inquiry* 23:127–137.
Tenny, Carol. 1992. The aspectual interface hypothesis. In *Lexical matters*, ed. Ivan A. Sag & Anna Szabolcsi, 490–508. Stanford: CSLI Publications.
Van Valin, Robert D., Jr., & David P. Wilkins. 1996. The case for 'effector': Case roles, agents, and agency revisited. In *Grammatical constructions: Their form and meaning*, ed. Masayoshi Shibatani & Sandra A. Thompson, 289–322. Oxford: Oxford University Press.
Wijayawardhana, G. D., Daya Wickramasinghe, & Theodora Bynon. 1991. Passive-related constructions in Colloquial Sinhala. In *Subject, voice and ergativity*, ed. David C. Bennet, Theodora Bynon, & B. George Hewitt, 105–141. London: Routledge.
Wunderlich, Dieter. 1997. Cause and the structure of verbs. *Linguistic Inquiry* 28:27–68.
Zubair, Cala. 2008. Doxastic modality as a means of stance taking in Colloquial Sinhala. In *Proceedings of SALSA 16*, ed. Amy Brown & Josh Iorio, 174–190.

John Beavers & Cala Zubair
jtbeavers@utexas.edu & calazubair@gmail.com

Causation in Hindi-Urdu: Care for your Instruments and Subjects*

Sakshi Bhatia

University of Massachusetts, Amherst

1. Introduction

This paper seeks to highlight novel differences between direct and indirect causatives in Hindi-Urdu with respect to 'optional' *-se* phrases and 'causer' subjects. As opposed to the singular instrument *-se* phrase in the case of direct causatives, indirect causatives allow multiple *-se* phrases - two instruments and an intermediate agent. While the licensing of the intermediate agent *-se* phrase in indirect causatives has been the cornerstone of various theoretical accounts of Hindi-Urdu causatives (Kachru 1980, Saksena 1982, Bhatt & Embick 2003, Ramchand 2008, 2010 & Richa 2011 among others), the additional instrument *-se* phrase in indirect causatives has not been discussed in detail hitherto.

In this paper I demonstrate that this observation regarding instruments can be best accounted for under a conceptualization of indirect causation as being structurally and semantically more complex than the direct causative. In particular, it is the bi-eventiveness of indirect causation, as opposed to the mono-eventiveness of direct causation, that mediates the licensing of an additional instrument. In this way this paper provides evidence against the analysis of indirect causatives as involving an event structure which lacks an interpretive boundary between the causing and caused events (Ramchand 2008, 2010).

The relativization of argument licensing to event structure, in the vein of a long line of work from Carlson (1984) to Williams (2015) recently, allows for a uniform analysis of *-se* phrases in causatives. The proposed unification of the broad class of 'optional' *-se* phrases, thus, seeks to take forward an enterprise initiated by Ramchand (2010), albeit with different theoretical assumptions, tools and consequences. Rather than having the interpretation of *-se* phrases be sensitive to the presence of 'implicit' sub-events, I argue for an analysis which cashes out the varying interpretations of *-se* phrases as a consequence of the mod-

*I would like to thank Kyle Johnson, Rajesh Bhatt, Angelika Kratzer, Lyn Frazier, John Beavers, Barbara Partee, Seth Cable, Ellen Woolford, Vincent Homer, Michel Clauss, Leland Kusmer, Jyoti Iyer, Ekaterina Vostrikova, Ethan Poole,Jon Ander Mendia, Mioko Miyama, Yutaka Ohno, audiences at the Syntax Workshop, the Syntax-Semantics Reading Group, the Second Year Mini-conference at the University of Massachusetts, and participants of FASAL 5 for helpful discussion and feedback. Parts of this project were funded by a Summer Research Grant (2015) from the University of Massachusetts, Amherst.

ification of distinct syntactico-semantic objects in each case. The analysis presented here has the advantage of being straightforwardly able to capture the licensing and interpretative conditions on instrument 'causer' subjects across the causative verbal alternations, without taking recourse to any additional stipulations regarding causers themselves.

The paper is organized as follows: §2 provides a brief overview of the causative verbal alternations and *-se* phrases along with recapitulating previous accounts of the phenomena. §3 discusses the additional data motivating the movement away from pre-existing analyses and presents the analysis for *-se* phrases in causatives. §4 discusses the consequences of the proposal for 'causer' subjects and §5 concludes the paper.

2. A short tour of Hindi-Urdu causatives

Hindi-Urdu has a productive morphological causativization paradigm. This involves suffixation of causativizing morphemes to verbal roots (Kellogg 1876, Kachru 2006, inter alia). Two types of causatives are identified: The **direct causative (DC)** realized with the suffix *-aa* or stem alternation and the **indirect causative (IC)** realized with the suffix *-vaa*. This three way alternation is illustrated in (1) with an unaccusative verb as the first member of the paradigm and transparent suffixation indicating causativization. Unergatives, transitives and ditransitives also participate in the causative alternation, but this paper does not explore these additional paradigms due to space constraints.

(1) *The causative paradigm.*

Unaccusative	Direct causative	Indirect causative
jal	jal-aa	jal-vaa
burn	burn	cause to burn

The use of the terms 'direct causative' and 'indirect causative' is based on Saksena (1982). The use of the direct causative suffix *-aa* 'signifies that the verb is a causative with a personally involved causer' (Saksena 1982: 2) and the indirect causative suffix *-vaa* signifies that the 'verb is a causative with a non-involved causer'. Thus, in (2) the ergative marked nominal *zamindaar* 'landlord', is understood to be directly involved in the burning of the house as its agent when the corresponding verbal form is that of the *-aa* suffixed direct causative. In contrast, when the verbal form is that of the *-vaa* suffixed indirect causative, the landlord is not understood to be involved in the burning of the house itself, with that task having been delegated to an unmentioned intermediate agent in this case.

(2) a. ghar **jal**-aa
House burn-PFV
'The house burned.'

b. zamindaar-ne ghar **jal-aa**-yaa
Landlord-ERG house burn-DC-PFV
'The landlord burned the house.'

c. zamindaar-ne ghar **jal-vaa**-yaa
Landlord-*erg* house burn-IC-PFV
'The landlord had the house burned.' (Based on Bhatt & Embick 2003)

Causativization can be realized through stem alternation for some predicates, for instance, *khul* 'open (unaccusative)', (3). See Bhatt & Embick (2003) for a detailed discussion.

(3) *Stem alternation.*

Unaccusative	Direct causative	Indirect causative
kh**u**l	kh**o**l	kh**u**lvaa
open	open	cause to open

In addition to the 'core' arguments discussed above, causative predicates permit a range of 'optional' arguments bearing the case marker *-se* (*-se phrases*). These *-se* phrases may introduce an *instrument* and an *intermediate agent* into the syntax, see *mashaal* 'torch' in (4-a) and *Dakait* 'bandit' in (4-b) respectively. The intermediate agent is distinct from an instrument in that it is an animate entity capable of performing the action described volitionally and is licensed only when the corresponding verb form is that of the indirect causative, (4-c). Arguably, it is this distinction regarding the licensing of the intermediate agent that has been almost definitional of indirect causatives (Saksena 1982). The basic distribution of *-se* phrases is summarized in (5).

(4) a. ye ghar *mashaal*-se jal-aa thaa
this house torch-INST burn-PFV be.PAST
'This house was burned with a torch.'
(i.e. the burning happened with a torch)

b. zamindaar-ne (**Dakait-se**) (*mashaal-se*) ghar jal-vaa-yaa
Landlord-ERG bandit-INST torch-INST house burn-IC-PFV
'The landlord had the house burned by the bandit.'

c. zamindaar-ne (***Dakait-se**) (*mashaal-se*) ghar jal-**aa**-yaa
Landlord-ERG bandit-INST torch-INST house burn-DC-PFV
'The landlord had the house burned by the bandit.'

(5) *Distributions of -se phrases.*

-se phrase	Unaccusative	Direct causative	Indirect causative
Instrument	✓	✓	✓
Intermediate agent	×	×	✓

2.1 Previous accounts of Causatives

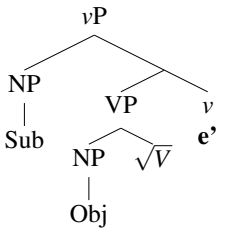
Previous accounts of Hindi-Urdu causatives have varying empirical coverage - Bhatt & Embick's (2003) account concentrates more on the intermediate agent *-se* phrase, while Ramchand's (2010) proposal also deals with the interpretation of one instrument *-se* phrase.

I will briefly present their ideas before turning to additional data that neither of their proposals sufficiently captures.

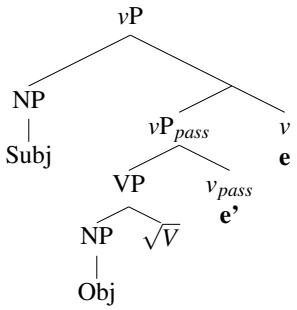
2.1.1 Bhatt & Embick 2003: The embedded passive

In Bhatt and Embick's (2003) proposal the DP agent of the transitive (direct causative) is licensed by an agentive *v*. This argument is added to the structure via *Event Identification* (Kratzer 1996). The syntactic structure for a basic direct causative under their account is given in (6). Like the direct causative the DP agent of an indirect causative is also licensed by an agentive *v* under their proposal. This *v* embeds *a passive complement*, which is a *vP* that contains an agentive *v*, but no case feature and no DP in the specifier of this head. With the embedded vP lacking an external argument, the resultant structure has the matrix subject be the agent of the causing event but not the embedded event and the embedded event does not have an explicitly realized agent, giving us the structure in (7). The agent of the embedded event - the intermediate agent - can then be introduced in a *-se* phrase modifying the passive *vP*.

(6) *Direct causative.*



(7) *Indirect causative.*



The corresponding semantic denotation for the direct causative is in (8-a) and for the indirect causative in (8-b). There is no overt causative component for the direct causative in this account. By having two agentive *v* heads as part of the indirect causative structure, Bhatt & Embick's proposal allows for the introduction of two events since each agentive *v* introduces a new event. Event identification is not available for events introduced by agentive *v* heads as that would be semantically anomolous and a causation relation holds between the two event arguments introduced by these two *v*'s.

(8) a. *Direct Causative.*
$[[subj\ obj\ \sqrt{V}\ v]] = \lambda e_s$ Agent(subj)(e) & V(obj)(e')]

b. *Indirect Causative.*
$[[subj\ obj\ \sqrt{V}\ v_{pass}\ v]] = \lambda e_s$Agent(subj)(e) & $\exists e'_s$[CAUS(e,e')
& $\exists x_e$[Agent(x)(e')] & V(obj)(e')]

Even though Bhatt & Embick do not engage with the issue of licensing instruments, we will see in §3.1 that the distribution of instruments also pushes us towards a bieventive

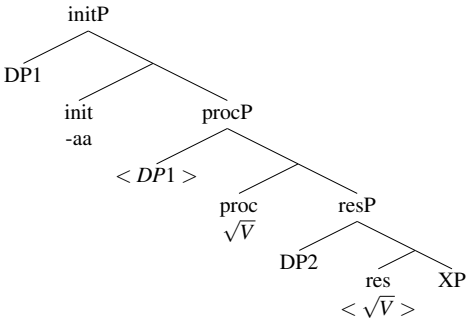
account of indirect causatives with an interpretive distinction between and causation event and a caused event.

2.1.2 Ramchand 2008, 2010: Modifying Overt and Covert Subeventualities

Ramchand's (2010) analysis of *-se* phrases builds on her (2008) proposal and takes the functional sequence corresponding to V to have been put together from a recursive embedding of eventuality descriptors - init(iator)P, proc(ess)P and res(ult)P - whose specifiers are systematically interpreted locally as the 'thematic' element of each sub-description. The account assumes that the *cause/leads to* relation holds between sub-events such that there are two loci of causation in a maximally complex event.

Based on the additive nature of the morphology of alternation, she argues for a structure building analysis where the direct causative (transitive) alternant is structurally larger than the intransitive version. The direct causative counterpart of an unaccusative verb would have the structure in (9). The indirect causative does not differ from the direct causative in having additional overt syntactic structure, (10). Rather, it differs from the direct causative in that there is no temporal overlap or common lexical content that is asserted for the *proc* and *res* subevents in the case of indirect causative, such that the whole event will be interpreted as involving an 'indirectly caused' result.

(9) *Direct Causative.*



(10) *Indirect causative.*

Furthermore, with the *-vaa* suffix multiply inserting into both *init* and *proc*, any verb root that combines with it will have to leave some of its own category features unassociated. This has the consequence that there exists an underassociated *proc* in the verbal phase whose encyclopedic content is still accessible to the semantics. Thus, rather than arguing for the presence of an implicit argument (for example, the intermediate agent) in a verb's argument structure, she argues in favour of implicit sub-eventual structure for the indirect causative.

The unified analysis of *-se* phrases based on this event-structural account, has *-se* phrases be sub-event modifiers, such that the 'instrument' interpretation arises if the overt *proc* is modified, while the 'intermediate agent' interpretation arises if the underassociated *proc* is modified. In providing an analysis where both instruments and intermediate agents form a uniform class of adjuncts in that they are 'non-volitional direct causers', she argues

against the Bhatt & Embick (2003) style analysis where the intermediate agent *-se* phrase has a privileged status as it functions as the by-phrase agent for an embedded passive.

However, since the implicit *proc* is distinct from the overt *proc* only in terms of the former lacking its independent projection or lexicalization and there is no difference between the implicit and overt *proc* in terms of their encyclopedic content, there are some concerns regarding the implementation of Ramchand's (2010) analysis. If both instruments and intermediate agents are 'non-volitional direct causers', as Ramchand suggests, then what prevents a *-se* marked intermediate agent from modifying the overt *proc* in a direct causative structure and have it be interpreted (and licensed) as an intermediate agent in that case, contra the observation in (4-c)? Similarly, in the case of the indirect causative, her account does not prevent a derivation where the *-se* marked intermediate agent modifies the covert *proc* and the instrument modifies the overt *proc* from converging.

Thus, in the absence of further delineation of differences between the two types of *proc*, the empirically attested interpretive differences between intermediate agents and instrument do not follow through. Furthermore, as we will see in the next section, there are additional challenges with respect to extending this account to include the licensing of the second instrument *-se* phrase that is available with indirect causatives.

3. The syntax-semantics of *-se* phrases

In this section I explore the multiplicity of instruments made available by indirect causation and what that implies about the event space for different types of predicates. I also look at additional evidence supporting the Bhatt & Embick proposal which treats the intermediate agent like the agent of the passive. Finally, I present my account of *-se* phrases in causatives.

3.1 Motivation 1: Being *the* instrument of *an* event

The syntactic distribution of instruments is not so much a factor of the valency of a predicate, but rather the predicate's status in the causativization paradigm. Unaccusatives and their (transitive) direct causatives forming a distinct grouping compared to the indirect causatives, since the former group only permits a single instrument -se phrase, which I refer to as an **instrument of the result** (instrument$_{res}$). The greater valency of the transitive is unaccompanied by the availability of a greater number of instrument *-se* phrases, see (11). In contrast, the indirect causative permits an additional instrument, *threats*, in (12). This additional instrument will be referred to as an **instrument of causation** (instrument$_{caus}$).

(11) *zamindaar-ne **maachis**-se *mashaal*-se ghar jal-aa-yaa
landlord-ERG matchstick-INST torch-INST house burn-DC-PFV
'The landlord burned the house with a torch with a matchstick.'

(12) zamindaar-ne apni **dhamkii**-se Dakait-se mashaal-se ghar jal-vaa-yaa
Landlord-ERG SELF'S threat-INST bandit-INST torch-INST house burn-IC-PFV
'The landlord used his threats to get the house burned by the bandits with a torch.'

While an instrument$_{caus}$ is intuitively distinguishable from an instrument$_{res}$ - the former category largely includes abstract means used to make an intermediate agent perform a desired action and the latter category includes concrete means to effect an action affecting the theme argument - there is no motivation for claiming the status of distinct semantic primitives for them and both are considered realizations of the instrument role.

The availability of an instrument$_{caus}$ as a *second* instrument in indirect causatives flies in the face of the long held observation that a semantic relation associated with one dependent cannot be repeated by another, irrespective of whether it is an adjunct or an argument, that is 'each dependent is interpreted as *exhausting* its semantic role, naming all of its satisfiers' (Williams 2015). This observation has formed the basis of the various conceptions of *thematic uniqueness* in the literature (Carlson 1984 inter alia) and has recently been recast as the principle of Role Exhaustion by Williams (2015).

(13) **Role Exhaustion**:
When a dependent is assigned a relation to some *(group of) event(s)*, it identifies *all and only* the individuals in that relation to *that (those) event(s)*.

As an illustration, let us look at (14-a) and (15-a). Here, the event of *smacking* has two roles associated with it: a *smacker* and a *smackee*. Given (13) , (14-a) and (15-a) entail their respective counterparts in (b). However, assigning the role assigned to *the wall* in (14-a), to two objects *the wall* and *the floor* in (15-a) leads to an inconsistency, that is, (15-a) cannot be used to express the meaning in (15-b).

(14) a. Nik smacked the wall.
b. There was smacking, and in it only the wall was smacked, and only Nik was a smacker. *(Williams 2015)*

(15) a. * Nik smacked the wall the floor.
b. There was smacking, and in it only the wall was smacked, only the floor was smacked, and only Nik was a smacker. *(Williams 2015)*

As a semantic principle that governs how any dependent is interpreted, the Role Exhaustion principle requires that a single dependent refer to the entirety of a given relatum, thereby accounting for the unacceptability of (15-a), as well as (11) where having two dependents referring to a single role - that of the instrument - is unacceptable as well.

The co-occurence of two instruments in indirect causatives, (12), would be in clear violation of the Role Exhaustion principle unless, in keeping with the formulation of the principle, which relies on the connection between events and argument structure, we have two (groups of) events at play in indirect causatives. The conclusion, therefore, is that each of these events has an independent existence in the semantics which allows for each of them to be independently modifiable. This in turn allows two instrument roles to be licensed.

Furthermore, this predicts that multiple exponents of other dependents bearing the same relation might be attested. Location is one such candidate. Thus, in (16), two independent

locations, London and Delhi, indicate the site of John's initiation of the entire causation event and the site of Mira's initiation of the caused event respectively.

(16) a. Context: John owns property in Dehradun. Since he is in London, he has to get his lawyer Mira, who is in Delhi, to divide it up for his family.
b. john-ne **london**-se miiraa-se **dilli**-se jaaydaad baNt-vaa-yi
john-ERG london-INST Mira-INST delhi-INST property divide-IC-PFV
From London, John got Mira to get the property divided up from Delhi.

Thus, multiple instruments (and multiple locations) bolster the conceptualization of indirect causation in Hindi-Urdu involving multiple events. This position has been explored independently in the semantic literature on causation. For instance, Kratzer (2005: 27) alludes to indirect causation as being a relation which involves 'possibly very long causal chains connecting the mentioned cause to the mentioned effect'. Therefore, indirect causation is bi-eventive at the very least and by extrapolation it is this property that is responsible for the differentiation observed between indirect causatives and other predicates in Hindi-Urdu as evidenced in the domain of licensing of instruments.

The bi-eventive analysis of indirect causation for Hindi-Urdu has been challenged by Ramchand (2008, 2010) who argues for a complex event structure decomposition for these causatives, but without an interpretive boundary between the causation event and the caused event. However, the modulation of semantic relations such as *instrument* by the event structure argued for above suggests that the event structure decomposition in indirect causatives is in fact associated with an interpretive differentiation. Furthermore, accommodating the second instrument in indirect causatives - instrument$_{caus}$ - while remaining true to the spirit of Ramchand's account is not straightforward. Two possibilities are explored here - either the instrument$_{caus}$ modifies an additional overt *procP* or it modifies the singular overt *procP* of the indirect causative along with the instrument$_{res}$. The former line of thought would necessarily have to be accompanied by constraints on the number of *procP*'s since empirically the number of instruments is not unrestricted. This option also has the consequence that the indirect causative structure would be syntactically more complex than the direct causative, and there would be more than two loci of causation in this sort of event.

The second line of thought - multiple instrument *-se* phrases modifying the same overt *procP* - would have to be accompanied by guidelines for the interpretation of two instruments relative to one sub-event since this is not the standard state of affairs given the Role Exhaustion principle. Furthermore, having the instruments be in roughly the same syntactic position would also fail to capture the effect of the verb root on an instrument$_{res}$ but not an instrument$_{caus}$. It has been noted in the literature that the kind of singular instrument a transitive predicate takes is determined by the lexical meaning of the verb root in question (Rissman 2012) or is encoded as part of the lexical entry of a verb (Erteschik-Shir & Rappaport 2007). Thus, what constitutes a valid instrument varies from one predicate to another - *burning* events requires *torches* and *matchsticks* and *cutting*, *knives*. In Hindi-Urdu, we only see this variation with respect to the instrument$_{res}$ across the verbal inventory while the class of possible instrument$_{caus}$ is not constrained by the properties of the caused event.

This suggests that there exists an asymmetry between instrument$_{res}$ and instrument$_{caus}$, which is not captured by this second extension.

3.2 Motivation 2: Intermediate agents are agents, not instruments

In this section I provide support for the claim that the intermediate agent differs from instruments and not (passive) agents. For one, intermediate agents are able to control into *-kar* adjunct clauses both when overt (Clauss 2014) and when *implicit* while this is not available even to *overt* instrument *-se* phrases.

(17) a. kabir$_i$-ne (mira$_j$-se) john$_k$-ko [**PRO**$_{i/j/*k}$ cilla-kar] jag-vaa-yaa
Kabir-ERG Mira-INST John-DAT scream-do wake-IC-PFV
'Kabir$_i$ got John$_k$ woken up (by Mira$_j$) through his/her$_{i/j/*k}$ shouting.'
(Based on Clauss 2014)

b. kabir$_i$-ne kainchi$_j$-se kapRa$_k$ [**PRO**$_{i/*j/*k}$ fisal-kar] cir-vaa-ya
Kabir-ERG scissors-INST cloth slip-do tear-IC-PFV
'On slipping, Kabir tore the cloth on the scissors.
Unavailable: 'The scissors slipped and Kabir tore the cloth on them.'

Another domain where the intermediate agent patterns with agents is that of binding the possessive anaphor *apnaa*. This anaphor has been traditionally described to be subject oriented in Hindi-Urdu (see for instance Dayal 1994). However, as Srishti (2014) notes the intermediate agent is also a possible binder, as in (18-a). The passive counterpart of Srishti's example also allows for this binding possibility, as in (18-b).

(18) a. miinaa$_i$-ne miikuu$_j$-se **apnaa**$_{i/j}$ darwaazaa khul-vaa-yaa
Mina-ERG Miku-INST SELF'S door open-IC-PFV
'Mina made Miku open his/her door.' (Srishti 2014)

b. miinaa$_i$-dwaaraa miikuu$_j$-se **apnaa**$_{i/j}$ darwaazaa khul-vaa-yaa ga-yaa
Mina-BY Miku-INST SELF'S door open-IC-PFV go-PFV
'Mina made Miku open his/her door.'

Finally, intermediate agents can bear the marker *dwaaraa* 'by' used for agents of regular passives instead of *-se*, and the use of this marker on instruments is very marked.

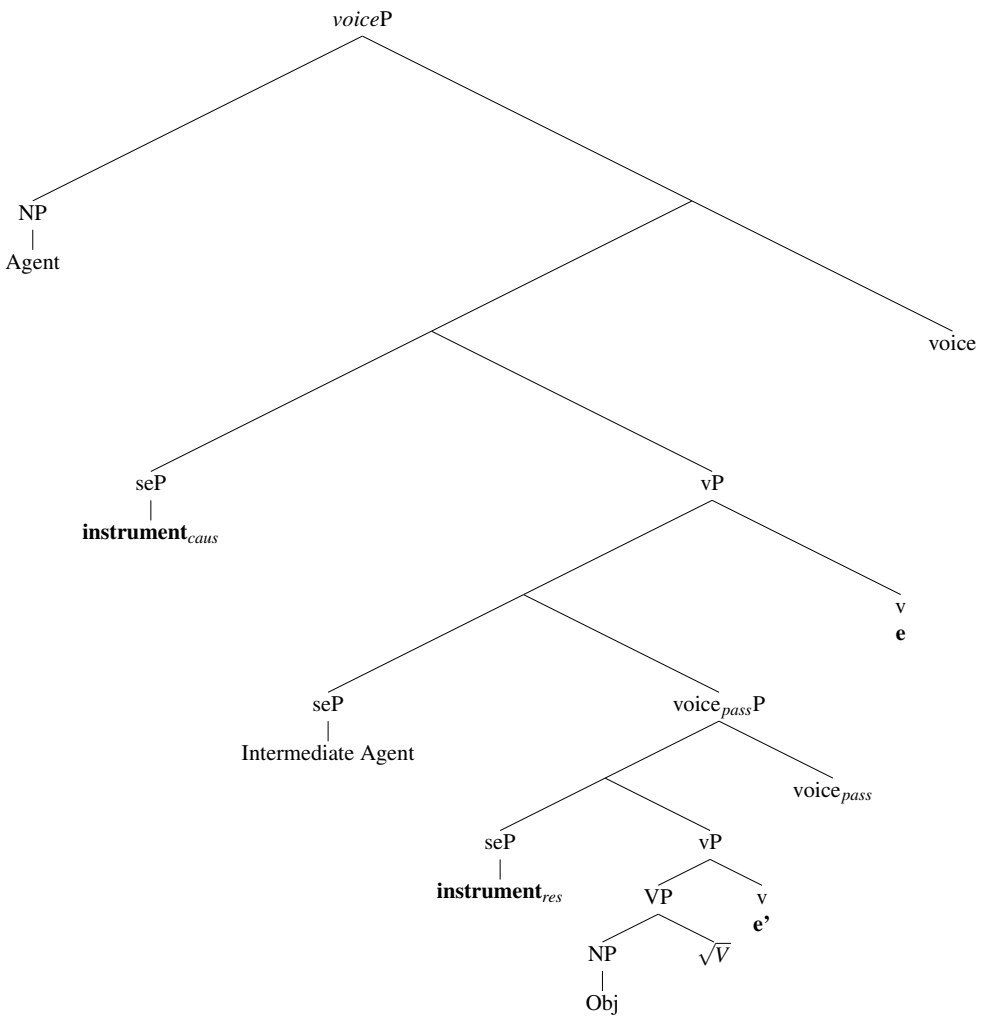
(19) miiraa-ne *raam -dwaaraa*/-se **aarii ??-dwaaraa**/-se lakRii kat-vaa-yii
Mira-ERG Ram -BY/INST saw -BY/INST wood cut-IC-PFV
'Mira had the wood cut by Ram with a saw.'

Together, all of the evidence presented above suggests that the similar syntactic treatment of (passive) agents and the intermediate agent *-se* phrase has its merits.

3.3 The proposal

Given the structural and interpretive differences between direct and indirect causatives with respect to instrument *-se* phrases - (a) unaccusatives and their direct causatives license a single instrument *-se* phrase - instrument$_{res}$; (b) bieventive indirect causatives license an additional *-se* phrase - instrument$_{caus}$; (c) the range of possible instruments$_{res}$ is constrained strongly by the meaning of the verb root as expressed by the caused event - I propose an alternative unified account of *-se* phrases which takes these distinctions into consideration. Under my proposal the indirect causative has the structure in (20).

(20)



Here the (stative) verb root composes with the theme argument. This VP is selected by an event argument **e'** introducing *v* head. The intermediate agent *-se* phrase is merged in the specifier of the *passive voice* head which selects this vP and is interpreted as the agent of the embedded event as in Bhatt & Embick (2003). However, following the arguments presented in Srishti (2011) I assume that *voice* and *v* are not bundled (see Pylkkänen 2008). The embedding *v* introduces a second event argument **e** and selects for the voice$_{pass}$P.

The configuration of two event arguments which cannot be identified with one another is interpreted through the CAUSE relation standardly defined as in (21).

(21) $[[\text{CAUSE}]] = \lambda f_{st} \lambda e_s \exists e'_s [CAUS(e, e') \& f(e')]$

With both the embedded and embedding *v* introducing event arguments into the semantics, instrument *-se* phrases get a uniform treatment since both instrument$_{res}$ and instrument$_{caus}$ are modifying *vPs* and are thereby interpreted as intersective event modifiers (as in Davidson 1967 and Parsons 1990). The interpretive differences between the two instruments - such that only the class of possible instruments$_{res}$ is determined independently by each predicate - follows from the structural asymmetry of the two *v* heads themselves and the consequent relativization of the interpretation of instruments to the different events.

4. Implications for causer subjects

In this section we look at the 'subjects'[1] of direct and indirect causatives with the aim of highlighting how my account of instruments in causatives derives restrictions on possible candidates in this slot. Beginning with direct causatives, we can see that these predicates permit animate *agents*, as well as inanimate eventive or instrument *causers* . This is in line with cross-linguistically attested patterns of there being a range of argument types that can be subjects (see Fillmore 1968, Parsons 1990, Schäfer 2012 inter alia). Thus, in direct causatives, instruments can be introduced either in *-se* phrases, see *caabi* 'key' in (22-b), OR subject positions, see (22-a), with no observable restriction other than a single utterance being barred from having both an instrument subject and an instrument *-se* phrase simultaneously in line with the Role Exhaustion principle.

(22) a. [anu / jinn ke jaadu / caabi]-ne taalaa khol-aa
anu / djinn GEN magic / hammer-ERG lock open.DC-PFV
'Anu / the djinn's magic/ the key opened the lock.'

b. anu-ne **caabi**-se taalaa khol-aa
John-ERG key-INST lock open.DC-PFV
'Anu unlocked the door with a key.'

In contrast, as first noted by Ramchand (2010), there seem to be some additional restrictions at play with respect to indirect causatives. In (23) *kettle* is not a suitable subject for the Indirect causative verb *ubal-vaa* 'cause to boil' even as it constitutes a licit subject for the Direct causative variant in the *boil* paradigm. This contrast may appear to be surprising given that other types of causers, for instance, eventive causers continue to be acceptable subjects in indirect causatives, (24).

[1]Diagnostics for 'subjects' in Hindi-Urdu: (a) Participation in the nominative/ergative split; (b) Binding of subject oriented anaphoric possessives; and (c) Anti-subject orientation effects for pronominal possessives.

(23) kettle-ne paani (jaldi-jaldi) [ubaal-aa / *ubal-vaa-yaa]
kettle-ERG water (quick-quick) boil-DC-PFV boil-IC-PFV
'The kettle boiled the water/ had the water boil quickly.'
(based on Ramchand 2010)

(24) paagalpan ke daure-ne ravi-se na jaane kya-kya kar-vaa-yaa
madness GEN bout-ERG ravi-INST not know what-what do-IC-PFV
'The bout of madness caused Ravi to do all sorts of things.'

Ramchand (2010) takes these empirical facts to suggest that inanimate and stative causers are systematically impossible as subjects of indirect causatives, and that in effect, the subject of a direct causative is *different* from the subject of the indirect causative, in that the former is a pure *initiator* while the latter must be an *undergoer-initiator*. However, there is reason to question this reading of the empirical observations given the additional data in (25) where we do find inanimate causers - *a glass of water* and *a lost key* - to be licit with indirect causatives.

(25) a. [**ek gilaas paani**-ne] kar-vaa-ya talaaq
One glass water-ERG do-IC-PFV divorce
'One glass of water caused there to be a divorce.'
Unavailable: 'A glass of water was the means of the divorce'
(Article Headline, *Patrika* News, Peter Hook, p.c. via Rajesh Bhatt)

b. (khoyi hui) **caabii**-ne karan-se taalaa khul-vaa-yaa
lost be.PFV key-ERG Karan-INST lock open-IC-PFV
'The (lost) key caused Karan to unlock the lock.'
Unavailable: 'The key was the means of the unlocking.'

Based on the data in (25), animacy and stativity of the causer appear not to be key here. I argue that the unlicensed instrument subjects in indirect causatives has its basis in the specific properties of the indirect causative. Recall the discussion in §3.1 which highlighted that indirect causatives can license two distinct kinds of instruments - an instrument$_{res}$ and an instrument$_{caus}$. Re-examining (23), where *kettle* is clearly an instrument$_{res}$, would then suggest that the restriction observed here is an interpretive one.

Furthermore, not all instrument subjects are banned in indirect causatives. Comparing (12) with (26) shows that the distributions of different subject types are better stated as in (27). The emergent generalization would then be of the form in (28).

(26) zamindar-ki **dhamkii**-ne Dakait-se mashaal-se ghar jal-vaa-yaa
Landlord-GEN threat-ERG bandit-INST torch-INST house burn-IC-PFV
'The landlord's threat caused the bandit to burn the house with a torch.'

(27) *Subjects of Causatives*

	Direct causative	Indirect causative
Animate agent	✓	✓
Eventive causer	✓	✓
Instrument$_{res}$	✓	×
Instrument$_{caus}$	-	✓

(28) Instruments$_{res}$ can be subjects of direct causatives but not an indirect causatives. Instrument$_{caus}$ can be subjects of indirect causatives.

Thus, rather than distinguishing between the basic properties of subjects themselves, I argue that the interpretation of an instrument subject is relativized to the event structure of the predicate in question. Given the bi-eventive analysis of indirect causatives and the semantic operation of existential closure, (29), the caused event is not available for modification once the CAUSE relation comes into play.

(29) Existential Closure (EC) saturates open argument positions by existentially quantifying over them.EC is the default mechanism for saturating event argument positions. (Davidson 1967 *via* Chung & Ladusaw 2004)

This means that all of the participant roles of the caused event - intermediate agent, instrument, affected object - are existentially closed if they remain unsaturated once the causation operation is applied and therefore cannot be modified further. Thus, any element merged into the structure at this stage will be evaluated only with respect to the causing event **e** and not the caused event **e'**. This allows us to account for the interpretive restriction exemplified in (25-b), where the only well-formed interpretation is one where *key* is understood to be modifying the causing event **e** i.e. as a causer of **e**, and interpreting the *key* to be modifying the caused event, that is the actual unlocking, is illicit.

5. Conclusion

This paper focused on two differences between *direct* and *indirect* causatives in Hindi-Urdu regarding their instruments. On examining the domain of *-se* phrases we saw that in addition to licensing an intermediate agent, which is the traditional identifier of indirect causation in the language, the indirect causative also allows an additional instrument *-se* phrase. This is in contrast to the unaccusative and the direct causative which are lacking in this regard. The availability of the instrument$_{caus}$ was argued to be an indicator of the bi-eventiveness of indirect causatives. In addition to *-se* phrases this paper also examined the restrictions on the 'subject' slot of direct and indirect causatives. It was observed that the direct causative permits its subject slot to be occupied by an instrument$_{res}$, while the indirect causative permits its subject slot to be occupied by an instrument$_{caus}$ but not an instrument that would ordinarily be construed to be modifying the result state. Thus, it was concluded that the complexity of indirect causation, in particular its bi-eventiveness,

predicts the distribution of causer subjects and additional differentiation of the agents of direct and indirect causatives is not required.

References

Bhatt, Rajesh, & David Embick. 2003. Causative derivations in hindi. Ms., University of Texas at Austin and University of Pennsylvania.

Chierchia, Gennaro. 2004. A semantics for unaccusatives and its syntactic consequences. In *The unaccusativity puzzle*, ed. E. Anagnostopoulou A. Alexiadou & M. Everaert, 22–59. Oxford: Oxford University Press.

Chung, Sandra, & William A. Ladusaw. 2004. *Restriction and saturation*. MIT Press.

Clauss, Michael. 2014. Subject orientation in hindi-urdu adverbial clauses. Ms., University of Massachusetts, Amherst.

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

Kellogg, Samuel H. 1876. *A grammar of the hindi language*. Thacker, Spink and Co.

Kratzer, Angelika. 1996. Severing the external argument from the verb. In *Phrase structure and the lexicon*, ed. J. Rooryck & L. Zaring, 109–137. Kluwer.

Kratzer, Angelika. 2005. Building resultatives. In *Event arguments in syntax, semantics, and discourse*, ed. C. Maienborn & A. Wöllstein-Leisten, 178–212. Niemeyer.

Masica, Colin. 1976. *Defining a linguistic area: South asia*. University of Chicago Press.

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

Ramchand, Gillian. 2008. *Verb meaning and the lexicon.*. Cambridge University Press.

Ramchand, Gillian. 2010. Licensing of instrumental case in hindi/urdu causatives. Ms., University of Tromsø, https://sites.google.com/site/gillianramchand01/multani.

Rissman, Lilia. 2013. Event participant representations and the instrumental role. Doctoral dissertation, Johns Hopkins University, Baltimore.

Saksena, Anuradha. 1982. *Topics in the analysis of causatives with an account of hindi paradigms*. University of California Press.

Schäfer, Florian. 2012. Two types of external argument licensing - the case of causers. *Studia Linguistica* 66:128–180.

Srishti, Richa. 2011. *Hindi verb classes and their argument structure alternations*. Cambridge Scholars Publishing.

Srishti, Richa. 2014. The case of causee in south asian languages (hindi, kashmiri, punjabi and manipuri). In *The lexicon-syntax interface: Perspectives from south asian languages*, ed. P. Chandra & R. Srishti, 197–216. Amsterdam: John Benjamins.

Williams, Alexander. 2015. *Arguments in syntax and semantics*. Cambridge University Press.

Sakshi Bhatia
sakshibhatia@umass.edu

Perspectival reflexivity (or what makes reflexives special): a case study from Tamil*

Sandhya Sundaresan

University of Leipzig

1. Overview

The goal of this paper is to analyze the nature of the perspectival dependency between an anaphor and its antecedent when the two are arguments of the same predicate (i.e. are co-arguments) — using the Dravidian language, Tamil, as a case-study. I will henceforth reserve the term "reflexivity" for this type of relation. Like cases of reflexivity in many languages (see Reinhart & Reuland 1993, Jayaseelan 1997, Reuland 2001b, 2011, for an overview), this dependency is distinguished from other cases of anaphora in the language by being specially marked. This in turn suggests that reflexivity is special and requires recourse to additional grammatical devices beyond what is needed by other types of anaphora where the antecedent and anaphor are not co-arguments.

Reflexive structures in many dialects of Tamil, and in other Dravidian languages like Kannada (see, for instance, work by Lidz 2001, 2004, et seq.) are obligatorily marked by a morpheme "*koḷ*" which is suffixed onto the predicate which the anaphor and its antecedent are arguments of. The sentence in (1) shows a non-reflexive sentence which is licit in the absence of *koḷ*. The minimal pair in (2)-(3) shows reflexive variants of this sentence without and with *koḷ*, respectively, and illustrates that *koḷ* cannot be licitly omitted in a standard reflexive construction:

(1) Kalpana Siva-væ kiḷḷ-in-aaḷ.
Kalpana.NOM Siva-ACC pinch-PST-3FSG
"Kalpana pinched Siva."

(2) * Kalpana$_i$ tann-æ$_i$ kiḷḷ-in-aaḷ.
Kalpana.NOM ANAPH-ACC.SG pinch-PST-3FSG

*I am grateful to Tom McFadden for discussing various aspects of this paper with me on numerous occasions and to the audiences at FASAL 5 at Yale University, at the Perspectivization Workshop at GLOW 39, Göttingen and at the Linguistics Colloquium at the IIT Delhi, where versions of the ideas herein were presented, for their feedback and questions. Special thanks to John Beavers for his very detailed and helpful comments on an earlier draft, which helped in no small measure to improve the current one, and to Rahul Balusu for his (co-)editorial patience. This paper presents a revised version of an analysis of perspectival reflexivity developed in my dissertation (see Sundaresan 2012, particularly the chapters in Part III). While much of the core data and theoretical background remain the same, the main analysis has been extended and (hopefully) improved in significant ways, in particular by sharpening the discussion of perspective-holding with testable inference patterns and discourse contexts, data involving non-anaphoric perspectival expressions and perspective-shifting, and motivating the ban on co-argument antecedence in purely structural terms.

"Kalpana$_i$ pinched herself$_i$." (Intended)

(3) Kalpana$_i$ tann-æ$_{\{i,*j\}}$ kiɭɭ-i-ko-ɳɖ-aaɭ.
Kalpana.NOM ANAPH-ACC.SG pinch-ASP-*koɭ*-PST-3FSG
"Kalpana$_i$ pinched herself$_{\{i,*j\}}$."

Non-reflexive anaphora in Tamil — i.e. structures where the anaphor and its antecedent are not co-arguments as in cases of long-distance anaphora and logophora — successfully obtains even in the absence of *koɭ*, however. In (4), (2) is embedded under an attitude verb; unlike (2), however, the resulting complex sentence in (4) is perfectly grammatical. Reflexive anaphora is still ruled out: i.e. *Kalpana* still cannot antecede the anaphor *ta(a)n*; but the matrix subject *Siva* denoting the attitude-holder may licitly "long-distance" antecede the anaphor, despite the absence of *koɭ*:

(4) Siva$_i$ [$_{CP}$ Kalpana$_j$ tann-æ$_{\{i,*j\}}$ kiɭɭ-in-aaɭ-ŭnnŭ] nene-tt-aan.
Siva Kalpana ANAPH-ACC.SG pinch-PST-3FSG-COMP think-PST-3MSG
"Siva$_i$ thought that Kalpana$_j$ pinched him$_{\{i,*j\}}$."

That said, nothing prevents *koɭ* from being present in such structures. Thus, we could come up with a minimal variant to (4) — as in (5) below — which differs from (4) only in that there is a *koɭ*-morpheme marking the embedded verb:

(5) Siva$_i$ [$_{CP}$ Kalpana$_i$ tann-æ$_{\{i,j\}}$ kiɭɭi-ko-ɳɖ-aaɭ-ŭnnŭ] nene-tt-aan.
Siva Kalpana ANAPH-ACC.SG pinch-PST-3FSG-COMP think-PST-3MSG
"Siva$_i$ thought that Kalpana$_j$ pinched him$_i$/herself$_j$."

This sentence is also grammatical, with the only difference lying in the range of possible antecedents for the anaphor *ta(a)n*. Where in (4) the matrix attitude-holder *Siva* is the only possible antecedent, in (5), both *Siva* and the co-argument *Kalpana* are possible antecedents for the anaphor. Given our prior observation that *koɭ* makes reflexive antecedence possible, this indeed is exactly what we expect. Taken by themselves, the minimal pairs presented in (2)-(3) and (4)-(5) suggest that anaphoric dependencies show a clear demarcation with respect to their distribution with *koɭ*: co-argument anaphora (or reflexive dependencies) requires the concomitant presence of *koɭ*, but all other types of anaphoric dependency do not.

However, reflexive structures involving psych predications such as those in (6) and (7) complicate this simple, binary picture. Consider the minimal pairs below:

(6) PSYCH REFLEXIVE WITH DATIVE SUBJECT:

a. Kalpana-vŭkkŭ$_i$ tann-æ$_{\{i,*j\}}$ piɖikkæ-læ.
Kalpana-DAT ANAPH-ACC.SG like-NEG
"Kalpana$_i$ didn't like herself$_{\{i,*j\}}$."

b. * Kalpana-vŭkkŭ$_i$ tann-æ$_i$ piɖi-ttŭ-kkoɭɭæ-læ.
Kalpana-DAT ANAPH-ACC.SG like-ASP-*koɭ*-NEG
"Kalpana$_i$ didn't like herself$_i$." (Intended)

(7) PSYCH REFLEXIVE WITH NOMINATIVE SUBJECT:

a. Abinaya$_i$ tann-æ$_{\{i,*j\}}$ virŭmbŭ-gir-aaɭ.
Abinaya.NOM ANAPH-ACC.SG love-PRS-3FSG
"Abinaya$_i$ loves herself$_{\{i,*j\}}$."

b. * Abinaya tann-æ$_i$ virumb-i-*koɭ*ɭŭ-gir-aaɭ.
Abinaya.NOM ANAPH-ACC.SG love-ASP-*koɭ*-PRS-3FSG
"Abinaya$_i$ loves herself$_i$." (Intended)

The minimal pairs in (6) and (7) involve reflexive structures with "quirky" dative and nominative subjects, respectively. These show precisely the opposite behavior with *koɭ* from that exhibited by non-psych reflexives like those in (2)-(3): i.e. reflexive anaphora obtains in the obligatory *absence* of *koɭ*. It can, furthermore, be shown that the ban on *koɭ* stems not from restrictions imposed by the reflexive dependency, but from properties of the psych predicates. This can be gleaned from the fact that *koɭ* is disallowed in the non-reflexive counterparts of the psych predications in (6)-(7), as well, as shown below:

(8) Kalpana-vŭkkŭ Siva-væ piḍikkæ-læ/*piḍittŭ-kkoɭɭæ-læ.
Kalpana-DAT Siva-ACC like-NEG/*like-ASP-*koɭ*-NEG
"Kalpana didn't like Siva."

(9) Abinaya Dhanush-æ virŭmbŭ-gir-aaɭ/*virŭmb-i-kkoɭ-gir-aaɭ.
Abinaya.NOM Dhanush-ACC love-PRS-3FSG/love-ASP-*koɭ*-PRS-3FSG
"Abinaya loves Dhanush."

We thus have potentially three classes of anaphora in Tamil: Structures involving standard reflexives (i.e. reflexives under non-psych predicates) which require the presence of *koɭ*, those involving non-reflexive anaphora (i.e. long-distance anaphora and logophora) which obtain in the absence of *koɭ* (though its presence is not banned), and those involving reflexives under psych predicates which require the absence of *koɭ*(though, again, this ban on *koɭ* seems to be independent of reflexivity and driven by an incompatibility with psych predicates more generally).

The central question that this paper asks is why reflexivity alone (as opposed to other types of anaphora in this language) requires special marking in the form of *koɭ*, and how this relates to the nature and grammatical representation of perspective. A corollary point of investigation has to do with understanding what makes perspectival reflexivity crosslinguistically special. The line of argumentation that I will pursue here consists of the following analytic pieces. The basic idea that I will motivate is that grammatical perspective is structurally instantiated. In languages with perspectival anaphora, the anaphor must be syntactically bound within its local Perspectival Phrase or PerspP (a binding domain with the additional restriction that the binder has to denote a perspective-holder). The antecedent, on the other hand, must be outside the perspectival domain. Reflexives, I will then propose, fundamentally differ from other types of anaphora in the following way: they instantiate the only structure where the intended antecedent is also a co-argument of the anaphor. This has the consequence that, in reflexives, the antecedent is also contained inside the local PerspP of the anaphor, which yields an anti-locality effect. Many languages, I believe, simply avoid such a configuration altogether — which may in turn help explain why perspectival reflexives are typologically so uncommon. Other languages, like Tamil, have recourse to special means for modifying the offending configuration (thereby allowing perspectival reflexivity to obtain after all). This, I propose, is precisely what the addition of *koɭ* helps to do.

First, I will show that reflexivity in Tamil (both with *koɭ* and in the context of psych predicates) is perspectivally regulated, like other types of anaphora in this language. For the standard *koɭ* reflexives, given the model of anaphora motivated here, this means that *koɭ* selects a PerspP in its complement. Second, I will argue that *koɭ* spells out a head with an affectedness semantics

above a resultative aspectual head (Asp_{res}) which itself is merged above Kratzer (1996)'s Voice. Third, *koɭ* thematically raises (in the sense of Ramchand 2008) the external argument in the Spec of Voice to its own specifier and assigns it another θ-role. These observations have the following consequences. If the raised external argument is the co-argument DP of an anaphor in a reflexive structure, this has the (entirely epiphenomenal) consequence that, enroute to being raised to Spec, *koɭ*P, this DP now escapes the binding domain (PerspP) it previously shared with the anaphor. From its new raised position in Spec, *koɭ*P, the external argument can thus now antecede the anaphor without violating anti-locality. Psych predicates, I suggest, are already lexicalized with this much functional structure, and thus the addition of *koɭ* is not necessary to license reflexivity.

2. The perspectival nature of anaphora in Tamil

Long-distance anaphora, in languages that display this phenomenon, has typically been characterized as being "subject-oriented" in the literature (see Koster & Reuland 1991, and the citations therein for an initial description). Such a characterization was supposed to capture the restriction that such anaphors could typically be anteceded by syntactic subjects but not by objects in a number of languages, like Icelandic (Sigurðsson 1990, 2010, Reuland 2001a, a.o.), Italian (Bianchi 2003, Giorgi 2006, 2010), Malayalam (Jayaseelan 1997), Chinese (Huang & Tang 1991, Huang & Liu 2001), Norwegian (Hellan 1988) and others.

At first glance, non-local anaphora in Tamil also seems to be subject-oriented in this manner. Thus, in (10), the medial object *Kristin* may not antecede *ta(a)n* in the unmarked discourse scenario; only the superordinate subjects *Sandhya* and *Sudha* may do so:

(10) [$_{CP}$ Sandhya$_i$ nene-tt-aaɭ [$_{CP}$ Sudha$_j$ Kristin-kiʈʈæ$_k$ [$_{CP}$ Champa
Sandhya.NOM think-PST-3FSG Sudha.NOM Kristin-OBL Champa.NOM
tan$_{\{i,j,*k\}}$ viiʈʈŭ-kkŭ mu:ɳŭ maasatt-ŭkkŭ var-ŭv-aaɭ-ŭnnŭ
ANAPH.GEN house-DAT three month-DAT come-FUT-3FSG-COMP
so-nn-aaɭ-nnŭ]]].
say-PST-3FSG
'Sandhya$_i$ thought [$_{CP}$ that Sudha$_j$ told Kristin$_k$ [$_{CP}$ that Champa will come to her$_{\{i,j,*k\}}$ house for three months]]'

However, there are systematic exceptions to this subject-restriction: an observation that is incidentally crosslinguistically robust, having been made for anaphoric systems in Italian (Giorgi 2006, 2010), Malayalam (Jayaseelan 1997), Japanese (Sells 1987) and even English (Minkoff 2003) among others, as well. In psych-predications, for instance, the chosen antecedent is not the syntactic subject, but the experiencer object, as in (11):

(11) [$_{CP}$ [$_{DP}$ Taan$_{\{i,*j\}}$ avvaɭavŭ eeɻæ-jaagæ irŭnd-adŭ] Raman-æ$_i$
ANAPH[NOM] so poor-ADJ be-PST-3NSG.NOM Raman-ACC
rombæ-vee baadi-jirŭ-kkir-adŭ.]
very-EMPH affect-be-PRS-3NSG
"[$_{DP}$ His$_{\{i,*j\}}$ having been so poor] has really affected Raman$_i$ very much."

(11) shows that it is not necessary to be a subject to antecede an anaphor in Tamil. (12) shows that it is not sufficient either — if the subject is non-sentient, it cannot antecede (unless, of course, it is anthrophomorphized, as in a fairy-tale scenario):

(12) Tan-akkŭ$_{\{i,*j\}}$ pinnaalæ iru-nd-æ maratt-æ koɻendæ$_i$/*vaɳɖi$_i$ iɖi-tt-adŭ.
ANAPH-DAT behind be-PST-REL tree-ACC child.NOM/car.NOM hit-PST-3NSG
"[The child]$_i$ hit [$_{DP}$ the tree [$_{CP}$ that was [$_{PP}$ behind itself$_{\{i,*j\}}$]]]."
"*[The car]$_i$ hit [$_{DP}$ the tree [$_{CP}$ that was [$_{PP}$ behind itself$_i$]]]." (Intended)

Part of what made the subject-orientation proposal appealing was that it lent itself readily to a standard syntactic analysis of long-distance binding (see e.g. the I-to-I movement analyzes of Pica 1987, Huang & Tang 1991, and relativized subject hypothesis of Manzini & Wexler (1987), Progovac (1993)). But such an approach is sharply undermined by logophoric sentences like that in (13) below — where the anaphor refers to a discourse-salient antecedent across the sentence boundary (see e.g. Clements 1975, Sells 1987, Kuno 1987, Hellan 1988, Koopman & Sportiche 1989, Sigurðsson 1990, Pearson 2013, a.o. for crosslinguistic evidence to the same effect):

(13) Koɻændæ$_i$ aɻŭ-d-adŭ. Tan-akkŭ$_{\{i,*j\}}$ romba pasittadŭ.
child.NOM cry-PST-3NSG. ANAPH-DAT very hungry.
"[The child]$_i$ wept. It$_{\{i,*j\}}$ was very hungry."

In fact, the unifying property of anaphoric antecedence in Tamil (and potentially also in languages like Icelandic, Italian, Malayalam, Japanese, and others) is not syntactic subjecthood, but perspective-holding: in particular, "A potential antecedent of *ta(a)n* is a nominal which [denotes an individual that has] a mental, temporal or spatial perspective with respect to a CP, PP, or DP in which the anaphor is a participant (i.e. thematic argument)" (Sundaresan 2012, 70, see also Sundaresan 2016b). To understand what this means more concretely, consider a simplified version of the sentence in (10), as in (14) below:

(14) [$_{CP}$ Sudha$_j$ Kristin-kiʈʈæ$_k$ [$_{CP}$ Champa tan$_{\{j,*k\}}$ viiʈʈŭ-kkŭ mu:ɳŭ
Sudha.NOM Kristin-OBL Champa.NOM ANAPH.GEN house-DAT three
maasatt-ŭkkŭ var-ŭv-aa[-ŭnnŭ] so-nn-aa[].
month-DAT come-FUT-3FSG-COMP say-PST-3FSG
'Sudha$_j$ told Kristin$_k$ [$_{CP}$ that Champa will come to her$_{\{j,*k\}}$ house for three months]'

(14) combines the use of the anaphor *ta(a)n* in the innermost clause with that of another perspective-sensitive item, namely 'come'. Relative locative expressions like 'come' and 'go' have long been known to be perspective-sensitive items (or PSIs) in the sense that the truth or falsity of propositions containing such expressions is relative to the point-of-view or perspective of a perspective-center (PC) or judge (see Mitchell 1986, Fillmore 1997, Oshima 2006, a.o. for discussion). Thus, given that I live in Leipzig, I cannot utter (16); I must say (15); however, when embedded under an attitude-verb, either 'come' or 'go' may be used (17) (as long as the attitude-holder — in this case *Champa* — does *not* live in Leipzig):

(15) Sudha is coming to Leipzig next month.
(16) # Sudha is going to Leipzig next month.
(17) Champa said [$_{CP}$ that Sudha is going/coming to Leipzig next month]

I'll assume the following discourse context for (14): Sudha, like her friend Kristin, lives in Berkeley; Champa lives in Chennai; the sentence is uttered by me, in Leipzig. To understand the relevance of perspective for *ta(a)n*, we need to contrast this sentence with the minimally varying one in (18):

(18) * [$_{CP}$ Sudha$_j$ Kristin-kiʈʈæ$_k$ [$_{CP}$ Champa tan$_{\{j,*k\}}$ viiʈʈŭ-kkŭ mu:ɳŭ
Sudha.NOM Kristin-OBL Champa.NOM ANAPH.GEN house-DAT three
maasatt-ŭkkŭ poo-v-aa[-ŭnnŭ] so-nn-aa[].
month-DAT go-FUT-3FSG-COMP say-PST-3FSG
'Sudha$_j$ told Kristin$_k$ [$_{CP}$ that Champa will go to her$_j$ house for three months]'

In (18), the PSI 'come' has been replaced with another, namely 'go'. But this yields ungrammaticality. What's even more interesting is that such ungrammaticality does not obtain when 'go' co-occurs with a coreferent pronoun instead of the anaphor *ta(a)n*, as in (19):

(19) [$_{CP}$ Sudha$_j$ Kristin-kiʈʈæ$_k$ [$_{CP}$ Champa aval$_{\{j,*k\}}$ viiʈʈŭ-kkŭ mu:ɳŭ
Sudha.NOM Kristin-OBL Champa.NOM she.GEN house-DAT three
maasatt-ŭkkŭ poo-v-aa[-ŭnnŭ] so-nn-aa[].
month-DAT go-FUT-3FSG-COMP say-PST-3FSG
'Sudha$_j$ told Kristin$_k$ [$_{CP}$ that Champa will go to her$_j$ house for three months]'

These patterns are precisely what we expect if *ta(a)n* is also a PSI and its antecedence governed by the condition on perspective-holding described above. (19) is fine because 'go' can be interpreted from the spatial perspective of the utterance-context speaker (me) who lives in Leipzig, and the deictic pronoun *aval* ('she') places no independent perspectival restrictions. In contrast, (18) is ungrammatical because 'go' again must be interpreted from the spatial perspective of me in Leipzig, but the use of *ta(a)n* places independent perspectival restrictions that force the chosen antecedent (*Sudha*) to be the perspective-holder. We thus have two clashing perspectives in a local domain, yielding ungrammaticality. Finally, in (14), there is no clash at all, because the locative PSI used here is 'come' not 'go', which is compatible with *Sudha*, the chosen antecedent of the anaphor, being the perspective-holder.[1]

The mental perspective-holding condition on antecedence correctly explains the subject-orientation in the unmarked discourse scenario (seen in sentences like (10)) and excludes the syntactic objects because, for *independent* reasons having to do with how grammatical functions are mapped onto thematic roles, subjects — functioning as Agents and Experiencers — tend to denote perspective-holders in natural language (see Mitchell 1986, for discussion of the correlation between θ-roles and perspective-taking). The advantage of this approach is that it can also be extended to account for antecedence in structures involving logophoric dependencies, as in (13): here again, the antecedent is the entity denoting the mental perspective holder with respect to the proposition containing the anaphor. The object antecedence in (11) is also no longer puzzling: the experiencer object denotes the mental perspective holder with respect to the predication containing the THEME anaphor, so this is the chosen antecedent for the anaphor in the unmarked discourse scenario. Finally, the ban on non-sentience is also explained, assuming that perspective-holding requires some kind of sentience. Building on prior work concerning the semantics of self-ascription, Sundaresan & Pearson (2014) propose that all perspectival predicates quantify over elements of a set that are designated by a sentient entity as candidates for the actual time, location or world of that entity. The

[1] For detailed data and discussion showing that PSIs in a local domain must "shift together", i.e. must always denote the same perspective-holder — akin to the Shift Together generalization proposed for shifted indexicals (Anand & Nevins 2004) — see Bylinina et al. (2014).

difference between spatial, temporal, and attitudinal/psych predicates, lies merely in the choice of this coordinate.[2]

Formal, theoretical accounts of perspectival anaphora still remain relatively sparse in the literature and, as far as I am aware, focus primarily on the semantic and pragmatic aspects of this phenomenon. Part of what makes the Tamil data relevant is that it provides evidence that perspectival dependence on anaphora must already be encoded at the level of syntax. Such evidence comes from verbal agreement. Normally, agreement on the verb is triggered by the nominal marked nominative; but when this nominative DP is the anaphor *ta(a)n*, the agreement is not obviously triggered by the anaphor (e.g. the agreement may be 1st person even though the anaphor itself may never take 1st-person antecedents); it is also not triggered by the antecedent of the anaphor (which may have different ϕ-features from the agreement and also be non-local to it); nonetheless, it seems to track the antecedent (i.e. its features vary as a function of its identity).[3] So I conclude that, in these special cases, the agreement must be triggered by a third element in the local domain — e.g. a silent pronoun (or *pro*). But in addition to triggering agreement, this *pro* must also mediate the dependency between the anaphor and its antecedent: this would explain why the agreement tracks the antecedent. Given that anaphora in Tamil is perspectival, this *pro* must then also be perspectival: more broadly, then, perspectival information must be syntactically accessible (and available at the point of triggering ϕ-agreement).

In particular, I propose that the perspectival *pro* is introduced in the specifier of a Persp head (in a Perspectival Phrase or PerspP) and encodes the perspectival center (PC). Although it is by default set to denote the utterance-context speaker, it can shift to denote other salient perspective-holders under relevant circumstances, e.g. in attitude complements. In clauses with a successfully bound perspectival anaphor it denotes the individual corresponding to the antecedent. The real binder of the anaphor is then not the apparent antecedent but the *pro* in its local Spec, PerspP (see Koopman & Sportiche 1989, for an earlier proposal along these lines), which then counts as its local binding domain. This *pro* mediates the relationship between the anaphor and its antecedent in the evaluation context, which thus corefer by transitivity (see Sundaresan 2012, 2016b, for a lot more detail, data and discussion of these points, and Sundaresan 2016a for more on the agreement patterns).

A central further component of the proposal, which will turn out to be crucial in the pages ahead, is that there is one unique Persp (and thus one unique perspectival *pro*) per (structural) domain (see

[2]Thus, a PSI like 'behind' in (12) would have a lexical entry like that in (i), with *Khorastic* being analogous to *Doxastic*, but for locations (Sundaresan & Pearson 2014, 15):

i. $[\![behind]\!]^{c,x,t,w,g} = \lambda x \lambda y \lambda z. \forall s' [s' \in Khorastic_{x,w,t} \rightarrow y$ is behind x in w at t relative to $s']$. Where:
   a. $Khorastic_{x,w,t} = Khorastic_{perceptual_{x,w,t}}$ or $Khorastic_{imagined_{x,w,t}}$;
   b. $Khorastic_{perceptual_{x,w,t}} = \{s'$: it is compatible with x's perceptual experience in w at t for $s_{x,w,t}$ to be $s'\}$, where $s_{x,w,t}$ = the spatial coordinate of x in w at t.
   c. $Khorastic_{x,w,t} = Khorastic_{imagined_{x,w,t}}$ only if:
      a. $x = Speaker(c)$, and
      b. there is some contextually salient entity u such that for every element s' of $Khorastic_{imagined_{x,w,t}}$, it is compatible with what x believes in w at t for u to be located at s'.

[3]Specifically, 1st-person agreement on the verb is triggered only when the anaphoric antecedent is the agent of a speech predicate in the immediately superordinate clause. In all other cases, obligatory 3rd-person agreement is triggered on the verb. What this shows is that the 1st-person agreement is somehow sensitive to the structural position and thematic properties of the anaphoric antecedent though not its actual ϕ-features.

Bylinina et al. 2014, for crosslinguistic evidence from perspective shifting for this point). What counts as a perspectival domain may well be parametrized. With respect to Tamil, at least (certain) CPs, PPs, and DPs count as perspectival domains. I will turn to the status of VP/*v*P below, since these will be central to the discussion of reflexives. Evidence for this comes from the fact that *ta(a)n* can be perspectivally anteceded across all these domains, and that anaphors can be shown to have different antecedents just in case they are in different domains. Under the proposal briefly sketched here, this means that the extended projections of (certain) CPs, PPs, and DPs must contain Persp. More recently, Nishigauchi (2014) and Charnavel (2015) have argued, on the strength of data from "empathic" binding in Japanese and "exempt" anaphora in French, respectively, that grammatical perspective must be syntactically represented in this manner in these languages as well.

3. Back to reflexivity in Tamil

We have just seen that anaphora in Tamil is perspective-driven: i.e. a DP cannot serve as an antecedent unless it denotes a perspective-holder along the mental or spatio-temporal dimensions toward the minimal PerspP containing the anaphor. We can now return to cases of reflexivity in this language and see how they fare against this baseline.

3.1 Reflexivity is also perspectival

A survey of the descriptive conditions on antecedence in reflexive structures in Tamil makes it apparent that reflexivity, too, is perspective-driven. First, the nominals that are allowed to serve as reflexive antecedents in *koɭ* constructions are AGENTS (as in (3) or EXPERIENCERS (as in (6a) and (7a): DPs that, by virtue of their thematic roles, readily denote perspective holders (the latter invariably along the mental dimension, the former along the mental or spatio-temporal dimensions) in the unmarked discourse scenario. The non-sentience restriction on antecedence, observed in cases of long-distance anaphora (see again (12)) obtains in reflexive structures in Tamil as well. (21) is degraded to the point of ungrammaticality in the discourse scenario in (20):

(20) Scenario: There is a vibrating alarm clock on a small, rickety bedside table. This morning, the alarm clock vibrated violently and, as a consequence of its own vibrations, slid to the edge of the table and fell down to the floor.

(21) * Gaɖigaaram$_i$ kiiɻæ viɻ-ŭndŭ tann-æ$_i$ tuḷḷam-tuḷḷam-aagæ
clock[NOM] down fell-ASP ANAPH-ACC small-small-ADJ
oɖe-čču-ko-ɳɖ-adŭ.
smash-ASP-*koɭ*-PST-3NSG
"[The clock]$_i$ fell down and smashed itself$_i$ into smithereens." (Intended)

But if the clock in (21) is magically made to come alive, as in context of the *Beauty and the Beast* fable, say — a reading we can accentuate by replacing 'fall' with an agentive verb like 'jump' — the sentence becomes licit; the same DP *gaɖigaaram*, denoting this anthropomorphized, suicidal clock, may now indeed antecede the anaphor:

(22) Gaɖigaaram$_i$ kiiɻæ kudi-čč-ŭ tann-æ$_{\{i,*j\}}$ tuḷḷam-tuḷḷam-aagæ
clock[NOM] down jump-ASP ANAPH-ACC small-small-ADJ
oɖe-čču-ko-ɳɖ-adŭ.
smash-ASP-*koɭ*-PST-3NSG

"[The clock]$_i$ jumped down and smashed itself$_{\{i,*j\}}$ into smithereens."

Assuming, as we did earlier, that non-sentience is banned because it is incompatible with perspective-holding, this again underscores the relevance of perspective-holding for reflexivity in Tamil. For these reasons, I will propose that reflexivity, or co-argument anaphora, just like all other kinds of anaphoric dependency in Tamil, is perspective-driven. That is, the antecedent of the anaphor *ta(a)n* must denote some individual who holds a perspective, mental and/or spatio-temporal, towards some predication containing the anaphor. Given the background of structural perspective described so far, this entails the following:

(23) The complement of *koḷ* contains a Perspectival Phrase (or PerspP) with a perspectival *pro* that locally binds the anaphor. In *koḷ*-reflexives, the co-argument of the anaphor corefers with this perspectival *pro*.

Right at the outset, it should be noted that the idea that reflexivity in Tamil is perspectival describes a somewhat striking state of affairs. Reflexivity in English, for instance, does not seem to be perspectival in the same way. Thus, under the scenario given in (20), the English counterpart of the Tamil sentence in (21) would be perfectly licit. But I believe that English and English-like languages are the norm, while Tamil and Tamil-like languages are the exception. Reflexives in Tamil are able to be perspectival only because it has recourse to the morpheme *koḷ*.

3.2 A structural restriction: or why perspectival reflexivity is special

We started this paper with the observation that, in the general case, reflexivity in Tamil obtains only in the presence of the *koḷ* morpheme suffixed to the main predicate, yielding the minimal pair repeated below:

(24) * Kalpana$_i$ tann-æ$_{\{i,*j\}}$ kiḷḷ-in-aaḷ.
Kalpana.NOM ANAPH-ACC.SG pinch-PST-3FSG
"Kalpana$_i$ pinched herself$_i$." (Intended)

(25) Kalpana$_i$ tann-æ$_{\{i,*j\}}$ kiḷḷ-i-kko-ɳɖ-aaḷ.
Kalpana.NOM ANAPH-ACC.SG pinch-ASP-*koḷ*-PST-3FSG
"Kalpana$_i$ pinched herself$_{\{i,*j\}}$."

Informally, (24) shows that the co-argument of the anaphor is unable, by itself, to antecede the anaphor: i.e. *Kalpana* is unable to denote a perspective-holder relative to some predication (whatever this may be) containing the anaphor. (25) shows that the addition of *koḷ* fixes this problem: i.e. in (25), *Kalpana* is suddenly able to denote a perspective-holder with respect to some predication containing the anaphor, thus is now able to antecede it. In Section 2, I argued that the perspectival *pro* in Spec, PerspP corefers with the antecedent in the evaluation context and binds the anaphor *ta(a)n* at LF; thus, the anaphor and antecedent corefer indirectly, by transitivity. Formally, therefore, the observations regarding (24)-(25) entail (26):

(26) The perspectival *pro* that binds *ta(a)n* and a co-argument of the anaphor typically cannot corefer. Exceptions: *koḷ* reflexives; psych predications.

I propose that (26) ensues from a seemingly trivial (in fact, definitional) property of reflexives — namely that they constitute the only instance where the antecedent of the anaphor is also its

co-argument. Relevant empirical evidence has been brought to bear in recent work by Bylinina et al. (2014), Bylinina & Sudo (2015) based on crosslinguistic data involving perspective-shifting with respect to various structural domains, which can help us figure out why this might yield (26). In particular, they argue that VP, i.e. the constituent containing the V + internal argument (but not the external argument) is not a shifting domain because, when a PSI appears as the main predicate, it cannot shift its perspectival center to the subject of that sentence.[4] Under the current proposal, this would translate into the following:

(27) There is no Persp between VP and Voice, i.e. between the internal and external arguments. In reflexive structures, the anaphor and its co-argument are contained inside the same minimal PerspP.

I would like to argue that this essentially leads to an anti-locality effect, where the anaphor and its co-argument are simply too close together for the latter to serve as antecedent. To see why, consider what would happen if, in a sentence like (24), the co-argument of the anaphor *Kalpana could* indeed antecede the anaphor. I.e. *Kalpana* would corefer with the perspectival *pro* (that binds *ta(a)n*) in Spec, PerspP, in contradiction of (26). But we have just seen in (27) that the anaphor and its co-argument are both contained inside the same local PerspP. This means that *Kalpana* would not only corefer with the perspectival *pro* in the evaluation context, it would also be asymmetrically c-commanded by it in the structure. This would lead to a Condition C violation. Replacing *Kalpana* with a pronoun like *avaḷ* ('she') wouldn't improve matters much, because it would yield a Condition B violation instead. Either way, the derivation would crash. To generalize:[5]

(28) **Explanation for (26)** (but not its exceptions): In perspectival anaphora, an antecedent is a nominal that corefers with the perspectival *pro* in Spec, PerspP that binds the anaphor. In reflexive structures, the intended antecedent is also the co-argument of the anaphor, and is thus contained in the same local PerspP as the anaphor (27). So, the antecedent not only corefers with *pro*, it is also asymmetrically c-commanded by it, yielding violations of Conditions B or C, causing a crash.

The goal of the rest of this paper is to explain the exceptions to this rule, in particular to investigate what special properties *koḷ* brings to the table that allow the anti-locality ban on reflexivity to be lifted. At the end of the paper I will turn, though more briefly, to the second kind of exception, namely that of psych reflexives.

[4]The authors provide examples like "John is handsome", where the (perspectival) TASTE- predicate 'handsome' has to be evaluated from the utterance-context speaker's perspective and cannot be evaluated from that of John. In contrast, in "If a handsome man comes in, John will be startled", the PSI 'handome' is ambiguous and may be evaluated either from the speaker's perspective or from John's.

[5]It must, of course, be noted that violations of Conditions B and C are tolerated to a much greater degree than are violations of Condition A, and can be significantly ameliorated by factors like contrastive focus. This has led to speculation that the former do not involve transgressions of narrow-syntactic principles but of interface conditions or, perhaps, even Neo-Gricean principles. A discussion of these issues is outside the scope of this paper (see Hicks 2009, for an excellent summary). What is relevant here is, simply, that having the co-argument be properly embedded within the minimal perspectival domain creates a configuration that is independently dispreferred, however this may be implemented.

4. Interactions between *koḷ* and reflexivity

The argumentation in (28) allows us to clarify in structural terms the kind of repair strategy that the presence of *koḷ* must make possible, which in turn feeds reflexivity. In *koḷ*-reflexives like (25), the co-argument and the anaphor must not be contained inside the same local PerspP. Rather, the co-argument must be outside the minimal PerspP containing the anaphor. Here, I will argue that this becomes possible because *koḷ* has the following properties. It spells out a head above Voice, and introduces below it a PerspP that is also above Voice. It is a thematic raising predicate (Ramchand 2008) with an affectedness semantics: it raises the external argument from Spec, VoiceP to its own Spec. In *koḷ*-reflexives, this external argument is the co-argument of the anaphor. But in its new raised position in Spec, *koḷ*P, it is no longer asymmetrically c-commanded by the *pro* in the Spec, PerspP introduced below *koḷ*. There is thus no Condition B/C violation when *pro* and the co-argument corefer. Such coreference may thus licitly obtain, yielding perspectival reflexivity.

4.1 The structural properties of *koḷ*

In this section, I will motivate the first piece of the proposal, namely that *koḷ* instantiates a head above VoiceP, so that the PerspP it introduces is also above VoiceP.[6] As part of this, I will briefly show that *koḷ* is not itself an instantiation of Voice.

As it happens, *koḷ* may also be suffixed onto unaccusatives, as in (29) below:

(29) Marakkiḷæ (sumaj-læ) vaḷæ-nʧŭ-kko-ɳʈ-adŭ.
Tree branch.NOM weight-LOC bend-ASP.INTR-*koḷ*-PST-3NSG
'The tree branch became bent (under its weight).' (Rough translation)

The distribution of *koḷ* in Tamil thus seems initially reminiscent of the (partial) syncretism between reflexive and unaccusative structures observed in Greek, Slavic, Romanic, and German (Embick 2004, Sportiche 1998, Schäfer 2008, Medová 2009, a.o., and see Lidz 2001 for such a proposal for *koḷ* in the related Dravidian language, Kannada). However, closer inspection reveals that the distribution of *koḷ* in Tamil is independent of the valency of the predicate. First, *koḷ* is actually optional on unaccusatives (29):

(30) Marakkiḷæ sumaj-læ vaḷæ-nʧ-adŭ.
Tree branch.NOM weight-LOC bend-ASP.INTR-3NSG
"The tree branch bent (under its weight)."

Second, *koḷ* not only appears on reflexives and unaccusatives but may also be suffixed, again optionally, onto non-reflexive transitives (31):[7]

(31) Sudha marakkiḷæ-jæ vaḷæ-ččŭ-kko-ɳʈ-aaḷ.
Sudha.NOM tree branch-ACC bend-ASP.TR-*koḷ*-PST-3FSG
'Sudha bent the tree-branch.'

Third and finally, there is morphological evidence that *koḷ* spells out a head distinct from Voice, coming from gemination yielding morphophonological voicing contrasts in the verbal stem (see Sundaresan & McFadden To appear, for data and discussion). Consider (32) below:

[6] Recall that there can be no PerspP *under* VoiceP.

[7] In the example in (31), I explicitly gloss the transitivity morpheme as such as this is the focus of the discussion, but don't do so for the other examples in this paper for reasons of perspicuity.

(32) **Linear sequence of verb-forms with *koɭ*:**

vaɭæ-nʤ̆ŭ-ko-ɳɖ-adŭ = ROOT-ASP.INTR-KOɭ-PST-3NSG

vaɭæ-ččŭ-ko-ɳɖ-aaɭ = ROOT-ASP.TR-KOɭ-PST-3FSG

Thus, *vaɭæ-nʤ̆ŭ-kko-* in the intransitive (29), contrasts with *vaɭæ-ččŭ-kko-* in the transitive (31) with respect to gemination and voicing on the morpheme above the verbal root. This suffix amalgamates transitivity (representing the Voice head) with an aspect head (Amritavalli & Jayaseelan 2005). In Sundaresan (2012), I present arguments that what is involved is a head Asp_{res} that yields a derived result state from the main event encoded by its complement (VoiceP in transitives). I.e. $[\![Asp_{res}]\!] = \lambda R_{<s,t>}\lambda s_s \exists e.R(e) \wedge Result(e,s)$. Informally, It existentially binds off the event in its complement and introduces a result state to it.

What is important to note here is that *koɭ* in these forms appears *after* the morpheme *-nʤ̆/čč-* where the gemination alternation appears, which marks transitivity and thus realizes the Voice head. That is, *koɭ* is independent of Voice, realizing a distinct syntactic head which I call Mid (see Sundaresan & McFadden To appear, for detailed argumentation with respect to these points). Based on the discussion so far, we may summarize the properties of *koɭ* as follows:

(33) *koɭ* spells out a head Mid, which is above Asp_{res}, which is above Persp, which is above Voice. I.e. Mid > Asp_{res} > Persp > Voice.[8]

4.2 The meaning contribution of *koɭ*

In this section, I turn to the question of the meaning contribution of *koɭ*. Specifically, I will motivate the idea that it is a thematic raising predicate in the manner described above with an affectedness (Jackendoff 1990, Beavers 2011a,b) semantics.

4.2.1 Affectedness reading

The addition of *koɭ* to most verbs in Tamil is actually optional. This allows us to consider a wide cross-section of different transitive, unergative and unaccusative verbs (examples of verbs were taken from Levin 1993) and construct sentential minimal pairs around them with and without *koɭ*. Such minimal pairs then give us a way to tease apart the meaning contribution of *koɭ* by itself. As an example, consider the transitive pair below:

(34) Mansi paal-æ uutt-in-aaɭ.
Mansi milk-ACC pour-ASP--PST-3FSG
'Mansi poured the milk.'

(35) Mansi paal-æ uutt-i-kko-ɳɖ-aaɭ.
Mansi milk-ACC pour-ASP-*koɭ*-PST-3FSG

[8]The relative hierarchy of Persp with respect to Asp_{res} is unclear. I assume that Asp_{res} is merged above Persp but nothing serious hinges on this choice, as far as I can see. A different option might be to assume that Asp_{res} includes additional perspectival properties and is thus a semantically mixed category of sorts: thanks to John Beavers (p.c.) for this idea. A different point that should be mentioned here has to do with selection. In particular, I do not assume that *koɭ* selects PerspP, but that their relative position is handled in terms of a rigid functional sequence or Extended Projection.

'Mansi poured the milk for herself.' READING 1
'Mansi poured the milk on herself.' READING 2

(34) has the straightforward meaning that Mansi poured milk. The addition of *koɭ* to the verb in (35) adds the reading that Mansi poured the milk *for* herself or, alternatively, that she poured the milk *on* herself (accidentally). I.e. *koɭ* contributes a reading of affectedness — namely, that the AGENT Mansi became affected by the end result (or outcome) of the pouring event in some sense. Readings 1 and 2 share the core meaning of affectedness but differ in terms of whether this affectedness reading is interpreted in a mental or spatial sense. Reading 1 has a *mental* affectedness reading: i.e. Mansi poured the milk and the end result of the pouring event benefitted her in some way (see Jackendoff 1990, for the idea that benefectiveness is a type of affectedness). Reading 2 has a *spatial* affectedness reading: i.e. Mansi poured the milk and the end result of the milk physically affected her in some way (e.g. she spilled the milk on herself).

I believe the lexico-semantic meaning of *koɭ* is itself underspecified as to whether the affectedness is mentally or spatially interpreted. This will then be restricted both by the discourse-pragmatics and by the inherent meaning of the main predicate to which *koɭ* attaches. For instance, the addition of *koɭ* to an inherently directed motion verb is quite degraded with the reading that Raman was spatially/physically affected by the outcome of this event. Given that the event of falling down already involves a notion of affectedness to the physical body of Raman, any additional affectedness reading contributed by *koɭ* is superfluous, thus marked. However, an affectedness reading along the *mental* dimension may still be contributed by *koɭ* (to the extent that *viɻŭ* ('fall'), like its English counterpart, may be coerced into an agentive reading) — e.g. in a scenario where Raman deliberately fell down (or, perhaps more precisely, dropped down) to avoid being seen.

(36) Raman kiiɻæ viɻŭ-nd-aan/??viɻŭ-ndŭ-kko-ɳɖ-aan.
Raman[NOM] down fall-PST-3MSG/??fall-ASP-*koɭ*-PST-3MSG
"Raman fell down."

Other types of affectedness reading are simply ruled out on pragmatic grounds. For instance, when *koɭ* attaches to the verb *aɖipaɖu* ('injure'), the resulting sentence is degraded with *koɭ* altogether: a spatial affectedness reading is ruled out for the same reasons as that in (36). But a benefactive meaning is pragmatically marked too since one doesn't typically injure oneself voluntarily.

The affectedness contributed by *koɭ* thus places restrictions on the types of verb that *koɭ* may attach to. As mentioned, with most classes of predicate, *koɭ*-suffixation is actually optional, but some verbs occur more readily with the suffix than others. Verbs that are readily compatible with an affectedness reading such as grooming, postural, and self-benefactive verbs frequently co-occur with *koɭ*, as with *vaaɻŭ* ('comb') in (37) below:

(37) Krishnan talai-jæ vaar-i-ko-ɳɖ-aan.
Krishnan[NOM] hair-ACC comb-ASP-*koɭ*-PST-3MSG
1. PHYSICAL AFFECTEDNESS READING: "Krishnan combed the hair and came to be physically affected by the outcome of this event." I.e. "Krishnan combed his (own) hair."
2. MENTAL AFFECTEDNESS READING: "Krishnan$_i$ combed his (or someone else's) hair and came to be mentally affected by the outcome of this event."

The structure in (37) with *koɭ* is, in fact, the standard way to express the reading that Krishnan combed his own hair. However, we may also understand the contribution of *koɭ* in (37) along the

mental dimension: (37) would then mean that Krishnan combed the hair (his own or someone else's) but that the result of the hair-combing event mentally benefited him in some way.

In contrast, verbs which don't take thematic arguments at all, like weather verbs and raising predicates, are incompatible with *koɭ*; verbs of creation and disappearance ('die') are also degraded with *koɭ*, which is also unsurprising, given that the argument is not present through all the relevant stages of the event.[9] Also incompatible are verbs whose meaning is at odds with the kind of affectedness semantics that *koɭ* contributes, e.g. *kuɖŭ* ('give'). This fails to readily combine with *koɭ* for the same reason that *self*-benefactive verbs are so readily compatible with it: the affectedness reading in *koɭ*-structures applies to the external argument in the transitive structures under discussion, so a structure that already involves a distinct affected internal argument (i.e. the GOAL), is at odds with this (see also Lidz & Williams 2005, for discussion of related Kannada facts).[10]

4.2.2 *koɭ* vs. psych verbs

We had observed at the beginning of this paper that *koɭ* is incompatible with psych verbs (see again the examples in (8)-(9); the latter is repeated below). A closer look shows us that *koɭ* is significantly more degraded with stative psych predicates than with eventive ones. Consider a stative psych predication with and without *koɭ*:

(38) Abinaya Dhanush-æ virŭmbŭ-gir-aaɭ/*virŭmb-i-kkoɭ-gir-aaɭ.
Abinaya.NOM Dhanush-ACC love-PRS-3FSG/love-ASP-*koɭ*-PRS-3FSG
"Abinaya loves Dhanush."

Recall (cf. (33)) that *koɭ* itself attaches to a resultative aspectual head (Asp_{res}) which is merged with the main event (VoiceP in transitives) and creates a result state out of it. This means that the meaning that *koɭ* introduces applies not to the main predication (encoded by VoiceP), but to the outcome (or derived result state) of that predication — this is precisely what we observed with the *koɭ*-sentences seen so far. But a result state can only be created out of an event, not out of a state (a point that is implicit in the denotation for Asp_{res}). This entails that *koɭ* cannot be combined with stative predicates like *viɻŭmbŭ* 'love' in (38). Confirming this point is the fact that *koɭ* is actually licit with certain eventive psych predicates, like *baja-* ('fear') below:[11]

(39) Raman baja-nd-aan.
Raman[NOM] fear-PST-3MSG
"Raman got scared." (Rough translation)

[9] E.g. with 'die', the individual denoted by the argument cannot be (mentally or spatially) affected by the outcome her death simply because she, by definition, ceases to exist after her death.

[10] Interestingly, and as may be expected, *koɭ* is still possible in these structures as long as we can find a discourse context where the external argument may *also* be construed as affected, as in a context where Sudha gives her boss, Champa, an expensive book for her birthday to win her favor in (i) below:

i. Sudha Champa-kiʈʈæ pustagatt-æ kuɖŭ-ttŭ-kko-ɳɖ-aaɭ.
Sudha Champa-TO book-ACC give-ASP-*koɭ*-PST-3FSG
"Sudha gave the book to Champa."

[11] (39) and (40) have nearly identical meanings. *koɭ* is presumably still contributing an additional mental affectedness reading that then pertains to the experiencer, but it is as yet unclear what this is.

(40) Raman baja-ndŭ-kko-ɳɖ-aan.
Raman[NOM] fear-ASP-*kol*-PST-3MSG
"Raman got scared."

4.2.3 Thematic raising and the rebounding effect

Regardless of how the affectedness reading of *kol* is actually formalized, it is clear that it involves a kind of "rebounding" effect commonly noted with certain types of middles crosslinguistically (see Kemmer 2003). Informally, the outcome of the main event predication that *kol* attaches to comes back to affect one of the participants of that same event. In the transitive structures we have primarily considered here, this has been the external argument of that event.[12] This, indeed, is why *kol* was seen to be degraded with predicates which apply the affectedness reading to a new argument – as with verbs like 'give'. In this sense, *kol* is not an applicative head (à la Pylkkänen 2008): it doesn't introduce a *new* argument into the structure and assign it an affectedness reading. Rather, the affectedness semantics of *kol* applies to an argument that has already been merged in the structure — namely to the external argument in its complement, in Spec, VoiceP. Here, I'll model this observation in two steps. First, I'll propose that the affectedness semantics of *kol* applies as a θ-role to the argument that is merged in its specifier. Second, this argument must be internally, not externally, merged: i.e. it must be *raised* into this position from within the structure. This, indeed, is what yields the rebounding effect, ensuring that the meaning of *kol* will affect an individual that is already a participant of the main event (Ramchand 2008). In transitives, this is the external argument in Spec, VoiceP: this DP gets an Agent θ-role from Voice and then raises up to Spec-MidP to get an additional affectedness role from Mid.

Many aspects of the precise meaning of *kol* remain to be worked out and formalized. For instance, it is not entirely clear what the unified notion of affectedness is that underlies the contribution of *kol* in unaccusatives (see again (29) with 'bend') and transitives. A promising line of investigation might be the proposal in Beavers (2011b) which argues that affectedness involves a transition between states on some scale, with different kinds of affectedness involving different scales. I leave this for future research.

5. Back to the beginning: reflexives and *kol*

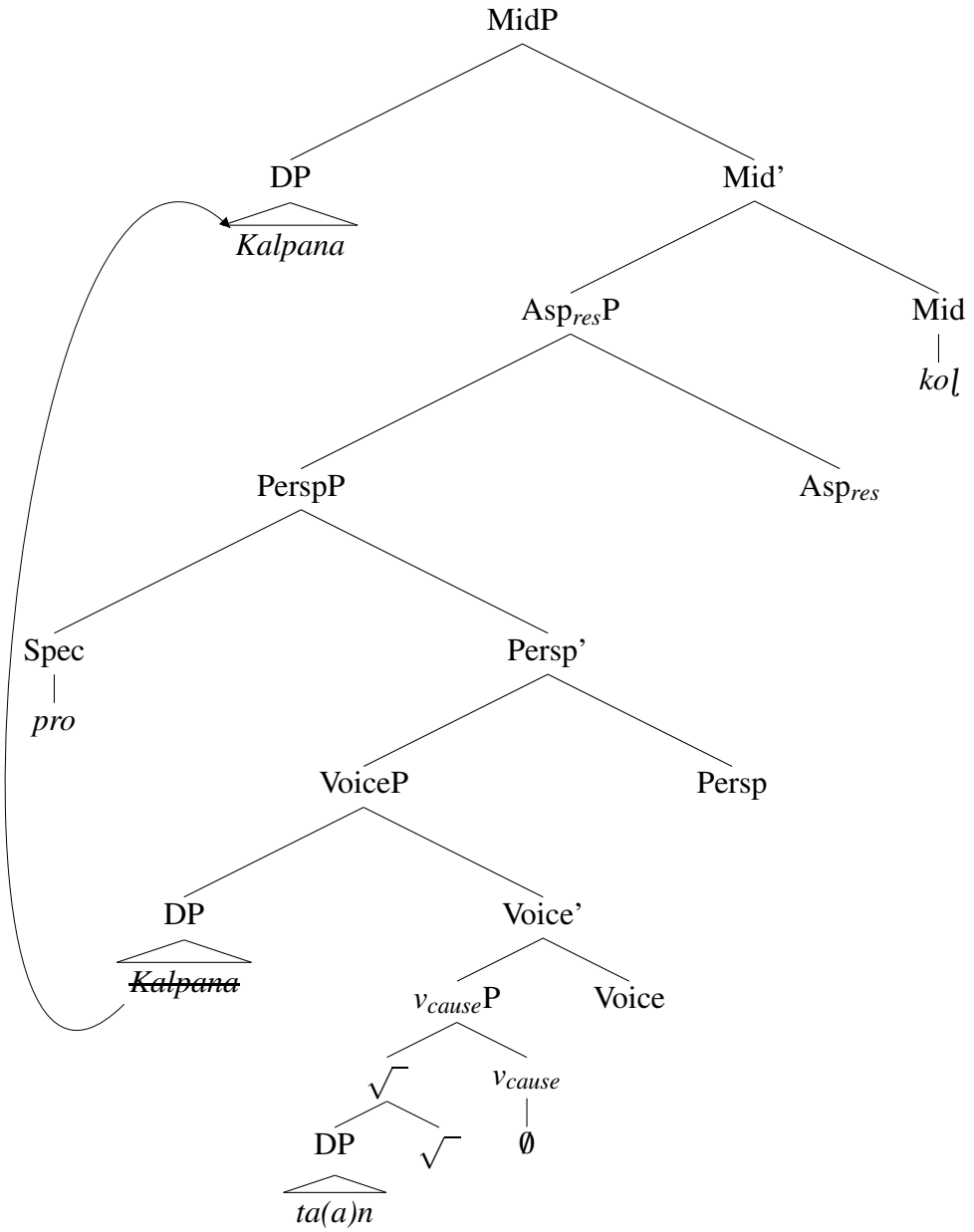
The discussion so far has established that *kol* spells out a Mid(dle) head that selects a PerspP in its complement, which is crucially above VoiceP. Furthermore, I have argued that Mid is a thematic raising predicate. We now have all the pieces of the puzzle needed to understand why *kol* is required for reflexive anaphora in the standard case, i.e. with non-psych predicates.

Consider again a *kol*-reflexive like that in (3), repeated below which, given what we've argued, must have the structure given in (42):

(41) Kalpana$_i$ tann-æ$_{\{i,*j\}}$ ki[[i-ko-ɳɖ-aa[.
Kalpana.NOM ANAPH-ACC.SG pinch-ASP-*kol*-PST-3FSG
"Kalpana$_i$ pinched herself$_{\{i,*j\}}$."

[12] This restriction is, incidentally, absolute. The affected argument in a transitive structure must be the external argument and cannot be the internal one. This can simply be modelled as a function of Minimality: the external argument is closer to the Mid head that *kol* spells out that is the internal one.

(42)



The matrix subject *Kalpana* is externally merged in Spec, VoiceP where it is assigned an AGENT θ-role. It is then thematically raised up to the Spec, MidP where it is assigned an additional Affectee θ-role by Mid, identifying it as the affected argument of the result state of the main event in the scope of Mid.[13] As a result of this, *Kalpana*, which had earlier been properly contained inside the minimal PerspP containing the anaphor, now bears scope outside the PerspP. From this new raised position in Spec, MidP, it can thus corefer with the *pro* in Spec, PerspP without inducing a Condition C violation. As such, it can lictly antecede *ta(a)n* in (41).[14] When *Kalpana* is the only salient antecedent available, as in (41), it is also the chosen antecedent, yielding reflex-

[13]Here, the salient reading is one of mental affectedness since 'pinch' lexically already subsumes a reading of physical affectedness.

[14]The precise syntactic properties of the perspectival *pro* in Spec, PerspP need to be clarified further. What is absolutely crucial to my account here is that it not count as "(pro)nominal" for the purposes of Principle B, which would otherwise be violated by a coreferring co-argument (R-expression or pronoun) in Spec, MidP. Such an approach

ive anaphora. When there are other potential antecedents available, as in (5) – repeated below — then either might be chosen, yielding either reflexive anaphora due to antecedence by *Kalpana* or non-reflexive anaphora due to antecedence by *Siva*:

(43) Siva$_i$ [$_{CP}$ Kalpana$_i$ tann-æ$_{\{i,j\}}$ kiɭ-i-kko-ɳɖ-aaɭ-ŭnnŭ] nene-tt-aan.
Siva Kalpana ANAPH-ACC.SG pinch-ASP-*koɭ*-PST-3FSG-COMP think-PST-3MSG
"Siva$_i$ thought that Kalpana$_j$ pinched him$_i$/herself$_j$ (for his/her own benefit)." (Rough translation)

Finally, to complete the paradigm, let us consider how other types of perspectival anaphora in Tamil are able to obtain in the absence of *koɭ*. The answer is something we have noted already. Reflexive anaphora (definitionally) instantiates the only structural configuration where the targeted antecedent of the anaphor is its co-argument; this is thus also the only configuration where the targeted antecedent starts out in the same minimal PerspP as the anaphor (recall that there is no minimal PerspP in the clausal domain that is smaller than VoiceP). In all other cases of anaphora — i.e. logophora and long-distance anaphora across CPs, PPs, and DPs — the targeted antecedent already begins its life outside the minimal PerspP containing the anaphor, thus can denote a perspective-holder, assuming independent thematic and discourse constraints on this are satisfied,[15] thereby qualifying as a potential antecedent for the anaphor.

We have thus explained the core *koɭ* patterns in reflexives and non-reflexive anaphoric structures in Tamil that we started this paper with. It is important to note, in this context, that *koɭ*'s interaction with reflexivity — in particular, the idea that it imbues the co-argument of the anaphor with extra properties that allow it to serve as its antecedent — is an entirely *incidental* by-product of its thematic raising property, which itself follows from the inherent meaning of the Mid head that *koɭ* spells out. There is no *direct* connection between *koɭ* and reflexivity: while (non-psych predicate) reflexives must occur with *koɭ*, *koɭ* can freely occur with non-reflexive transitives, unaccusatives and unergatives. In this sense, it is also misleading to classify *koɭ* as a reflexive marker.

There is one last point that still needs to be clarified. This has to do with the obligatory absence of *koɭ* in psych reflexives, as illustrated by the patterns repeated below:

(44) Kalpana-vŭkkŭ$_i$ tann-æ$_{\{i,*j\}}$ piɖikkæ-læ/*piɖi-ttŭ-kkoɭɭæ-læ.
Kalpana-DAT ANAPH-ACC.SG like-NEG/*like-ASP-*koɭ*-NEG
"Kalpana$_i$ didn't like herself$_{\{i,*j\}}$."

(45) Abinaya$_i$ tann-æ$_{\{i,*j\}}$ virŭmbŭ-gir-aaɭ/*virumb-i-*koɭ*ɭŭ-gir-aaɭ.
Abinaya.NOM ANAPH-ACC.SG love-PRS-3FSG/*love-ASP-*koɭ*-PRS-3FSG
"Abinaya$_i$ loves herself$_{\{i,*j\}}$."

There are two theoretical aspects to these empirical patterns. The first is the fact that *koɭ* is incompatible with stative psych predicates. We have already explained why this is the case, arguing that it follows from the fact that most psych-predicates are stative and that the Asp$_{res}$ head in the complement of *koɭ* can only combine with eventives. As we saw, *koɭ* is, in fact, licit with eventive psych predicates as in (39)-(40).

may also be necessary to explain why it is the external argument in Spec, VoiceP and not the perspectival *pro* (which is minimally closer) that is raised to Spec, MidP.

[15] Recall that, in the unmarked discourse scenario, subjects tend to be able to denote perspective-holders more readily than objects, with the exception of EXPERIENCER objects.

The second has to do with the availability of reflexive anaphora in the absence of *kol*. Here, we must reason backwards. Reflexive anaphora in psych predications is also regulated by perspective: in particular, the co-argument EXPERIENCER antecedent denotes a mental perspective holder with respect to the PerspP containing the anaphor. This, in turn, must mean that psych verbs involve a structure containing a PerspP, and that the EXPERIENCER argument of a psych verb is merged (or perhaps moved) above this PerspP. If it were properly contained inside the minimal PerspP containing the anaphor, we would get a Condition B or C violation if it also anteceded the anaphor, as we have already argued.

Neither of these is an unreasonable conclusion to draw. Since psych predicates denote a mental or psychological experience, it seems reasonable to posit that they involve a mental PerspP as part of their argument-structure (much like attitude verbs do). One way to ensure that the EXPERIENCER is higher than the minimal PerspP (which would contain the anaphoric THEME argument if there is one) would be to say the argument-structure of a psych-predicate is structurally larger and more complex than those of other types of verbs and that, in particular, the EXPERIENCER is merged higher than Spec, VoiceP. Such a proposal is actually in line with others — Adger & Ramchand (2006) e.g. argue that psych predication in Scottish Gaelic involves experiencers that are base-generated higher than other stative subjects. Positing a larger structure may crucially also help explain the hitherto puzzling possibility of backward binding (Minkoff 2003) in psych-predicate structures. This is a matter for future research.

References

Adger, David, & Gillian Ramchand. 2006. Psych nouns and predication. In *Proceedings of NELS*, volume 36.

Amritavalli, R., & K.A. Jayaseelan. 2005. Finiteness and negation in Dravidian. In *The Oxford Handbook of Comparative Syntax*, ed. Guglielmo Cinque & Richard S. Kayne, 178–220. Oxford: OUP.

Anand, Pranav, & Andrew Nevins. 2004. Shifty operators in changing contexts. In *Proceedings of SALT*.

Beavers, John. 2011a. An aspectual analysis of ditransitive verbs of caused possession in English. *Journal of semantics* 28:1–54.

Beavers, John. 2011b. On affectedness. *Natural Language and Linguistic Theory* 29:335–370.

Bianchi, Valentina. 2003. On finiteness as logophoric anchoring. In *Temps et point de vue/Tense and Point of View*, ed. Jacqueline Guéron & L. Tasmovski, 213–246. Nanterre: Université Paris X.

Bylinina, Lisa, Eric McCready, & Yasutada Sudo. 2014. The landscape of perspective shifting. Talk given at Pronouns in Embedded Contexts at the Syntax-Semantics Interface, Tübingen.

Bylinina, Lisa, & Yasutada Sudo. 2015. The semantics of perspective sensitivity. ESSLLI 2015 course lectures.

Charnavel, Isabelle. 2015. Apparent exemption from Condition A: a perspective-based theory. Http://ling.auf.net/lingbuzz/002683.

Clements, G.N. 1975. The logophoric pronoun in Ewe: its role in discourse. *Journal of West African Languages* 2:141–177.

Embick, David. 2004. Unaccusative syntax and verbal alternations. In *The unaccusativity puzzle:*

Explorations of the syntax-lexicon interface, ed. E. Anagnostopoulou A. Alexiadou & M. Everaert, Oxford Studies in Theoretical Linguistics, 137–158. Oxford: OUP.
Fillmore, Charles. 1997. *Lectures on deixis*. Stanford: CSLI.
Giorgi, Alessandra. 2006. From temporal anchoring to long distance anaphors. *Natural Language and Linguistic Theory* .
Giorgi, Alessandra. 2010. *Towards a syntax of indexicality*. Oxford Studies in Theoretical Linguistics. Oxford: Oxford University Press.
Hellan, Lars. 1988. *Anaphora in Norwegian and the theory of grammar*, volume 32 of *Studies in Generative Grammar*. Dordrecht: Foris Publications.
Hicks, Glyn. 2009. *The derivation of anaphoric relations*. Linguistik Aktuell. John Benjamins.
Huang, C.C. James, & C.T. Jane Tang. 1991. The local nature of the long-distance reflexive in Chinese. In *Long-distance anaphora*, ed. Eric Reuland Jan Koster, 263–282. Cambridge University Press.
Huang, James, & Luther Liu. 2001. Logophoricity, attitudes, and *ziji* at the interface. In *Long-distance reflexives*, ed. Peter Cole, Gabriella Hermon, & James Huang, 141–195. New York: Academic Press.
Jackendoff, Ray. 1990. *Semantic structures*. Cambridge, MA: MIT Press.
Jayaseelan, K.A. 1997. Anaphors as pronouns. *Studia Linguistica* 51:186–234.
Kemmer, Suzanne. 2003. *The middle voice*. Typological studies in language. Philadelphia: John Benjamins.
Koopman, Hilda, & Dominique Sportiche. 1989. Pronouns, logical variables, and logophoricity in Abe. *Linguistic Inquiry* 20:555–588.
Koster, Jan, & Eric Reuland, ed. 1991. *Long-distance anaphora*. Cambridge: CUP.
Kratzer, Angelika. 1996. Severing the external argument from its verb. In *Phrase structure and the lexicon*, ed. Johan Rooryck & Laurie Zaring, 109–137. Dordrecht: Kluwer.
Kuno, Susumo. 1987. *Functional syntax – anaphora, discourse and empathy*. Chicago: Chicago University Press.
Levin, Beth. 1993. *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago Press.
Lidz, Jeffrey. 2001. The argument structure of verbal reflexives. *Natural Language and Linguistic Theory* 19:311–353.
Lidz, Jeffrey. 2004. Causation and refexivity in Kannada. In *Clause structure in South Asian languages*. Springer.
Lidz, Jeffrey, & Alexander Williams. 2005. On the absence of reflexive benefactives in Kannada. In *The Yearbook of South Asian Languages and Linguistics*, ed. Tanmoy Bhattacharya Rajendra Singh, 237–254. Berlin: Mouton de Gruyter.
Manzini, Rita, & Ken Wexler. 1987. Parameters, binding theory, and learnability. *Linguistic Theory* 18:413–444.
Medová, Lucie. 2009. Reflexive clitics in the Slavic and Romance languages. Doctoral dissertation, Princeton University.
Minkoff, Seth. 2003. Consciousness, backward coreference, and logophoricity. *Linguistic Inquiry* 35:485–494.
Mitchell, Jonathan. 1986. The formal semantics of point of view. Doctoral dissertation, University of Massachusetts, Amherst, MA.
Nishigauchi, Taisuke. 2014. Reflexive binding: awareness and empathy from a syntactic point of

view. *Journal of East Asian Linguistics* 23:157–206.
Oshima, David. 2006. Perspectives in reported discourse. Doctoral dissertation, Stanford University, Stanford, CA.
Pearson, Hazel. 2013. The sense of self: Topics in the semantics of de se expressions. Doctoral dissertation, Harvard University, Cambridge.
Pica, Pierre. 1987. On the nature of the reflexivization cycle. In *Proceedings of NELS 17*, ed. J. McDonough & B. Plunkett. University of Massachusetts, Amherst, GLSA.
Progovac, Liljana. 1993. Long-distance reflexives: movement to Infl versus relativized SUBJECT. *Linguistic Inquiry* 24:755–772.
Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge, Mass: MIT Press.
Ramchand, Gillian. 2008. *Verb meaning and the lexicon: a first phase syntax*. Cambridge: Cambridge University Press.
Reinhart, Tanya, & Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24:657–720.
Reuland, Eric. 2001a. Anaphors, logophors, and binding. In *Long-distance reflexives*, ed. Peter Cole, Gabriella Hermon, & C.-T. James Huang, 343–370. London: Academic Press.
Reuland, Eric. 2001b. Primitives of binding. *Linguistic Inquiry* 32:439–492.
Reuland, Eric. 2011. *Anaphora and language design*. Cambridge, Mass.: MIT Press.
Schäfer, Florian. 2008. *The syntax of (anti-)causatives. external arguments in change-of-state contexts*, volume 126 of *Linguistik Aktuell*. Philadelphia: John Benjamins.
Sells, Peter. 1987. Aspects of logophoricity. *Linguistic Inquiry* 18:445–479.
Sigurðsson, Halldór Ármann. 1990. Long distance reflexives and moods in Icelandic. In *Modern Icelandic syntax*, ed. Joan Maling & Annie Zaenen, 309–346. New York: Academic Press.
Sigurðsson, Halldór Ármann. 2010. Mood in Icelandic. In *Mood in the languages of Europe*, ed. Björn Rothstein & Rolf Thieroff, 33–55. Amsterdam: John Benjamins.
Sportiche, Dominique. 1998. *Partitions and atoms of clause structure: subjects, agreement, case, and clitics*. Psychology Press.
Sundaresan, Sandhya. 2012. Context and (co)reference in the syntax and its interfaces. Doctoral dissertation, University of Tromsø and University of Stuttgart, Tromsø.
Sundaresan, Sandhya. 2016a. Anaphora vs. agreement: a new kind of Anaphor Agreement Effect in Tamil. In *The impact of pronominal form on interpretation*, ed. Patrick Grosz & Pritty Patel-Grosz. de Gruyter, Mouton.
Sundaresan, Sandhya. 2016b. Perspective is structural: evidence from anaphora. Lingbuzz: http://ling.auf.net/lingbuzz/002893.
Sundaresan, Sandhya, & Thomas McFadden. To appear. The articulated *v* layer: evidence from Tamil. In *The verbal domain*, ed. Roberta D'Alessandro, Irene Franco, & Ángel Gallego. Oxford: Oxford University Press.
Sundaresan, Sandhya, & Hazel Pearson. 2014. Formalizing linguistic perspective: insights from spatial anaphora. Paper based on talk given at LAGB 2014: http://ling.auf.net/lingbuzz/002270.

Sandhya Sundaresan
sandhya.sundaresan@uni-leipzig.de

Pronoun Agreement Mismatches in Telugu*

Troy Messick

University of Connecticut

1. Introduction

How languages express *de se* attitude reports in finite clauses is subject to a large amount of cross linguistic variation. For example, English and other Indo-European languages do not distinguish *de se* attitudes from *de re* attitudes. Thus, the utterance in (1) can be used to report an attitude with the attitude holder fully aware that the said attitude is about himself (i.e., *de se*) or unaware that the attitude is about himself (i.e., *de re*). This allows (1) to be used to felicitously report both the scenario in (2a) and (2b).

(1) John said that he is smart.

(2) a. John said, "I am smart."
b. John$_i$ said, "he$_i$ is smart."

However a number of authors have recently noted that in many languages, *de se* attitude reports are expressed via *indexical shift*, where a first person pronoun is used to refer to the attitude holder. This is shown in (3) for Zazaki (Anand & Nevins 2004, 21, see this work for evidence that we are not dealing with a quoted clause here.)

(3) Hɛseni$_j$ va [kɛ ɛz$_j$ dɛwletia]
Hesen.OBL said [that I rich.be-PRES]
'Hesen said that he was rich.'

A large body of literature has tried to account for such variation (e.g., Schlenker 1999, 2003, von Stechow 2002, 2003, Anand 2006, Sudo 2012). Dravidian languages have shown

*Thanks to Radhika Shiradkar for her patience and generosity in sharing their language with me. I also thank Jonathan Bobaljik, Željko Bošković, Jon Gajewski, Magdalena Kaufmann, and the FASAL 5 reviewers and audience for comments and questions about the topics presented here. Finally thanks to the FASAL 5 proceedings editors and an anonymous reviewer for suggestions that lead to many improvements of this paper. All errors are my responsibility.

an interesting different pattern. For instance, in Tamil, overt first person pronouns do not shift, but the agreement morphology can be first person when the subject refers to the matrix subject under verbs of speech. (Asher 1985, Woolford 1999, Sundaresan 2012). This is shown in (4).

(4) Murukeesan [taan var-r-een-nnŭ] so-nn-aarŭ
Murugesan [ANPH come.PRES-1SG-COMP] say-PAST-3MSG
'Murugesan said that he would come.'

In this paper, I investigate this phenomenon in Telugu.[1] As (5) shows, a *de se* attitude report in Telugu[2] has a third person pronoun controlling first person agreement on the embedded verb very similar to what we find in Tamil.[3]

(5) Rani [t̪anu exam pass ajj-aa-n-ani] nam-mu-t̪undi.
Rani [3SG exam pass happen-PAST-1SG-COMP] believe-PAST-F.SG
'Rani believed that she passed the exam.'

Following Sundaresan, I will refer to this type of agreement pattern as *monstrous agreement*. I will show that monstrous agreement has the same interpretive properties and syntactic distribution as indexical shift. Despite these similarities, previous accounts of indexical shift cannot straightforwardly account for agreement shift. In this paper, I propose a uniform analysis for the two. The basic idea of the analysis is that in languages like Telugu, when a third person pronoun is interpreted *de se*, it is semantically first person but morphologically third person (in a way to be made explicit later). The agreement probe, responsible for verbal morphology, targets the semantic features of the controller (Corbett 1979, 1983, 2006). Since my analysis attempts to cover similar data to the analysis given in Sundaresan (2012), I will briefly compare the two. I will then show that the analysis can be expanded to account for indexical shift and other forms of *de se* marking while also giving a principled explanation for a previously unnoticed typological gap.

[1]The Telugu data presented here comes from multiple elicitation sessions with a linguistically untrained Telugu consultant. The semantic judgments were obtained using a Truth Value Judgment Task procedure where the consultant was presented with a scenario and a sentence and was asked to judge whether the sentence was true in the given scenario. These judgments were stable across multiple consultation sessions.

[2]A reviewer notes that there is cross linguistic variation in which embedding predicates allow for indexical shift/logophors/monstrous agreement, with some languages only allowing these phenomena to happen in a subset of attitude contexts. As we will see in section 2.3, Telugu appears not have these restrictions and monstrous agreement can occur in any attitude context.

[3]Curnow (2002) notes what seems to be the same agreement pattern in the Nilo-Saharan languages Karimojong and Lotuko. I also present data that the related language Nuer also has this phenomenon in Messick (in preparation). There are also languages like Donno Sɔ (Culy 1994), where first person agreement appears to be controlled by logophors. I discuss this in section 5.

2. Agreement in Telugu embedded attitudes

In this section, I will introduce the Telugu monstrous agreement data. Where relevant, I will compare agreement shift to the better studied cases of indexical shift. I will show that like other cases of indexical shift, monstrous agreement in Telugu can only be used to express *de se* attitudes; also, as with indexical shift in Uyghur, the syntactic distribution of agreement shift is tied to a special type of complementizer.

2.1 Background on Telugu agreement

Telugu displays verbal agreement with non-case marked subjects. The agreement paradigm for matrix clauses is given in (6) (putting aside number).

(6) a. neenu parigeṯṯ-ææ-nu
1SG run-PAST-1SG
'I ran.'

b. nuvvu parigeṯṯ-ææ-vu
2SG run-PAST-2SG
'You ran.'

c. ṯanu parigeṯṯ-ææ-Du
3SG run-PAST-M.SG
'He ran.'

d. ṯanu parigeṯṯ-in-di
3SG run-PAST-F.SG
'She ran.'

Before we continue, I would like to note that what I gloss as the third person pronoun *ṯanu* is cognate to *ṯa(a)n* found in other Dravidian languages such as Malayalam (Anand 2006) and Tamil (Sundaresan 2012). *Ṯa(a)n* in these languages is usually not treated as a third person pronoun, but a logophoric pronoun or a long-distance anaphor. *Ṯanu* was evidently also once logophoric, however in current usage, speakers use it as a non-logophoric third person pronoun (Krishnamurti & Gwynn 1985, 73).

This can be seen by examining the distribution of *ṯanu*. Logophoric pronouns are typically found in embedded attitude reports; they cannot be the matrix subject of an out of the blue sentence. This is shown in (7) for the logophoric pronoun *yè* in Ewe. *Yè* can be used in attitude reports (7a), but not as the matrix subject of an out-of-the-blue context (7b) (data taken from Pearson in press).

(7) a. kofi be yè-dzo
Kofi say LOG-leave
'Kofi$_i$ said that he$_i$ left'

b. *yè dzo
LOG leave
Intended: 'He left'

Ṯanu, on the other hand, can be used in both environments as shown in (8). Not only can *ṯanu* be used in embedded attitudes (8a), but also in matrix clauses in out-of-the-blue contexts (8b).

(8) a. Raju ṯanu parigeṯṯ-ææ-nu ani čepp-ææ-Du
Raju 3SG run-PAST-1SG COMP say-PAST-M.SG
'Raju said that he ran.'

b. ṯanu parigeṯṯ-ææ-Du
3SG run-PAST-M.SG
'He ran.'

(8b) can even used deictically (i.e., accompanied by a pointing gesture). So I take the treatment of *ṯanu* as a third person pronoun to be empirically well-founded.

As noted in section 1, Telugu allows for monstrous agreement with pronouns with embedding in attitude reports. When the report expresses an attitude about the attitude holder, the agreement on the embedded verb can be either third person (9a) or first person (9b).

(9) a. Raju [ṯanu parigeṯṯ-ææ-Du ani] cepp-ææ-Du
Raju 3SG run-PAST-M.SG COMP say-PAST-M.SG
'Raju said that he ran.'

b. Raju [ṯanu parigeṯṯ-ææ-nu ani] cepp-ææ-Du
Raju 3SG run-PAST-1SG COMP say-PAST-M.SG
'Raju said that he ran.'

In order to rule out the possibility that the embedded clause is (partially) quoted, I provide two diagnostics from matrix question formation and NPI licensing. As has been noted in the literature on indexical shift (e.g., Anand & Nevins 2004, 21), grammatical dependencies cannot cross quotation marks. This is shown for English in (10). In (10a), a *wh*-element is moved out of the quoted clause into the matrix clause and the resulting utterance is ungrammatical. Likewise, the ungrammaticality of (10b) is caused by the fact that the matrix negation cannot license the NPI in the quoted clause.

(10) a. *What$_i$ did Bob say, "I ate t_i"?
b. *Bob didn't say, "I ate any bananas."

As is the case with monstrous agreement in Tamil and indexical shift in languages like Zazaki, Telugu allows such dependencies between the embedded and matrix clauses indicating that the embedded clause is not a quotation. This is shown in (11). In (11a), a

wh-element *eemi* in the embedded clause can scope into the matrix clause and receive matrix question interpretation. In (11b), negation in the matrix clause can license the NPI in the embedded clause.

(11) a. Raju [t̪anu eemi tinn-aa-nu ani] cepp-ææ-Du
Raju [3SG what eat-PAST-1SG COMP] say-PAST-M.SG
'What did Raju say I ate?'

b. Raju [t̪anu ee aratipanD-lu tinn-aa-nu ani] cepa-leed̪u
Raju [3SG any banana-PL eat-PAST-1SG COMP] say-NEG
'Raju did not say that he ate any bananas.'

Having ruled out a quotation analysis of monstrous agreement in Telugu, let us turn to investigating what elements can control monstrous agreement. In addition to *t̪anu*, the second person pronoun *nuuvu* can also control monstrous agreement. As shown in (12), *nuuvu* can control either second person (12a) or first person (12b) agreement.[4]

(12) a. nuuvu parigeṯṯ-ææ-vu ani nuuvu cepp-ææ-vu
2SG run-PAST-2SG COMP 2SG say-PAST-2SG
'You said that you ran.'

b. nuuvu parigeṯṯ-ææ-nu ani nuuvu cepp-ææ-vu
2SG run-PAST-1SG COMP 2SG say-PAST-2SG
'You said that you ran.'

Monstrous agreement is only acceptable in embedded clauses. Mismatches are disallowed in matrix clauses, as in (13).

(13) a. t̪anu parigeṯṯ-ææ-Du
3SG run-PAST-M.SG
'He ran.'

b. *t̪anu parigeṯṯ-ææ-nu
3SG run-PAST-1SG
'He ran.'

A final note: what sets monstrous agreement apart from indexical shift is the fact that pronouns do not shift. In other words, first person pronouns must always refer to the current speaker and cannot refer to the attitude holder. This is shown in (14). The embedded first person pronoun, *neenu*, obligatorily refers to the current speaker.

(14) Raju neenu eemi tinn-aa-nu ani čepp-ææ-Du?
Raju$_i$ 1SG$_{*i/s}$ what eat-PAST-1SG COMP say-PAST-M.SG
'What did Raju say that I ate?'

[4] A reviewer notes that monstrous agreement with second person pronouns may also occur in some dialects of Tamil as well. I leave the investigation of this possibility a matter for future research.

2.2 The interpretation of monstrous agreement

Moving on to the interpretation of attitudes with monstrous agreement; monstrous agreement is only allowed if the report is a *de se* attitude. For that, it must meet the criterion in (15) (from Pearson 2012).

(15) a. *Aboutness condition*: the attitude is about the attitude holder *and*
b. *Awareness condition*: the attitude holder is aware that the attitude is about herself

To test whether monstrous agreement only occurs in *de se* attitudes, a scenario must be constructed where the condition in (15b) is not met to see if the sentence is judged felicitous in such a situation. This is done in (16). In this scenario, Rani is not aware that she has an attitude about herself; the sentence with monstrous agreement is judged to be infelicitous while the sentence without monstrous agreement is judged to be acceptable.

(16) SCENARIO: Rani took an exam, and later saw the top 10 scores with the scorer's student ID numbers. She forgot her own ID number, so did not know who was who. Looking to the top score, she thinks: "This student definitely passed!" But it turned out she was that student.

(17) a. raani [ṯanu exam pass ajj-in-and̪-ani] nam-mu-ṯundi
Rani [3SG exam pass happen-PAST-F.SG-COMP] believe-PAST-F.SG
.'Rani believes that she passed the exam.

b. #raani [ṯanu exam pass ajj-aa-n-ani] nam-mu-ṯundi
Rani [3SG exam pass happen-PAST-1SG-COMP] believe-PAST-F.SG
'Rani believes that she passed the exam.'

A similar interpretative restriction has been found for languages that allow for indexical shift: clauses with indexical shift can only express *de se* attitudes. This is shown for Amharic in (18) (Schlenker 1999, 97; see also Sudo (to appear) for Uyghur).

(18) SCENARIO: Jon, who is a candidate in the election, is so drunk he doesn't remember who he is. He watches TV and sees a candidate he finds terrific, thinking that this guy must be a hero. This candidate happens to be Jon himself though he doesn't realize it.

(19) a. #Jon ʤəgna nə-ññ yil-all
John hero COP.PF-1sO 3M.say-AUX.3M
'John says that he is a hero.'

b. Jon ʃwyew ʤəgna näw alä
John the-man hero is said

'John said the man is a hero.'

A question one may have at this point is: do attitude reports without monstrous agreement like those in (9a) and (12a) also have a *de se* reading or are they always read *de re*? This is a more difficult question than it appears at first because in simple cases, utterances with a *de se* attitude entail the one with a *de re* attitude. Despite this, there are ways to test whether an attitude has a *de se* reading. Below I deploy a test developed in Percus & Sauerland (2003). This test involves the scenario in (20).

(20) SCENARIO: Rani, Raju, Rahul, and Troy all took an exam. Later the exam scores were posted next to the student's ID numbers. Rani was the only confident one and thought, "I passed the exam." Raju and Troy had forgotten their ID numbers and both were pessimistic about how they did, thinking they had failed. They saw the two top scorers and thought that those students definitely passed. It turned out they were those students. Rahul also thought he had failed, but was confident about Rani and thought she had passed.

In this scenario, there are four individuals: one has a *de se* thought, two have *de re* thoughts about themselves, and one has a *de re* thought about the first individual. The test sentence then reports that only the first individual has the attitude. The prediction of the test is that if a report has a *de se* reading, then the sentence will be judged true in this scenario because it is true that she is the only one who has the *de se* attitude, but if the report only has a *de re* reading, then it would be judged as false because other people in the scenario have *de re* beliefs about themselves or the first individual. As indicated by the judgements in (21), both clauses with monstrous agreement and clauses without monstrous agreement are judged to be true in such scenarios, suggesting that both reports have a *de se* reading.

(21) a. raan-e [ṯanu exam pass ajj-in-aṉḍ-ani] nam-mu-ṯuṉḍi
Rani-FOC [3SG exam pass happen-PAST-F.SG-COMP] believe-PAST-F.SG
'Only Rani believes that she passed the exam.'

b. raan-e [ṯanu exam pass ajj-aa-n-ani] nam-mu-ṯuṉḍi
Rani-FOC [3SG exam pass happen-PAST-1SG-COMP] believe-PAST-F.SG
'Only Rani believes that she passed the exam.'

2.3 The distribution of agreement shift

As discussed in the previous section, monstrous agreement, like most cases of indexical shift, can only occur in embedded clauses. For many languages with indexical shift, indexicals only shift under certain attitude verbs. For example. indexicals in Amharic only shift under the verb meaning 'to say' (Anand 2006). This is demonstrated in (22). While the embedded first person morphology can refer to the attitude holder in (22a) when embedded under the verb *ələ*, this is not the case in (22b) where first person morphology must refer to the speaker of the current utterance.

(22) a. John ʤəgna nə-ññ yil-all
John hero COP.PF-1sO 3M.say-AUX.3M
'John says that he is a hero.'

b. John ʤəgna n-ññ yiSəllig-all
John hero COP.PRES-1sO think.IMPERF-3SM
'John thinks that I am a hero.'

Other languages appear to be more permissive. Uyghur, e.g., allows for shifting to occur under verbs of saying, belief, knowledge and direct perception (Sudo 2012). This is shown in (23).

(23) a. Ahmet *pro* kim-ni jaxshi kör-imen di-di
Ahmet *pro* who-ACC well see-IMPERF.1SG say-PAST.3
'Who did Ahmet say that he likes?'

b. Ahmet *pro* kim-ni jaxshi kör-imen dep bil-du
Ahmet *pro* who-ACC well see-IMPERF.1SG COMP believe-IMPERF.3
'Who did Ahmet believe that he likes?

c. Ahmet Aygül-din *pro* qaysi imtihan-din öt-tim dep
Ahmet Aygül-from *pro* which test-from pass-PAST.1SG COMP
angla-di
hear-PAST.3
'Which test did Amhet hear from Aygül that he passed?'

Sundaresan (2012) conjectures that the licensing environments for indexical shift fall on the implicational hierarchy developed by Culy (1994) for logophoric pronouns, given in (24). (24) should be read as stating that if indexical shift is licensed by a class of embedding verbs then all other verbs to its left will also license it.

(24) SPEECH > THOUGHT > KNOWLEDGE > DIRECT PERCEPTION

Sundaresan (2012) shows from survey work that there are four dialects of Tamil that differ on where in the hierarchy they fall.

Where does Telugu fall on this hierarchy? It appears to fall on the far right. Not only does it allow for agreement shift with verbs of saying and belief, as demonstrated in the previous section, but also with verbs of knowledge and direct perception. This is shown in (25).[5]

(25) a. raani [t̪anu exam pass ajj-aa-n-ani] t̪elusu-k̪un-di.
raani [3SG exam pass happen-PAST-1SG-COMP know-REFL-F.SG
'Rani found out she passed the exam'

[5]For a comprehensive list of verbs that license agreement shift see Messick (in preparation).

b. raani [t̪anu exam pass ajj-aa-n-ani] santošanga und̪i.
raani [3SG exam pass happen-PAST-1SG-COMP] happy COP
'Rani is happy that she passed the exam'

Interestingly monstrous agreement in Telugu only occurs in complements introduced by the complementizer *ani*. This again patterns with indexical shift in Uyghur. In Uyghur, complements introduced by the complementizer *dep* allow indexical shift. Interestingly, both *ani* and *dep* are forms of the verb meaning *to say* in Telugu and Uyghur.[6]

3. An analysis of agreement and indexical shift

In this section, I will propose an analysis of monstrous agreement that can also cover the basic cases of indexical shift. The basic idea is that when a pronoun is interpreted *de se*, it is semantically first person. In languages with indexical shift, the morphology allows for those features to be spelled out as first person, but in languages without indexical shift, the morphology forces the features to spelled out as a third person pronoun (putting aside logophors for the time being). What happens in Telugu agreement shift is that the semantic features of the pronoun are able to control agreement on the embedded verb. I will make all these intuitions explicit in the upcoming sections.

3.1 Semantic features and agreement

Descriptively, Telugu agreement shift is a mismatch between agreement controller and the target where it appears that the semantic interpretation of the controller is influencing the agreement target. In a series of typological studies, Corbett (1979, 1983, 2006) has shown that semantic features of nominals can control agreement; in fact, sometimes a nominal can control semantic *and* syntactic agreement in the same utterance. An example of such *semantic agreement* is given in (26). In (26a), a semantically plural noun *committee* can control plural agreement in British English. In (26b) a grammatically masculine noun can control feminine agreement in Russian when the referent of the noun is female. Finally, (26c) shows a case of so called unagreement in Spanish where a third person NP can control first person agreement when the speaker is included in the group the NP is referring to.

(26) a. The committee has/have decided.

b. Novyj vrač skazala
new.MASC doctor said.FEM
'New doctor said...'

c. Qué desgraciadas somos las mujeres!
how unfortunate be.1PL the women
'How unfortunate (we) women are'

[6]It has long been speculated that logophoric pronouns are licensed in complements introduced by complementizers that are forms of verbs of saying (e.g., Sells 1987). For the role such complementizers play in licensing indexical and agreement shift see Messick (in preparation).

The relevance to Telugu is apparent. I suggest that agreement shift is part of this larger paradigm where semantic features are available as agreement controllers. How do we model the ability of semantic features to control agreement? A common way is to allow for nominal elements to carry two sets of ϕ-features, one that interfaces with the semantics and the other with the morphology (e.g., Wechsler & Zlatić 2000, 2003, Smith 2015, Landau 2015). The same can be done to account for Telugu agreement shift. For concreteness, let's adopt the system developed in Smith (2015). Smith proposes that in the narrow syntax, all nominal elements come with interpretable features that interface with the semantics and uninterpretable features that interface with the morphology. In most cases, the two are the same, but this is not always the case. Thus, *committee* nouns in British English have the feature set in (27).

(27) ϕ_{number} = [*u*F:singular, *i*F:plural]

When the nominal is sent to spell-out, the features are split; the *u*Fs are sent to the PF interface, and the *i*Fs are sent to the LF interface.

Once we have two sets of features, we must explain how agreement probes can target both *u*Fs and *i*Fs. Following Chomsky (2000, 2001), let's assume that the locus of the agreement probe for subject agreement is on the T(ense) head, and also that ϕ-features on T are uninterpretable. In this system, this will mean that T only has one set of ϕ-features, which furthermore do not need to be sent to the LF interface. Following Arregi & Nevins (2012) and Bhatt & Walkow (2013), let's also assume that Chomsky's AGREE operation is decomposed into two sub-operations: MATCH and VALUATION. I assume the definition of MATCH in (28) (from Bhatt & Walkow 2013, 972).

(28) MATCHING is a relation that holds of a probe P and a goal G. Not every link induces VALUATION. To do so G must (at least) be in the domain D(P) of P and satisfy locality conditions. The simplest assumptions for the probe-goal system are shown below:

(29) a. Matching is feature identity.
b. D(P) is sister of P.
c. Locality reduces to "closest c-command"

While MATCH is a syntactic relation, the authors above argue that the other sub-operation VALUATION, the actual sharing of features between the probe and the goal, can occur either in syntax or PF. Smith (2015) argues that when VALUATION occurs in the syntax, it may target either the interpretable *i*F or the uninterpretable *u*F of the goal, but if it occurs in the PF component, then only the uninterpretable *u*F is available as a target.[7]

[7]For discussion of and extension to the locality of semantic agreement in the system, see Smith (2015).

3.2 De se attitudes

Following Lewis (1979), Chierchia (1989), Schlenker (1999), von Stechow (2002, 2003), Pearson (2012); a.o., I assume that the complement of an attitude verb is not a proposition, but rather a property of type $\langle e\langle st\rangle\rangle$, as shown in (30).[8]

(30) ATT [λx. λw. [. . . x . . . in *w*]]

Attitude verbs then quantify over individual-world pairs, or *centered worlds*. The denotation of *believe* in this system is given in (31).

(31) a. $[\![believe]\!]^g = \lambda P_{\langle e\langle s,t\rangle\rangle}.\lambda x_e.\ \lambda w_s.\ \forall\langle y, w'\rangle \in \text{DOX}(x, w)[P(y)(w')]$
b. $\text{DOX}(x,w) = \{\langle y, w'\rangle$: w' is compatible with x's beliefs in w and x identifies as y in w'$\}$

To illustrate how *de se* interpretations arise, I provide a sample LF and semantic derivation in (32). The abstractor over individuals in the embedded clause binds the pronoun in its scope; this, along with the meaning postulate in the verbal denotation in (31), will result in the pronoun being interpreted *de se*.

(32) a. [$_{CP1}$ λw_1. [w_1 Pete believes [$_{CP2}$ λx_2. λw_3. [w_3 he$_2$ is smart]]]]
b. $[\![CP2]\!]^g = \lambda$x. λw. x is smart in w
c. $[\![CP1]\!]^g = \lambda$w. $\forall\langle y, w'\rangle \in$ DOX(Pete, w)[y is smart in w'].

I will assume that the LF for *de se* attitude reports for languages that have indexical shift as well as languages with monstrous agreement is that of (32a). How then do we account for the variation? I follow here a modified version of the system developed in Schlenker (1999, 2003) (see also Baker 2008).[9] I assume, following Kratzer (2009), that bound pronouns may be born as variables without any feature values (Kratzer's minimal pronouns). The features are then valued during the course of the derivation. The minimal pronoun is shown in (33).

(33) $x_{[uF:\_;iF:\_]}$

When the minimal pronoun is bound by the abstraction over individuals in the left periphery of the embedded clause, its *u*F and *i*F are valued [AUTHOR; −C*] (cf. Kratzer's

[8] I am putting aside tense as it is inconsequential for my analysis.

[9] As a reviewer notes, there are also similarities between the proposed analysis and the analyses presented in von Stechow (2002, 2003). As noted above, both analyses assume the same LF for *de se* interpretations; however, the locus of variation differs in both analyses. In von Stechow, the locus is the ability for verbs in some languages to check features of embedded pronouns. In the current analysis, the locus of variation is the vocabulary insertion rules of the embedded pronouns.

discussion of relative pronouns and PRO receiving features when bound).[10] The AUTHOR feature marks the pronoun as the *de se* center of some speech context, while the −C* feature indicates that the pronoun is not the author of the actual speech context. Following Schlenker (1999, 2003), the variation between languages discussed here comes from how these features are spelled out. In Telugu, the spell out rule for the first person pronoun specifies that the pronoun must refer to the author of the actual speech act. In languages with indexical shift, the first person pronoun is unspecified for which speech act the pronoun refers to. This is shown in (34).

(34) a. *Telugu* [AUTHOR] [+*C**] ↔ *neenu*
b. *Zazaki* [AUTHOR]↔ *ɛz*

Since the first person pronoun cannot be used in Telugu to refer to the author of an embedded speech act, the pronoun must be out another way. (9b) and (12b) indicate that the embedded pronoun inherits the features of the matrix subject, i.e., when the matrix subject is third person, then the third person pronoun *ṯanu* is used and when the subject is second person the embedded pronoun is also the second person pronoun *nuuvu*. The relevant examples are repeated below in (35).

(35) a. Raju [ṯanu parigeṯṯ-ææ-nu ani] cepp-ææ-Du
Raju 3SG run-PAST-1SG COMP say-PAST-M.SG
'Raju said that he ran.'
b. nuuvu parigeṯṯ-ææ-nu ani nuuvu cepp-ææ-vu
2SG run-PAST-1SG COMP 2SG say-PAST-2SG
'You said that you ran.'

Although these features are transmitted to the embedded pronoun, they do not seem to be interpreted. For illustration, examine the embedded pronouns in (36). Even though John is a woman in all contexts compatible with his hopes, the masculine pronoun can be used in (36a). Similarly, in (36b), the matrix subjects have the singular *de se* thought (i.e., *I am the smartest student in the world*), but the plural pronoun can be used in the embedded clause.

(36) a. John, a transexual, hopes that he will become a woman and that society will accept him.
b. We all think that we are the smartest student in the world.

There are a number of proposals that attempt to deal with these facts (e.g., Schlenker 1999, 2003, von Stechow 2002, 2003, Rullman 2003, Anand 2006, Heim 2008, Kratzer 2009, Landau to appear). I will follow Heim (2008) and Landau (to appear) in assuming that these features are transmitted to the pronoun in the PF component of grammar; hence they

[10] In Baker (2008), the features are lexically specified on the pronoun, however, they must be licensed by being bound by an operator in the left periphery.

are invisible to the interpretation at LF. With the necessary background in place, we can now analyze monstrous agreement.

3.3 Putting it all together

In this section, I will give partial derivations for indexical shift and agreement shift to illustrate how the system developed in the previous sections works. I will begin with indexical shift, as it is simpler. Lets skip ahead to where the embedded TP is already constructed, as this is where the action begins. Assuming the language has subject agreement (like Amharic), the pronoun and T undergo MATCH.

(37) $[_{TP} \ldots \mathrm{T}_{u\phi\_} \ldots [_{vP}\ \mathrm{x}_{[uF:\_,iF:\_]}]]$

The pronoun will move from its thematic position and the the left periphery of the embedded clause is then constructed. The abstractor over individuals binds the pronoun and values its *i*F and *u*F with the AUTHOR and -C* feature values.

(38) $[_{CP}\ \lambda \mathrm{x}.\ \lambda \mathrm{w}.\ [_{TP} \ldots \mathrm{x}_{[uF:\mathrm{AUTHOR};\ \text{-}\mathrm{C}^*,iF:\mathrm{AUTHOR};\ \text{-}\mathrm{C}^*]} \ldots \mathrm{T}_{u\phi\_} \ldots]]$ (BINDING)

The pronoun and T can undergo VALUATION, copying the value of the *i*F of the pronoun onto T.[11]

(39) $[_{CP}\ \lambda \mathrm{x}.\ \lambda \mathrm{w}.\ [\ldots \mathrm{x}_{[uF:\mathrm{AUTHOR};\ \text{-}\mathrm{C}^*,iF:\mathrm{AUTHOR};\ \text{-}\mathrm{C}^*]} \ldots \mathrm{T}_{u\phi\mathrm{AUTHOR};\ \text{-}\mathrm{C}^* \ldots}]]$ (VALUATION)

The clause is then sent to spell out. The result is that the pronoun and the agreement morpheme are spelled out as first person.[12]

Moving to the more complicated case of agreement shift, recall from the discourse scenarios constructed in (16) and (20), that when a pronoun is interpreted *de se* in Telugu, it can either control first person agreement (i.e., *monstrous agreement*) or third person agreement. All the steps will be the same up until (39) for languages with agreement shift. So let us pick back up there. Recall that in Telugu the first person pronoun, *neenu*, can only be inserted with the feature values [AUTHOR] [+*C**], so unlike languages with indexical shift, the pronoun cannot be spelled out here in Telugu. I propose that in this situation feature transmission with the matrix subject can be invoked as a last resort measure to allow the

[11] One may have the worry that VALUATION occurs counter cyclically. There are several ways to overcome this technical problem: one way is to assume, following Chomsky (2008), that all operations within a phase occur simultaneously (or that the cycle is defined on phases); another possibility is to follow Frampton & Gutmann (2000) and assume that once the pronoun and T undergo MATCH (their *feature sharing*), their features can be valued simultaneously when the pronoun is bound.

[12] VALUATION could occur at PF as well, however the result would be the same because for languages with indexical shift there is never a mismatch between *u*F and *i*F.

pronoun to be spelled out (see Messick (in preparation) for discussion of the last resort nature of feature transmission). Following Heim (2008) and Landau (to appear), I assume that feature transmission is post syntactic feature sharing between two DPs.[13] This is shown for a third person matrix subject in (40).[14]

(40) $\overset{\text{FEATURE TRANSMISSION}}{\overbrace{\phantom{\text{Subject}_{\text{M.SG}}\ \text{ATT}\ [[\ldots x_{[uF:\text{M.SG}]}}}}$ $\text{Subject}_{\text{M.SG}}\ \text{ATT}\ [[\ldots x_{[uF:\text{M.SG}]} \ldots T_{u\phi\text{AUTHOR};\ \text{-C*}} \ldots]]$

As a result of feature transmission with the matrix subject, the pronoun can now be spelled out, but as a third person pronoun. However, the T probe and the pronoun had already undergone VALUATION in the syntax, resulting in the *i*F features of the pronoun being copied onto T. Since these features were [AUTHOR] [$-C$*], in this derivation, the agreement morpheme is spelled out as first person. Recall from the previous section that VALUATION can be delayed until PF. What would happen if VALUATION between the pronoun and the embedded agreement probe on T is delayed until PF? For languages with indexical shift, it is inconsequential (see footnote 12). For languages with agreement shift, since feature transmission results in a mismatch between *u*F and *i*F, if VALUATION took place in PF this means that it would only have access to the *u*F of the pronoun. In (40), this would result in the M.SG being copied onto T, as shown in (41).

(41) $\text{Subject}_{\text{M.SG}}\ \text{ATT}\ [[\ldots \overset{\text{VALUATION}}{\overbrace{x_{[uF:\text{M.SG}]} \ldots T_{u\phi\text{M.SG}}}} \ldots]]$

This will result in both the pronoun and agreement morpheme being spelled out as third person. The optionality of agreement shift occurring with *de se* pronouns then boils down to the timing of VALUATION: if VALUATION takes place in the syntax, the *i*Fs of the pronoun are copied onto T, resulting in the agreement morpheme being spelled out as first person. If VALUATION takes place in PF, the *u*Fs of the pronoun are copied onto T; this results in the agreement morpheme being spelled out as third person in the cases above, as the pronoun had its *u*F valued by feature transmission with the third person matrix subject. Now recall from (16) that when the pronoun is not read *de se*, it crucially cannot control agreement shift (i.e., it must control third person agreement). The system developed here also accounts for this; if the pronoun is not *de se*, this means that it was not bound by the abstraction over individuals in the embedded clause, hence cannot receive the AUTHOR feature. Hence, the

[13] A reviewer notes that there needs to be some way to ensure that the embedded pronoun will undergo feature transmission with the matrix subject and not some other DP (e.g., a matrix object). This could potentially be done through the use of Pearson (2012)'s [ATT] feature.

[14] One may worry that feature transmission appears to cross a finite CP boundary which constitutes a phase. There are several ways to get around this technical problem. One way is to follow Kratzer (2009) and assume that feature transmission chains are broken into smaller steps: the matrix subject shares its features with *v* which in turn shares them with C which can then transmit them to the pronoun. Another alternative is to follow Bošković (2007) and assume that agreement operations (but not movement operations) can occur across phases.

T probe will never have access to the first person features unless the pronoun is read *de se*.[15]

4. Comparison to Sundaresan (2012)

Since this paper attempts to cover similar data as Sundaresan (2012), I will briefly compare the two analyses. Sundaresan (2012) treats monstrous agreement as a sub type of the phenomenon known as the anaphor agreement effect (AAE) (Rizzi 1990) stated in (42).

(42) Anaphors do not occur in syntactic positions construed with agreement.

Sundaresan treats Tamil *taan* as a long distance reflexive anaphor, hence it is subject to the condition in (42). Sundaresan further assumes that the left periphery of complements of verbs of communication contains a perspective phrase that contains a null pronoun in its specifier. In the case of monstrous agreement, this null pronoun will have first person features. When the ϕ-probe on T undergoes search, it encounters *taan*, however it cannot agree with it due to (42). It continues to probe upwards until it reaches the null pronoun in the specifier of the Perspective projection. This null pronoun will value the ϕ-probe on T. This is shown schematically in (43).

(43) [*PerP* *pro*$_{\phi:1st}$ [*Per′* [*TP* taan [*T′* [*vP*...] T$_{\phi:_}$]] Per]]
AGREE

The main difference between the analysis presented in Sundaresan (2012) and the one presented here is what we take the controller of agreement to be. The analysis presented here allows for the matrix subject of the embedded clause to be the controller, while Sundaresan (2012) argues that it is a null pronoun. These different analytical choices lead to different empirical predictions. Specifically, Sundaresan (2012) predicts that monstrous agreement should only occur when the embedded subject is an anaphor as that is the only time the ϕ-probe on T would probe beyond the subject.

As I discussed in section 2.1, Telugu appears to violate this prediction since *ṯanu* is no longer used as an anaphoric element as evidenced by the fact it can used in larger range of environments than anaphors. Outside of Dravidian there are a number of other languages that allow for non-anaphoric third person pronouns to control first person agreement. Karimonjong for instance allows for this agreement pattern, as shown in (44) (Curnow 2002). (45) shows that the pronoun can be the subject of an out of the blue context suggesting again that this is not an anaphor.

(44) àbù papà tlim ɛbè àlózì iŋèz morotó.
AUX father say that 1SG-go-NEST 3SG Moroto
'The father said that he was going to Moroto.'

[15]For *de se* interpretations of *de re* pronouns, it is possible to have the relation between the attitude holder and the *res* be one of identity. This way we can also have the interpretation with the centered world semantics.

(45) íkítacapi iŋèz ìwɔn
weed.CAUSE 3SG 1INCL
'He will cause us to weed.'

Even putting aside the status of *t̪anu*, the Telugu data presented here is problematic for the account given in Sundaresan (2012) since it was shown that the second person pronoun *nuuvu* also had the ability to control monstrous agreement, as the repeated example shows below:[16]

(46) nuuvu parigett̪-ææ-nu ani nuuvu cepp-ææ-vu
2SG run-PAST-1SG COMP 2SG say-PAST-2SG
'You said that you ran.'

This type of example is unproblematic on the account presented here as the embedded pronoun and matrix subject will share features via the feature transmission mechanism discussed in section 3.[17]

5. Extended typology

Let us now examine two other ways that languages use to mark *de se* attitudes. One way is through the use of logophoric pronouns.[18] Now if logophors can also be read *de se*, this means that they can also be bound by the *de se* center. Under the current system, this would lead us to expect to find logophors which can control first person agreement in some languages, as they would also receive AUTHOR features when bound by the embedded abstraction over individuals. And indeed, we do find such languages, as shown in (47) for Donno Sɔ (Culy 1994).

(47) Oumar inyemɛ jɛmbɔ paza bolum miñ tagi
Oumar LOG sack.DF drop left.1SG 1SG.OBJ informed
'Oumar told me that he had left without the sack'

We have seen that the system in place can account for languages with indexical shift, languages with agreement shift, and languages with logophors (with and without first person agreement). Now how does the system handle a language like English, where there is no

[16]A reviewer points out that only bound uses of these pronouns are able to control monstrous agreement. In order for the data presented here to fall under the AAE, either the condition in (42) would have to be amended to block agreement with bound uses of pronouns or we would have to posit an ambiguity between bound uses of pronouns and their free uses.

[17]If the controller of monstrous agreement is indeed *taan* in Tamil, we still need a explanation of why it appears to violate the anaphor agreement effect. One option is to claim that the AAE is paramaterized and agreement with anaphors is allowed in Tamil. Another option would be to treat *taan* not as an anaphor but rather a logophoric like pronoun along the lines of *taan* in Malayalam (Anand 2006).

[18]Note that I am not saying that logophoric pronouns are *obligatorily de se*, as Pearson (in press) has shown that logophors can be read *de re*. I am making the weaker claim that such pronouns *can* be used to express *de se* attitudes.

marking of *de se*? There are two potential ways: one way is to follow Anand (2006), and assume that *de se* readings in English only occur as a special form of *de re*; another way is to assume that the LF for English *de se* is the same as for other languages, but that agreement in English only occurs in the PF component for person agreement. This would mean VALUATION of the probe on T would only have access to the *u*F of the pronoun. Both options are fully compatible with the system developed here. Taking into account all the languages discussed here, we have the following typology of embedded *de se* marking.

(48) *Typology of de se marking*

Language	*de se* marking
English	Third person pronoun
Amharic, Zazaki	Indexical shift
Ewe	Logophor
Donno Sɔ	Logophor with first person agreement
Telugu	Third person pronoun with first person or third person agreement

The system developed here has been shown to have the flexibility to account for all of this variation. Notice, however, that there is a gap in the typology in (48). In this hypothetical language, a *de se* attitude would be expressed with a first person pronoun and third person agreement, as shown in (49).[19]

(49) John said I is a hero.
Intended: 'John said that he$_{dese}$ is a hero'

In the current system, in order to be interpreted as *de se* and be spelled out as a first person pronoun both the *u*F and *i*F would be first person, so no matter when VALUATION takes place first person features will always be copied onto T. Hence, the agreement pattern from (49) is underivable under the system developed here, allowing for a principled explanation of this typological gap.

To conclude, this paper presented new data from Telugu that showed a new way of marking *de se* attitudes. It was shown that monstrous agreement behaves in many respects like indexical shift. The system developed here to account for monstrous agreement can also account for other ways languages mark *de se*, additionally explaining a hitherto unnoticed typological gap.

References

Anand, Pranav. 2006. De De Se. Doctoral dissertation, MIT, Cambridge, Massachusetts.
Anand, Pranav, & Andrew Nevins. 2004. Shifty operatos in changing context. In *Proceed-*

[19]It is not the case that this surface pattern does not exist; it does in e.g., Golin (Papuan). However, when a first person pronoun controls third person agreement in such languages, it is always interpreted as a *de re* attitude about the current speaker. See Messick (in preparation) for an analysis.

ings of SALT 14, ed. Robert B Young, 20–37.
Arregi, Karlos, & Andrew Nevins. 2012. *Morphotactics: Basque auxilaries and the structure of spellout*. Springer.
Asher, R. E. 1985. *Tamil*. London: Croom Helm.
Baker, Mark C. 2008. *The syntax of agreement and concord*. Cambridge: Cambridge University Press.
Bhatt, Rajesh, & Martin Walkow. 2013. Locating agreement in grammar. *Natural Language and Linguistic Theory* 31:951–1013.
Bošković, Željko. 2007. On the locality and motivation of move and agree:An even more minimal theory. *Linguistic Inquiry* 38:589–644.
Chierchia, Gennaro. 1989. Anaphora and attitudes *de se*. In *Semantics and contextual expressions*, ed. van Benthem Bartsch & van Ernde Boas, 1–31. Kluwer/Reidel.
Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step*, ed. Roger Martin, David Michaels, & Juan Urigareka, 89–115. Cambridge, Massachusetts: MIT Press.
Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 1–52. MIT Press.
Chomsky, Noam. 2008. On Phases. In *Foundational issues in linguistic theory*, ed. Robert Freidin, Carlos P. Otero, & Maria Luisa Zubizarreta, 133–166. Cambridge: MIT Press.
Corbett, G. G. 1979. The agreeement hierarchy. *Journal of Linguistics* 15:203–395.
Corbett, G. G. 1983. *Hierarchies, targets and controller: Agreement patterns in Slavic*. London: Croom Helm.
Corbett, G. G. 2006. *Agreement*. Cambridge: Cambridge University Press.
Culy, Christopher. 1994. Aspects of logophoric marking. *Linguistics* 32:1055–1094.
Curnow, Timothy Jowan. 2002. Three types of verbal logophoricity in African languages. *Studies in African Linguistics* 31.
Frampton, John, & Samuel Gutmann. 2000. Agreement is feature sharing. Ms., Northeastern University.
Heim, Irene. 2008. Features on bound pronouns. In *Phi theory: Phi-features across modules and interfaces*, ed. Daniel Harbour, David Adger, & Susana Bejar, 36–56. Oxford University Press.
Kratzer, Angelika. 2009. Making a pronoun. *Linguistic Inquiry* 40:187–237.
Krishnamurti, BH., & J.P.L. Gwynn. 1985. *A grammar of modern Telugu*. Delhi: Oxford University Press.
Landau, Idan. 2015. DP-internal semantic agreement: A configurational analysis. Ms., Ben Gurion University.
Landau, Idan. to appear. Agreement at PF. *Syntax* .
Lewis, David. 1979. Attitudes de dicto and de se. *Philosophical Review* 88:513–543.
Messick, Troy. in preparation. Agreement features at the syntax-semantics interface [working title]. Doctoral dissertation, University of Connecticut, Storrs, Connectitcut.
Pearson, Hazel. 2012. The sense of self. Doctoral dissertation, Harvard University, Cambridge, Massachusetts.
Pearson, Hazel. in press. The interpretation of the logophoric pronoun in Ewe. *Natural Language Semantics* .

Percus, Orin, & Uli Sauerland. 2003. On the LFs of attitude reports. In *Proceedings of Sinn und Bedeutung 7*, 228–242.

Rizzi, Luigi. 1990. On the anaphor-agreement effect. *Rivista di Linguistica* 2:27–42.

Rullman, Hotze. 2003. Bound variable pronouns and the semantics of number. In *Proceedings of WECOL 2002*, ed. Brian Agbayani, Paivi Koskinen, & Vida Samiian, 243–254.

Schlenker, Philippe. 1999. Propositional attitudes and indexicality. Doctoral dissertation, MIT.

Schlenker, Philippe. 2003. A plea for monsters. *Linguistics and Philosophy* 26:29–120.

Sells, Peter. 1987. Aspects of logophoricity. *Linguistic Inquiry* 18:445–479.

Smith, Peter. 2015. Feature mismatches: Consequences for syntax, morphology and semantics. Doctoral dissertation, University of Connecticut, Storrs, Connectituct.

von Stechow, Arnim. 2002. Binding by verbs. Ms., University of Tübingen.

von Stechow, Arnim. 2003. Feature deletion under semantic binding. In *Proceedings of NELS 33*, 397–403.

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

Sudo, Yasutada. to appear. Person indexicals in Uyghur indexical shift. In *Proceedings of BLS 36*.

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

Wechsler, Stepehen, & Larisa Zlatić. 2000. A theory of agreement and its application to Serbo-Croatian. *Language* 76:799–832.

Wechsler, Stepehen, & Larisa Zlatić. 2003. *The many faces of agreement*. Stanford, CA: CSLI Publications.

Woolford, Ellen. 1999. More on the anaphor agreement effect. *Linguistic Inquiry* 30:257–287.

Troy Messick
troy.messick@uconn.edu

Obligatory Additive Particle on Negation

Ishani Guha

MIT

1. Introduction

In Bangla, the negated complement of the existential modal predicate *par-* (in epistemic and in deontic readings) shows an obligatory additive particle '-o'.

Epistemic use of the modal:

(1) Ritu ækhon baṛi-te na-**#(o)** thak-te pare
Ritu now home-loc neg-O stay-inf may.pres.3
'It is possible that Ritu is not at home now.' [may > neg]
#'It is not possible that Ritu is at home now.' [# neg > may]

(2) Ritu ækhon baṛi-te na thak-te-**#(o)** pare
Ritu now home-loc neg stay-inf may.pres.3
'It is possible that Ritu is not at home now.'

While (2) is more general, (1) is particularly used for polarity contrast. I would concentrate on analyzing (1) in this squib.

Deontic use of the modal:

(3) tumi ice-cream-ṭa na-**#(o)** khe-te paro
you icre-cream-cla. neg-O eat-inf may.pres.2
'You are allowed to not eat the ice-cream.'

The paper tries to understand this particular instance of obligatoriness of the Additive particle.

2. Background

2.1 Background on Negation in Bangla

The position of sentential negation in Bangla usually depends on the presence or absence of the Tense inflection on the verb. The negation follows a verb form that is marked with

Tense ((4) below) while it precedes a verb form lacking Tense inflection in embedded clauses (see Simpson and Sourov (2014)). The protasis of conditional provides an exception to this distribution, as in this clause, the negation occurs before a verb form marked with Tense (see Ramchand (2014) for an analysis). For the purposes of this paper it is important to note that only the negation in pre-verbal position can host an Additive particle (or any emphatic clitic). Moreover, even among pre-verbal negation, the appearance of the Additive particle is limited to only two constructions: the cases illustrated in (1) through (3) and in Concessive conditionals. I refrain from discussing Concessive Conditionals in this paper for the sake of brevity.

In examples (1) through (3), the existential modal has scope over negation. To get the negation over modal reading, the negation has to be placed after the modal. Example (4) with post-verbal negation on the matrix modal predicate *pare* represents the negation over modal reading.

(4) Ritu ækhon baṛi-te thak-te pare **na**
Ritu now home-loc stay-inf may.pres.3 neg
'It is not possible that Ritu is at home now.' [neg > may]
#It is possible that Ritu is not at home now. [# may > neg]

2.2 Background on the Additive Particle

The Additive Particle *-o* is one of the two 'emphatic' clitics in the language. Bayer and Lahiri (1990) (among others) discusses the distribution of the particle and gives a syntactic account of its scope. The particle *-o* triggers an Additive presupposition about its Associate, which is often the constituent the particle cliticizes to.[1] The following examples illustrate the variation in meaning resulting from the different positions of *–o* in simple sentences.

On Subject:
Context: Manoj came to the party last night, and…

(5) Robi**-o** eSechilo.
Robi-O came.3
'$[\text{Robi}]_F$ came too.'

The sentence asserts *that Robi came* and the *–o* on *Robi* gives rise to the presupposition *that somebody other than Robi came.*

The presence of *–o* on the subject triggers alternatives of the sentence, which are of the form [came(x): $x \in D_e$] and the particle is anaphoric to atleast one such alternative salient in the context.

[1] The particle can also mean 'even' in appropriate contexts. It is a close parallel of Hindi *–bhii* as discussed in Lahiri (1998).

On Object:
Context: Robi read the magazine, and…

(6) Robi boi-ṭa-**o** poṛeche
Robi book-cla-O read.3
'Robi read [the book]$_F$ as well.'

The sentence asserts *that Robi has read the book* and the *–o* on *book* gives rise to the presupposition *that other than the book there is something else that Robi has read.*

With *–o* on the object in the given context, we are considering alternatives of the form [read(r, x): x ∈ D_e].

On Intransitive verb:
Context: Robi has done the assignment, he went to the party, and…

(7) Robi ghumiye-**o**-che
Robi sleep.pfv-O-asp.pres.3
'Robi has [slept]$_F$ as well.'

The sentence asserts *that Robi has slept* and the *–o* on *slept* gives rise to the presupposition *that other than sleeping, Robi has done something else (*i.e., *some other predicate is true of Robi).*

The additive particle marks alternatives of the form [R(r): R∈ $D_{<et>}$]

On a Transitive verb:
(8) Robi boi-ṭa poṛe-**o**-che
Robi book-cla read-pfv-O-pfv-pres.3
'Robi has [read]$_F$ the book as well.'

In (8), *-o* attaches to the verb, triggering relevant alternatives of the form [R(r, the book): R∈ $D_{<e,et>}$] . The construction is usable in contexts where there is an antecedent that entails that Robi did something else with the book. For example,

(9) Robi Sudhu boi-Ta kene-i-ni, Robi boi-Ta **poṛe-o-che**.
Robi only book-cla buy.pfv-I-neg-pfv, Robi book-cla read.pfv-O-asp.pres.3
'Robi not only [bought] the book, he [read]$_F$ it as well.'

In the cases above (5 to 9), *–o* clearly marks the constituent which has to be replaced with elements of the same type, to obtain the focus value of the sentence. Being an additive particle it presupposes that one of the (contextual) alternatives of the sentence is true.

There are constructions in which the constituent that is *–o* marked and the constituent that is actually focused do not match. The clearest example at hand is a construction like (6) above, except used in a different context.

On Object:
Context: Robi watered the plants and…

(10) … o boi-ṭa**-o** poṛeche
pron.3 book-cla-O read.3
'…he read the book as well.'

In (10), the Additive presupposition is *that Robi did something else, other than reading the book.* Therefore, even though *–o* appears on the object, the entire VP is focused, as is fitting to the context, and consequently, the relevant alternatives of the sentence are of the form $[R(r) : R \in D_{<et>}]$.

Considering just these cases, it seems the occurrence of *–o* is like an F-marker with usual F-projection properties.[2] It can be shown that *–o* behaves like the additive operator as well (Guha 2016ms.).

3. Additive particle on negation

3.1. The meaning of the construction

Assuming the Additive attribute of *–o,* the meaning of sentence (1) would be:

(1) Ritu ækhon baṛi-te **na-#(o)** thak-te pare
Ritu now home-loc neg-O stay-inf may.pres.3
'It is possible that Ritu is not at home now.'

a. Assertion: It is possible that Ritu is not at home now.
b. Presupposition: It is possible that Ritu is at home now.

So in effect, uttering (1) in Bangla amounts to uttering *it is also possible that Ritu is not at home now* in English.

Interestingly, the following sentence shows that this intuition is on the right track. Consider:

(11) Ritu ækhon baṛi-te **na-(#o)** thak-te pare **na**[3]
Ritu now home-loc neg-O stay-inf may.pres.3 neg
'It is not possible that Ritu is not at home now.'

Here *–o* cannot appear on the embedded negation, because if it did it would lead to a contradictory Additive presupposition as predicted by the meaning above. Let us see the meanings of the infelicitous version of (11), which is (12) with *na-o* in the embedded clause.

[2] *–o* does not appear on attributive adjectives (*[A-o NP]), or on nouns inside postpositional phrases (*[NP-o P]), and the idea of *–o* being an F-marker would require further qualifications for such cases. (see Bayer and Lahiri (1990), for a syntactic proposal)
[3] A metalinguistic use of the matrix negation might make the construction acceptable with *–o*.

(12) #Ritu ækhon baṛi-te **na-o** thak-te pare **na**
Ritu now home-loc neg-O stay-inf may neg

a. Assertion: It is not possible that Ritu is not at home now.
b. Additive presupposition: It is not possible that Ritu is at home now.

Clearly (12a) and (12b) are contradictory. Within the same domain of worlds, Ritu has to either be at home or not be at home. Both of these cannot be 'not possible' (impossible) as the assertion and the presupposition would yield. That is why the use of *–o* in this sentence is infelicitous.

3.2. The affirmative complement of *par-*

In contrast with the negative complement that we have discussed so far, the affirmative complement of the existential modal predicate *par-* does not appear with an obligatory additive particle.

(13) Ritu ækhon baṛi-te thak-te**-(o)** pare
Ritu now home-loc stay-inf-(O) may.pres.3
'It is possible that Ritu is at home now.'

3.3. Context dependency

A brief survey reveals that the obligatoriness of *–o* in the complement of the existential modal is dependent on the context. When the antecedent entails that *it is possible that Ritu is at home now*, the utterance must contain the additive particle in the negative complement.

Scenario 1:

A: Ami Ritur baṛi jacchi. O Sadharonoto ækhon baṛitei thake.
'I am going to Ritu's place. She is usually at home now.'

B: Right, but

(14) Ritu ækhon baṛi-te **na-#(o)** thak-te pare
Ritu now home-loc neg-O stay-inf may.pres.3
'It is also possible that Ritu is not at home now.'

In Scenario 1, A's utterance entails the proposition *it is possible that Ritu is at home now*, which is also what B's utterance (14) presupposes by the particle *–o*.

On the other hand, when the antecedent entails *it is possible that Ritu is not at home now*, the utterance must contain the additive particle in the affirmative complement.

Scenario 2:

A: Ritu Sadharonoto ækhon baṛi-te thakena.
'Ritu is usually not at home at this hour.'

B: Right, but,

(15) Ritu ækhon baṛi-te thak-te-**#(o)** pare
Ritu now home-loc stay-inf-O may.pres.3
'It is also possible that Ritu is at home now.'

In Scenario 2, A's utterance entails the proposition *it is possible that Ritu is not at home now*, which is also what B's utterance (15) presupposes by the particle *–o*.

To summarize, when the context takes it for granted that [◊ p], and the speaker wants to assert that [◊ (¬ p)], the additive particle has to be used to refer to the presupposed possibility. Similarly, in a context that takes [◊ (¬ p)] for granted, the assertion [◊ p] must signal that the other possibility is also available.

Note that, *p* and *¬(p)* are mutually exclusive and cannot hold in the same world.

3.4. The meaning contribution of *–o* as an anaphor

Kripke (1990ms./2009) argued that *too* is an anaphor that refers to parallel information in the `active' context. Heim (1992) had formally represented the proposal of Kripke by making *too* coindexed with the relevant constituent in the antecedent. For example, notice the indexing in the following example.

(16) Piku_i lives in Delhi and Anu_i lives in Delhi too_i.

which says: Anu lives in Delhi, in addition to Piku.

Heim takes the general rule for the interpretation of *too* to be:

(17) $\phi[\alpha_F]\text{too}_i$ presupposes $x_i \neq \alpha \wedge \phi(x_i)=1$

Chemla and Schlenker (2012), analyzed *too$_i$* to be a propositional anaphor. They defined *too$_i$* in Rooth's Alternative semantics. An expression *E* has an ordinary semantic value $[\![E]\!]_o$ and a focus semantic value $[\![E]\!]_f$. The interpretation function is relativized to an assignment function *g* and an evaluation world *w*. If too_i takes a clause ϕ as its argument, the result will have the value in (18). (`#' means `undefined')

(18) $[\![\phi\ \text{too}_i]\!]^{g,w}_o = \#$ unless,

a. $g(i)(w)=1$

b. $[\![g(i)]\!]_o \in [\![\phi]\!]^g_f$

c. $[\![g(i)]\!]_o \neq [\![\phi]\!]^g_o$

If $[\![\phi\ \text{too}_i]\!]^{g,w}_o \neq \#$, then $[\![\phi\ \text{too}_i]\!]^{g,w}_o = [\![\phi]\!]^{g,w}_o$

Among the definedness conditions, (18.a) says the proposition that *too$_i$* is anaphoric to is true; (18.b) says the ordinary semantic value of the proposition is an element of the focus

semantic value of the utterance ϕ; (18.c) says the ordinary semantic value of that proposition is different from the ordinary semantic value of ϕ. The last line says that *too*$_i$ does not have any contribution in the asserted content of ϕ.

For the additive particle *–o*, I adopt Heim's analysis in Chemla and Schlenker's formulation. That is, I take *-o* to be anaphoric to a constituent that is an alternative of the focused constituent in the utterance. In (14) above, the focused constituent is *NOT*.

In Bangla, when two embedded Intransitive-*vP*s are contrasted *-o* can appear only on the verb and not on the negation in the negated *vP*.

(19) Robi aste-o pare, na jete-o pare
Robi come.inf-O may, not go.inf-O may
`Robi may come and not go as well.'

(20) *Robi aste-o pare, na-o jete pare
Robi come.inf-O may, not-O go.inf may
`Robi may come and not go as well.'

However in Polarity contrast, the best option is to have *-o* on negation.

(21) Robi aste-o pare, na-o aste pare
Robi come.inf-O may, not-O come.inf may
`Robi may come and he may not come as well.'

The example (21) is like example (14) in Scenario 1, which motivates the conclusion that *NOT* is focused in (14).

Going back to the meaning of (14) in Scenario 1, the antecedent entails [It is possible that Ritu is at home], which can be represented as [MAY [AFF$_i$ [Ritu be at home]]], where `AFF$_i$' is the affirmative operator, the relevant alternative of *NOT*. Considering *–o* as the anaphor, *-o* bears the index '*i*'.

The utterance (14) in Scenario 1 can be represented as:

(22) [Not-o$_i$] λ_7 [MAY [t$_7$ [Ritu be at home]]]

Let, ϕ be the function [λ_7 [MAY [t$_7$ [Ritu be at home]]].

Then,

(23) $[\![(13)]\!]^{g,w}_o = [\![[\text{NOT- o}_i]\ \phi\ ]\!]^{g,w}_o$ = # unless,
a. ϕ (g(i))(w)=1
b. $[\![g(i)]\!]_o \in [\![\text{NOT}]\!]^g_f$

c. $[\![g(i)]\!]_o \neq [\![NOT]\!]^g_o$

… where g(i) = AFF

These definedness conditions form the meaning contribution of *–o* in (14)[4] :

(24) $[\phi\ (AFF)(w)=1 \wedge [\![AFF]\!]_o \in [\![NOT]\!]^g_f \wedge [\![AFF]\!]_o \neq [\![NOT]\!]^g_o]$

3.5. Asymmetry in availability of alternatives

There is an inherent asymmetry between [◊ (p)] and [◊ (¬ p)]. For [◊ (¬ p)] the additive particle seems obligatory, but for [◊ (p)] that is not the case, until one adds [◊ (¬ p)] in the context.

In terms of alternatives, this asymmetry can be restated as, [◊ (¬ p)] by default has [◊ (p)] as a formal alternative, but not vice versa. [◊ (p)] can have [◊ (¬ p)] as an alternative only when it is made salient as a contextual alternative.

The asymmetry in terms of alternatives can be readily captured in terms of the notion of (structural) Complexity as defined for the derivation of Formal alternatives in Fox and Katzir (2011).

(25) Focus-sensitive version of Complexity from Fox and Katzir (2011):

The set of formal alternatives of *S*, *F(S)* is defined as the set of all structures obtained from *S* by replacing focused constituents within *S* with constituents that are at most as complex as the original constituents.

(26) Source of substitution for a given constituent X in context C:
   a. The lexicon
   b. The sub-constituents of X
   c. The set of salient constituents in C

For the case under discussion, Complexity, will allow (27) to have (28) as one of its formal alternatives.

(27) [It is possible that John is not$_F$ at home]
(28) [It is possible that John is at home]

To implement Fox and Katzir (2011)'s notion of Complexity in this case appropriately, I would have to assume ∑P (Laka 1990), so that *NEG* can have *AFF* as an alternative.

(29) [It is possible that John is $[_{\Sigma P}$ $[not]_F$ at home]][5]
(30) [It is possible that John is $[_{\Sigma P}$ [Aff] at home]]

[4] In Guha(2016ms.) I discuss problems with this meaning (24) and suggest a modification.
[5] For Bangla, ΣP, above TP, would be the highest projection below CP. The clause that does not have ΣP will have TP as its highest projection.

But, Complexity will prevent (31) from having (32) as a formal alternative, as (32) is structurally more complex than (31), because according to Laka's proposal a simple affirmative sentence (31) does not have ∑P, but (32) does.

(31) [It is possible that John is at home]
(32) [It is possible that John is [$_{\sum P}$ not at home]]

Only when (32) is made salient in the context, (32) would be part of the set of alternatives of (31).

Note, if we assume that ∑P has a different type from TP, then a type-based theory of alternatives could also derive the asymmetry.

There is an interesting aspect of using ∑P syntax here. Laka argues that emphatic affirmatives (*John did come*) have ∑P in their structure. So the prediction would be, when [◊ (p)] is an emphatic affirmative, [◊ (¬ p)] would be available as a formal alternative. So for affirmatives, we have two options, [◊ (p)] and [◊ (p)-***o***]. The latter could represent the emphatic affirmative, which has [◊ (¬ p)] as alternative. For [◊ (¬ p)] we do not have such structural ambiguity.

3.6. Implicatures

For sentence (14) in Scenario 1, the utterance is :

(33) φ = [◊ (¬ p)] ∧ [◊ p]

Since *NOT* is F-marked, the only formal alternative of *φ* in context C that we consider is:

(34) ALT (φ,C) : [◊ p] ∧ [◊ p]

The ALT in (34) is entailed by the *φ*. Therefore, it does not lead to a S.I. (Secondary Implicature)

If we considered just the prejacent *ψ* (35) of the additive particle, and that the entire sentence is focused, so that both the modal and the negation are focused, then we derive mutually exclusive alternatives to *ψ*.

(35) ψ = ◊ (¬ p)

(36) ALT(φ,C) :
a. ◊ p
b. □ (¬ p)
c. □ p

Assuming an opinionated speaker,

(37) Quality Implicature(Q.I.) : B_S [◊ ¬(p)] (the speaker believes that [◊ ¬(p)])

(38) Primary Implicatures(P.I.) :

a. ¬ B_S [◊ p]
b. ¬ B_S [□ (¬ p)]

The ALTs (36.a) and (36.b) are mutually contradictory and (36.c) contradicts the Q.I. (and 36.b). Therefore, none of them would lead to Secondary Implicature (S.I.), and we would get Ignorance Inferences from (38.a) and (38.b).

However, the ALT (36.a) is presupposed, so that cannot become an Ignorance Inference. In that case, ALT (36.b) can lead to a S.I. that is consistent with the Q.I. and we don't derive any Ignorance Inferences.

(39) Secondary Implicature(S.I.) : B_S ¬ [□ (¬ p)]

As it happens, the S.I. (39) is equivalent to the presupposition [◊ p].

3.7. Deriving Obligatoriness: Maximize Presupposition!

The maxim of Maximize Presupposition states, "make your contribution presuppose as much as possible!" (from Sauerland 2008's interpretation of Heim 1991) The maxim applies at situations when we encounter a competition between a pair of contextually equivalent sentences S and S', such that S has a presupposition and S' does not. S' can be used only if the speaker is certain that the presupposition is not common ground or if the speaker is not certain whether the presupposition is common ground.

Applying the maxim to the case at hand:

The context of the utterance entails: [It is possible p]

Utterance:

(40) It is also possible ¬ p.

The prejacent of the Additive particle is the Assertion :

(41) ψ = [It is possible ¬ p].

The Presupposition is (42) which is the formal and contextual Alternative of (41):

(42) π = [It is possible p].

The utterance is:

(43) [ψ ∧ π]

If the speaker did not use the Additive particle, she would in effect utter just ψ.

Uttering ψ would implicate ¬ [φ ∧ π], which in turn would implicate that the presupposition *π* is not common ground.

(44) ψ ⇝ ¬ [ψ ∧ π] = ¬ ψ ∨ ¬ π

(¬ ψ) is ruled out by the utterance ψ. Therefore, ψ implicates (¬ π).

(45) [π] = [It is possible p]

[¬ π] here indicates that the speaker knows that π is not satisfied or more crucially, that the speaker does not know whether π is satisfied.

However, [It is possible p] is already part of the common ground. Hence, the Implicated Presupposition [Certain ¬ (It is possible p)] or [¬ Certain (It is possible p)] is incompatible with this discourse. Therefore, to avoid the Implicated Presupposition, the speaker has to Maximize Presupposition!

4. Extending the analysis

4.1.1. Discussion of other Mutually Exclusive cases

We have discussed the case of *p* and *¬(p)* under an existential modal predicate. Since, *p* and *¬(p)* cannot hold in the same world, in a context that takes *p* for granted one cannot felicitously assert [◊ (¬ p)]. In other words, the narrow scope of the Additive particle is contradictory in the case under discussion, because that would make the assertion and the presupposition contradictory.

Let us now look at cases involving mutually exclusive alternatives under the possibility modal, which are equally 'complex'.

Scenario 4:

News on TV:
'The World Chess Championship is being played between Anand and Carlsen. Carlsen has already won two games.'

A: ækhono pa͡c-Ta game baki ache.
'There are still five games to go.'

(46) **Anand-#(o)** Championship-Ta jit-te pare
Anand-O championship-cla win-inf may.pres.3
'Anand may win the Championship too.'

The context makes the possibility of Carlsen winning the Championship salient, since he is already two games up. But A thinks that Anand's chance of winning is still open. A takes the context to entail [◊ win (carlsen)] and utters: [◊ win (anand)]-also.

Since, the antecedent is [◊ win (carlsen)], *also* in the utterance (42) is anaphoric to it, and not just to the complement clause [win (carlsen)] at the exclusion of the modal. In other words, *also* has wide scope in this context.

(47) Wide scope:
   a. Assertion: It is possible that Anand will win.
   b. Additive presupposition: It is possible that Carlsen will win.

Given [win (anand)] and [win (carlsen)] are mutually exclusive alternatives, the narrow scope of *also* is ruled out.

(48) Narrow scope:
   a. Assertion: It is possible that Anand will win.
   b. (#) Additive presupposition: that Carlsen will win.

Thus effectively we will get the inference (due to 'local effect') *it is possible that Anand and Carlsen will win,* which is not permissible in the given scenario.

4.1.2. Regular mutually compatible cases

As is evident from the discussion above, if we do not set up mutually incompatible opposition then the so called narrow scope reading would become possible.

Scenario 5:
Ritu wants to visit Robi and Shomir now.

A: Robi is at home now.

B:

(48) **Somir-#(o)** ækhon baṛi-te thak-te pare
Shomir-O now home-loc stay-inf may.pres.3
'It is possible that Shomir is also at home now.'

(49) Narrow scope:
   a. Assertion: It is possible that Shomir is at home.
   b. Additive presupposition: Robi is at home.

In the context of what A says in the conversation above, (49.b) is the salient additive presupposition. This presupposition is obtained by the narrow scope of the Additive particle.

If the scenario was set appropriately, in this case the wide scope could be infelicitous. It is easy to see that with the deontic modal *par-* `may'.

Scenario 6:

A: Robi had the ice cream, though he was not allowed to have it.

(50) #Somir-o icecreamTa khe-te pare
Shomir-O icecream eat-inf may.pres.3
a. 'It is allowed that Shomir also eats the icecream.'
b. # 'It is also allowed that Shomir eats the icecream.'

The reading in (50.b) in Scenario 6, looks like the wide scope reading with the inference *it is allowed that Shomir and Robi eats the icecream*. However, this inference is slightly diffent from the wide scope reading and is due to what has been called the `obligatory local effect' (Tonhauser et al. 2013). The narrow scope of the additive along with the projection processes gives rise to such a meaning in this case.

4.2. Obligatoriness of additive particle in mutually exclusive scenarios

In Scenario 4, it is part of the common ground that [◊ win (anand)].

(51) Utterance: φ = ◊ win (carlsen) ∧ ◊ win (anand)

Since, [◊ win (anand)] is already part of the common ground, the additive particle has to be used in B's utterance to avoid the implicated presupposition [Certain ¬[◊ win (anand)]] or [¬Certain [◊ win (anand)]], which are incompatible with the discourse.

The discussion in section (4) shows that when the complement clause of the existential modal holds an Additive Particle, the shape of the presupposition triggered is dependent on whether we are considering mutually exclusive alternatives or mutually compatible ones.

Summary

In this squib I have discussed cases of mutually exclusive pairs of alternatives placed under the possibility modal. In Bangla (and in other South Asian languages), we see an obligatory additive particle showing up in such cases. Based on the literature on obligatory presuppositions, I have tried to reason that not using the additive particle with one possibility excludes the possibility of the other. So in scenarios, when all the possibilities are open, the speaker is obligated to use the additive particle to signal this openness. The paper also shows that in such constructions we always get a wide scope of the additive.

The squib gives rise to important questions that must be addressed in future work. One needs to understand the predicate *par-* as opposed to other predicates expressing existential modality. It is important to include the concessive conditionals in the discussion for a fuller analysis of the phenomenon.

Ishani Guha
guha@mit.edu

References

Amsili, P. and C. Beyssade. 2009. Obligatory Presuppositions in discourse. Ms.

Chemla, E. and P. Schlenker 2012. Incremental vs. symmetric accounts of presupposition projection: an experimental approach. Natural Language Semantics 20: 177-226.

Bayer, J. and A. Lahiri. 1990. Bengali emphatic clitics in the lexicon-syntax interface. In W.U. Dressler et al. Eds, Trends in Linguistics, Mouton de Gruyter.

Fox, D and I. Heim. 2015. Lecture-notes on Pragmatics.

Fox, D and R. Katzir. 2011. On the characterization of alternatives. Natural Language Semantics 19: 87-107.

Heim, I. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9: 183-221.

Kripke, S. A. 2009. Presupposition and Anaphora: Remarks on the Formulation of the Projection Problem. Linguistic Inquiry 40: 367-386.

Lahiri, U. 1998. Focus and Negative Polarity in Hindi. Natural Language Semantics 6: 57-123.

Laka, I. 1990. 'Negation in syntax: on the nature of functional categories and projections', PhD Dissertation, MIT, Cambridge, MA.

Ramchand. 2014. Deriving variable linearization. Natural Language and Linguistic Theory 32: 263-282.

Simpson, A and S. Sourov. 2014. Finiteness, negation and the directionality of headedness in Bangla. Natural Language and Linguistic Theory 32: 231-261.

Sauerland, U. 2004. Scalar Implicatures in complex sentences. Linguistics and Philosophy 27.

Sauerland, U. 2008. Implicated Presuppositions. In Anita Steube (Ed.), *Sentence and Context*. Mouton de Gruyter.

Tonhauser, J, D. Beaver, C. Roberts and M. Simons. 2013. Toward a taxonomy of projective content. Language 89: 66–109.

Interaction between prosody and focus types: Evidence from Bangla and Hindi*

Arunima Choudhury & Elsi Kaiser

University of Southern California

1. Introduction

This paper reports a series of psycholinguistic experiments investigating the prosodic encoding of information structure in Bangla and Hindi. We tested whether speakers of these languages encode the information-structural distinction between new-information focus and corrective focus prosodically, and if so, what prosodic dimensions (e.g. fundamental frequency (*F0*, duration, intensity) are used. We also tested to what extent listeners are able to perceive prosodic cues to information structure.

The term 'information structure' is used to refer to parts of the sentence that can be classified as given and topical information or as new/focused information (e.g. Chafe 1976, Vallduví 1990, Lambrecht 1994, Féry & Krifka 2008). The category of focus is often classified into two broad types: (a) New Information Focus, and (b) Contrastive Focus. New-information focus is primarily associated with new, non-presupposed information, and a commonly used diagnostic for new-information focus is the use of wh-questions, as in (1a), from Zimmermann & Onea (2011). The focus constituent (marked with square brackets and the subscript F) is the part of the sentence that corresponds to the answer to the wh-question (e.g. Jackendoff 1972, Gussenhoven 2008, Kanerva 1990).

(1) a. A: What color did Peter paint his bicycle? B: He painted it $[\text{blue}]_F$.

Following Zimmermann & Onea (2011), we define new-information focus as follows: A focused constituent (e.g. *blue* in 1a) expresses new information if the focused element introduces new information to the Common Ground (the mutually shared knowledge between speaker and addressee), when alternatives to the focused element had not been explicitly mentioned in the prior discourse. Contrastive focus, on the other hand, occurs when one or more of the alternatives to the focused constituent have been mentioned. For example, in (1b), one speaker claims that Peter painted his bicycle red.

* We are grateful to the audience at FASAL 5 (Yale, 4/2015) for thoughtful comments and suggestions. We also thank Prof. Andrew Simpson for his valuable feedback and support at various stages of this project.

The other speaker corrects this and says that out of the possible colors that Peter could have painted this bicycle, he painted it blue. The contrastively-focused element *blue* competes with another, explicitly-mentioned member in the set of possible bicycle colors (*red*) for introduction in the Common Ground (see Zimmermann & Onea 2011).

(1) a. A: Peter painted his bicycle red. B: No, he painted it $[\text{blue}]_F$.

The type of focus in (1b) is more specifically known as corrective focus, which is sometimes regarded as a subtype of the more general class of contrastive focus, and sometimes as distinct from contrastive focus (for discussion, see e.g. Zimmermann & Onea 2011, Dik 1997, Gussenhoven 2007). In this paper, we investigate the prosodic encoding of new-information focus and corrective focus, since these two focus types are intuitively distinct and reliably elicited by wh-questions and correction contexts.

Despite examples like (1a,b), researchers disagree whether contrastive focus constitutes an information-structural category of its own, distinct from new-information focus. Some researchers – largely with a syntactic focus – argue that contrastive focus is a focus type independent of new-information focus (e.g. Chafe 1967, Halliday 1967, Rochemont 1986, Kiss 1998, Valduví & Vilkuna 1998, Molnár 2002). On the other hand, others like Bolinger (1961), Rooth (1992), Krifka (1993) and Schwarzschild (1999) – largely with a semantic focus – have claimed that contrastive focus is not fundamentally different from information focus, as far as its semantics are concerned.

Work on prosody has identified different intonational properties for new-information focus and contrastive/corrective focus. Zubizarreta (1998) believes new-information focus and contrastive focus to be information-structurally distinct and to have different intonational realizations in Romance languages like Italian and Spanish. Pierrehumbert (1980) and others working within auto-segmental metrical phonology posit distinct pitch accents in English for new-information focus (H*) and contrastive focus (L + H*).

Psycholinguistic work suggests that speakers encode different focus types with different prosodic cues and that listeners are sensitive to this, but the mapping between focus types and prosody may not be straightforward. For example, a comprehension study by Watson et al. (2008) on English showed that L+H* accents are interpreted as having a contrastive interpretation, whereas H* accents can mark either contrastive referents or new-information referents. A series of production-and-perception experiments by Breen et al. (2010) found that speakers distinguish corrective and non-corrective focus reliably only when they are aware of the prosodic ambiguity present across different information structures. In that case, speakers produced contrastively focused elements with greater intensity, longer duration, and (perhaps surprisingly) lower mean and maximum *F0* than non-contrastively-focused elements.

On the perception side, Breen et al. observed that even when speakers' productions distinguished corrective and non-corrective focus, listeners did not successfully identify focus type. (Listeners' performance on focus types improved when an attributive phrase "I heard that" preceded the critical SVO sentence, which Breen et al. attribute to speakers prosodically marking "I" when the sentences were contrastive.) Recent production work by Katz & Selkirk (2011) found that contrastive focus and new-information focus are prosodically distinct, but Katz and Selkirk did not test perception.

In sum, within theoretical linguistics there is an on-going debate about whether we

should have a grammatical representation of contrastive focus that is distinct from that of new-information focus. Current psycholinguistic work exploring the distinctness (or lack thereof) of new-information focus and corrective focus has led to mixed results, especially when we look at both production and perception.

In the current study, we investigate the production and perception of new-information and corrective focus with native Bangla and Hindi speakers, to see how two languages which are typologically related to each other but which differ from English in some key respects, can contribute to this debate. As will become apparent, looking at two related languages turns out to be highly informative with respect to the question of what prosodic/acoustic dimensions languages use for encoding focus.

1.1 Background: Bangla and Hindi

Bangla and Hindi are a part of the modern-day Indic/Indo-Aryan branch of the Indo-European family of languages. Their canonical word order is SOV, but word order is relatively flexible. Both Bangla and Hindi have a 'default focus position' which immediately precedes the verb (e.g. Choudhury 2010 on Bangla, Kidwai 2000 on Hindi), and this position has been suggested to be the default position for new information focus in both languages (S $[O]_{new}$ V). Furthermore, native speaker judgments indicate that contrastively-focused elements can also occur in this position (S $[O]_{corr}$ V). Furthermore, in canonical SOV word order, the subject can be in new-information focus ($[S]_{new}$ O V) or contrastively focused ($[S]_{corr}$ O V).[1] Hence, the canonical SOV word order in both Bangla and Hindi can have the following interpretations:

(i) S $[\mathbf{O}]_{\mathbf{new}}$V - New-information focus on **object**; subject is unfocused
(ii) S $[\mathbf{O}]_{\mathbf{corr}}$V - Corrective focus on **object;** subject is unfocused
(iii) $[\mathbf{S}]_{\mathbf{new}}$ O V - New-information focus on **subject**; object is unfocused
(iv) $[\mathbf{S}]_{\mathbf{corr}}$ O V - Corrective focus on **subject**; object is unfocused

Now, let us turn to what is known about the prosodic properties of Hindi and Bangla. Previous studies (Hayes & Lahiri 1991, Féry 2010, Patil et al 2008, Moore 1965, Harnsberger 1994, Khan 2007) have shown that both languages have an L^*H_P pitch accent on the focused constituent (based on the autosegmental metrical phonology framework, see Pierrehumbert 1980). In both languages, each content word forms its independent phonological phrase and each phonological phrase receives a low tone and a high phrase boundary associated with the right edge of the prosodic word. Both Bangla and Hindi clauses also exhibit a downstep intonation pattern. However, not much is known about the quantitative values of fundamental frequency, duration and intensity of the focused constituents in Hindi and Bangla. Also, we do not know much about the prosodic differences between focus types in Hindi and Bangla, and whether Bangla and

[1] Noncanonical OSV order does not seem to allow the sentence-initial argument to be in new-information focus (* $[O]_{new}$ S V) although contrastive focus seems to be possible ($[O]_{corr}$ S V). See Choudhury (2015) for further discussion and experiments regarding OSV order in Hindi and Bangla. Here, we focus on SOV order. It is also worth noting that Bangla and Hindi have focus particles (e.g. Bhatt 1994, Sharma 2003 on Hindi, Choudhury 2010 on Bangla), but they are only used in certain contexts, and only with corrective focus. We do not investigate them here. Our experimental stimuli do not require focus particles.

Hindi encode the focus types distinctly using prosody. Our work aims to contribute to these questions.

1.2 Research Questions

The first research question that our experiments address is whether *speakers* of Bangla and Hindi encode the distinction between new-information focus and corrective focus prosodically, and if so, what prosodic dimensions (e.g. fundamental frequency *F0*, duration, intensity) are used, and are they the same in both languages? The second research question is whether *listeners* of these two languages are able to perceive the prosodic cues used to signal focus types. Are listeners able to differentiate between new information focus and corrective focus, when word order provides no cues?

These two questions will provide new data on the encoding and perception of information structure in Bangla and Hindi, and they also relate to the debate regarding the relationship between new-information focus and contrastive/corrective focus. If speakers of Hindi and Bangla distinguish prosodically between the two focus types and listeners are sensitive to these prosodic cues, this will provide us with further crosslinguistic evidence that focus types are encoded as separate categories.

Conducting a parallel investigation of two closely-related languages will allow us to gain insights into how variable languages are in terms of the specific prosodic dimensions that they use to encode information structure. Several studies on Chinese languages suggest that even typologically closely related language may employ different prosodic cues to encode focus (Xu 1999 for Mandarin, Wu & Xu 2010 for Cantonese, Chen et al 2009 for Beijing Mandarin, Taiwanese Mandarin and Taiwanese). Thus, we should not assume an overly simplistic view regarding the 'division of labor' between different prosodic dimensions such as *F0*, duration and intensity.

We conducted three sets of experiments. All three experiments consist of a production study followed by a perception study, where the production study recordings were used as stimuli for the perception study. Experiment 1 and Experiment 2 were conducted with native Bangla speakers, and Experiment 3 was conducted with native Hindi speakers. Experiment 2 (Bangla) and Experiment 3 (Hindi) use sentences with Adverb-Subject-Object-Verb word order. This was done based on the results of Experiment 1 (Bangla), which used Subject-Object-Verb sentences. In Experiments 2 (Bangla) and 3 (Hindi), the subject is no longer at the sentence-initial position, but the object is still at the default focus position. This allows us to see whether the results of Experiment 1 could be due to specific prosodic properties of sentence-initial elements. For all three studies, we first describe the design and methods used in the production phase, followed by the perception study and the results of the perception study, and finally the acoustic analyses of the stimuli.

2. Experiment 1: SOV word order in Bangla

2.1 Production phase

In the production phase, we elicited Bangla SOV sentences with new-information or corrective focus on the subject or the object. The data from this study was analyzed and

used as the stimuli for the perception study phase of Experiment 1. Five adult native Bangla speakers (3 female, 2 male) participated in the production phase. All were originally from Kolkata, India, and had been in the U.S. for at most 3 years.

Participants produced SOV sentences in contexts designed to elicit either new-information focus or corrective focus. We used wh-questions to elicit new-information focus (following Gussenhoven 2008, Kanerva 1989 and others), and yes/no questions to elicit corrective focus (following Breen et al 2010), as in (2a)-(2d). We manipulated focus type (new-information focus vs. contrastive focus) and the grammatical role of the focused element (subject vs. object). This resulted in four conditions: Sub-New, Obj-New, Sub-Corr, Obj-Corr (Table 1). The word order of the answer is the same in all conditions (SOV). The verbs were in the simple past tense, and nouns were 2.15 syllables long on average. All subjects were [+human], and all objects were inanimate.

Table 1. Experiment 1: Production study design. (Underlining indicates focus)

	Subject	Object
New-information Focus	S O V	S O V
Contrastive Focus	S O V	S O V

(2a) Subject wh-question to elicit new-information focus (Sub-New information)
Q: ke gari kinlo? A: **[baba]**$_{\textbf{newinfo}}$ gari kinlo
Who car bought father car bought
Who bought a car? 'Father bought a car'

(2b) Object wh-question to elicit new-information focus (Obj-New information)
Q: baba ki kinlo? A: baba **[gari]**$_{\textbf{newinfo}}$ kinlo
Father what bought father car bought
'What did father buy?' 'Father bought a car'

(2c) Subject yes/no question to elicit corrective focus (Sub-Corrective)
Q: protibeshi gari kinlo ki? A: **[baba]**$_{\textbf{Corr}}$ gari kinlo
Neighbor car bought Q father car bought
Did neighbor buy a car? 'Father bought a car'

(2d) Object yes/no question to elicit corrective focus (Obj-Corrective)
Q: baba kompyutar kinlo ki? A: baba **[gari]**$_{\textbf{Corr}}$ kinlo
Father computer bought Q father car bought
Did father buy a computer? 'Father bought a car'

Participants saw a question-answer pair on the screen, and said the answer out loud. They were encouraged to speak naturally, as if in a conversation.[2] The responses were recorded. This study was run on a PC using Paradigm software (Perception Research Systems). We created 20 target items and 32 filler items. Each speaker produced four targets per condition – for a total of 16 targets – as well as 16 fillers. We did two rounds of recordings for each speaker (all 32 trials in an initial round, and then all 32 trials again), to minimize disfluencies. The recordings from the second round were used for all

[2] When answering the yes/no questions, participants were instructed to avoid saying 'no' explicitly.

subsequent analyses. Each speaker produced 16 targets out of the full 20-item target set, and 16 fillers out of the full 32-item filler set; this kept the study at a reasonable duration. The full target and filler sets become relevant for the design of the perception study. In the next section, we present the perception study that used these sound files as stimuli. In Section 2.3, we present the acoustic analyses for these sound files.

2.2 Perception phase

The perception phase tests whether Bangla speakers perceive a distinction between new-information focus and corrective focus in the sentences generated during the production phase, and to see if the grammatical role of the focused element impacts perception. Twelve adult native speakers of Bangla participated (all originally from India, all had been in the U.S. for less than five years). None of these participants took part in the production study. The materials for this study were the audio files from the production phase. This study had 20 targets and 32 fillers, in a Latin-Square design. Every list had an equal distribution of target sentences spoken by all five speakers in all four conditions. The study was run using Paradigm software (Perception Research Systems).

On critical trials, participants saw a wh-question and a yes/no question on the screen (Figure 1). The grammatical role focused by the questions matched the grammatical role focused in the sound files: When participants saw a subject wh-question and a subject yes/no question, they heard a sound file elicited by (a) a subject wh-question or by (b) a subject yes/no question. Conversely, when they saw an object wh-question and an object yes/no question, they heard a sound file elicited by (a) an object wh-question or by (b) an object yes/no question. This allows us to test whether listeners can distinguish new-information focus and contrastive focus. Left and right locations of question types were counterbalanced. (The questions were shown in Bangla script.) Participants' task was to choose whether the sound file is an answer to the wh-question or the yes/no corrective question – i.e., which question is most appropriate for the sentence that they heard.

Figure 1. Experiment 1: Schematic example (Sub-New condition)

Ke gari kinlo?	**Protibeshi gari kinlo ki?**
Who bought a car?	Did neighbor buy a car?
(Subject wh-question)	*(Subject yes/no question)*

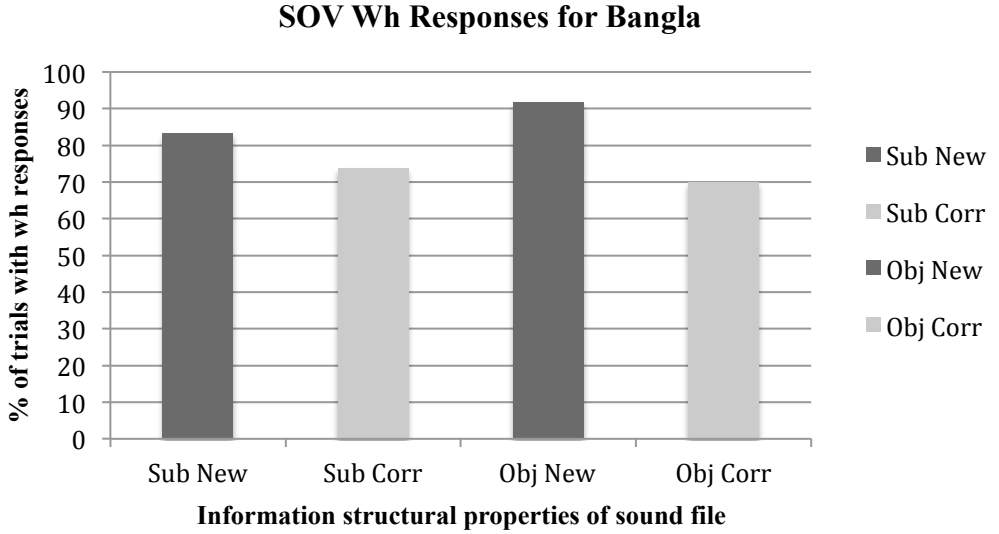
Presented auditorily (example of Sub-New condition)
[**baba**]$_{\text{NewInfo}}$ gaRi kinlo
father car bought

2.2.1 Results for perception phase

We present the results in terms of how often the participants chose the wh-question. Since wh-questions elicit new-information focus, this can be thought of as the *rate of new-information interpretations*. Figure 2 shows the rate of wh-question responses for each condition. Because participants only had two choices (wh-question, yes/no question), this means that whenever they did not choose a wh-question, they chose a yes-

no-question. The rate of yes/no-question responses can thus be inferred from Figure 2. (The wh-question is the 'right' answer for sounds elicited in new-information focus contexts, but the 'wrong' answer for sound files elicited in corrective focus contexts.)

Figure 2: Experiment 1: Rate of wh-question choices (indicating new-information focus)



As can be seen in Figure 2, the percentage of wh-question responses – i.e., the percentage of the time participants perceived the focus as being new-information focus – is well over 50% in all conditions. In other words, participants have an overall preference for a new-information interpretation (for discussion of why this might be the case, see Choudhury 2015). What is relevant for us here is that the preference for wh-questions is weakened for sound files elicited in a corrective context, when compared to sound files elicited in a new-information context. In object-focus conditions, the rate of wh-question choices is 92% for sound files elicited in a new-information focus context (i.e., preceded by a wh-question), but only 70% for sound files been elicited in a corrective focus context (i.e., preceded by a yes/no question). The same numerical pattern is observed in subject-focus conditions, but at a smaller magnitude: 83% of wh-question choices with new-information focus sound files; 73% wh-question choices with corrective focus sound files. Linear mixed-effects regressions confirm that the difference is significant in the case of object focus ($p<.05$) but not significant in the case of subject focus ($p=.14$).

We also tested for main effects and interactions here and in the other statistical analyses reported in this paper. However, due to space limitations, we mostly focus on the planned comparisons between the two different focus types on subjects and objects. Please see Choudhury (2015) for additional details and full results of all analyses.

Thus, our results indicate that Bangla listeners are able to distinguish corrective focus and new-information focus using only prosodic information when the focused element is the object, but not when it is the subject.

2.3. Acoustic analyses of production phase

We now turn to the prosodic acoustic properties of the focused constituents. We report on

two main prosodic cues, namely fundamental frequency (*F0* and duration). (We also looked at intensity, normalized over time, but found no significant effects of focus type, so we will not discuss the intensity data further. It seems that intensity does not play any role in encoding focus types in Bangla. Please see Choudhury 2015 for details.) The main purpose of the acoustic analyses is to determine which of these prosodic cues is facilitating the perception of the focus types, and to see if we can learn more about why listeners are not able to distinguish focus type differences at the subject position.

The production data was analyzed using Prosody Pro (Xu 2005-2012) and Praat (Boersma & Weenink 2009). We analyzed the time-normalized fundamental frequency and relative duration. For the time-normalized measures, we used Prosody Pro to divide each constituent into 10 equal time points/time segments. For the *F0* analyses reported throughout this paper, we focus on (the average of) the last 5 of these time points, because of the pitch accenting alignment of both Hindi and Bangla (Hayes & Lahiri 1991, Féry 2010, see Choudhury 2015 for additional discussion). We also analyzed the relative duration of the focused elements. To account for differences in speech rate, we followed Ito et al. (2006) and Kaland et al. (2011), and computed the relative duration of a constituent by dividing its absolute duration with the absolute duration of the sentence.

2.4 Results of production phase

Figure 3 shows the **fundamental frequency *(F0)*** patterns for the four conditions in Experiment 1, on SOV sentences in Bangla. As can be seen in the figure, it looks like the subject in all conditions ends with a high *F0* at the right edge of the word, regardless of whether or not it is focused. Objects, on the other hand, have high *F0*s at the right edge when they are focused and lower *F0*s when they are unfocused. Furthermore, we see that numerically, correctively focused objects (triangles) have a higher *F0* than new-information focused objects (diamonds). Statistical analyses (two-tailed paired t-tests) confirm that *F0*'s are significantly higher for corrective objects than new-information objects ($t(4)=4.019$; $P<0.05$), but corrective subjects vs. new-information subjects show no significant effect of focus type ($t(4)= -1.849$; $P=0.138$).

Figure 4 shows the **mean relative duration** of the focused constituent (subject or object) in all four conditions in Experiment 1. Correctively-focused constituents appear to be considerably longer than constituents in new-information focus. Statistical analyses (two-tailed paired t-tests) confirm that the mean durations are significantly higher for elements in corrective focus than for elements in new-information focus (subjects: $t(4)=-4.958$ $p<0.05$; objects: $t(4)=-7.217$ $p<0.01$). In sum, the duration analyses show that correctively-focused constituents are indeed significantly longer than constituents in new-information focus.

2.5 Discussion of Experiment 1 (Bangla SOV sentences)

Experiment 1 consisted of a production and perception study looking at SOV sentences in Bangla to test whether and with what prosodic dimension speakers encode the difference between new-information focus and corrective focus on subjects and objects. The acoustic analyses show that *F0* in Bangla encodes a difference between focus types on objects but not on subjects, whereas duration encodes a difference between focus types

on both subjects and objects. (Intensity does appear to reflect focus types.) However, in the perception phase of this study (which used the stimuli from the production phase as its stimuli), listeners were able to distinguish the two focus types reliably on objects but not on subjects. This is in line with the acoustic analyses for *F0*, but seems surprising in light of the duration data.

Before considering the implications of these findings further, it is important to mention a possible complication in Experiment 1: Could the asymmetrical *F0* results for subjects and objects be an artifact of the sentence-initial position of the subject, which is known to be associated with prosodic prominence. Could it be that this was masking potential effects of focus type on the subject? Experiment 2 addresses this issue, by testing SOV sentences that have an adverb in sentence-initial position.

Figure 3. Experiment 1: Time-normalized F0 contours of the SOV target sentences

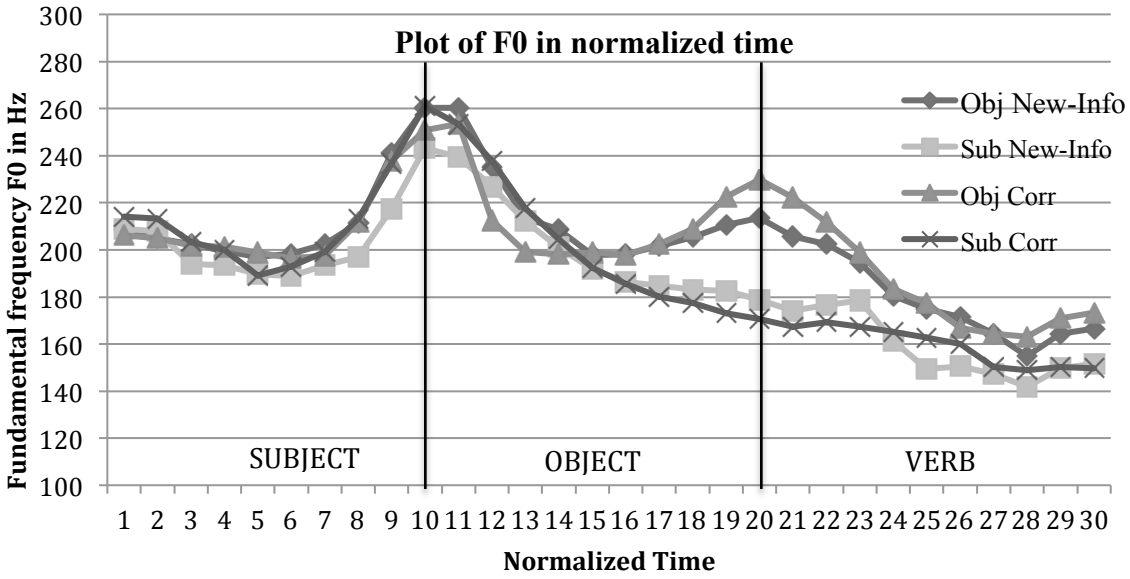
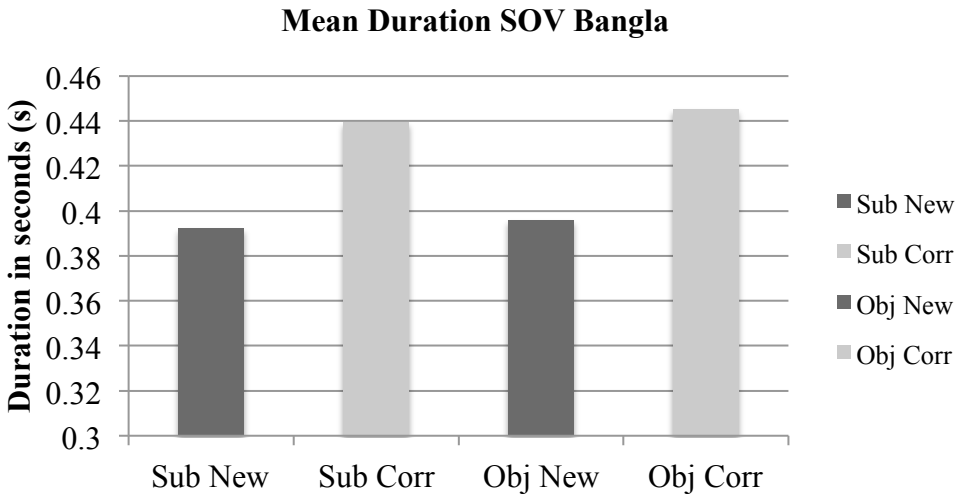


Figure 4: Experiment 1: Mean relative duration of the focused constituent (in seconds)



3. Experiment 2: Adv-SOV word order in Bangla

Experiment 2 tests if the absence of focus type effects on subjects exists even if the subject is not in sentence-initial position. Bangla is known to have a down-step pitch pattern (Hayes & Lahiri 1999, Féry 2010). This raises the question of whether the sentence-initial prominence could be masking potential differences between new-information focus and contrastive focus on the subject. Experiment 2 uses sentences where the subject is no longer in sentence-initial position: We added a sentence-initial adverb in front of the subject, as in example (3).

Similar to Experiment 1, this study is divided into two parts: a production study, followed by a perception study. The sentences generated during the production study were used as stimuli for the perception study.

(3a) Subject wh-question to elicit new-information focus (Sub-New)
Q: gotokal ke gari kinlo? A: gotokal [baba]$_{newinfo}$ gari kinlo
Yesterday who car bought yesterday father car bought
'Who bought a car yesterday? 'Father bought a car yesterday'

(3b) Object wh-question to elicit new-information focus (Obj-New)
Q: gotokal baba ki kinlo? A: gotokal baba [gari]$_{newinfo}$ kinlo
Yesterday father what bought yesterday father car bought
'What did father buy yesterday?' 'Father bought a car yesterday'

(3c) Subject yes/no question to elicit corrective focus (Sub-Corr)
Q: gotokal protibeshi gari kinlo ki? A: gotokal [baba]$_{Corr}$ gari kinlo
Yesterday neighbor car bought Q yesterday father car bought
Did neighbor buy a car yesterday? 'Father bought a car yesterday'

(3d) Object yes/no question to elicit corrective focus (Obj-Corr)
Q: gotokal baba kompyutar kinlo ki? A: gotokal baba [gari]$_{Corr}$ kinlo
Yesterday father computer bought Q yesterday father car bought
Did father buy a computer yesterday? 'Yesterday father bought a car'

3.1 Production phase

Similar to Experiment 1, the data from the production phase was acoustically analyzed and used as stimuli for the perception study. Five native Bangla speakers (3 female, 2 male; all originally from West Bengal, India, allhad been in the U.S. for at most 5 years) participated in the production phase. None of the participants participated in the previous experiment. The design was the same as in the production phase of Experiment 1, with four conditions (Sub-New, Obj-New, Sub-Corr, Obj-Corr). We used the same 16 targets as in Experiment 1. However, now participants produced Adv-SOV sentences instead of SOV sentences on target trials. The SOV part of the sentences was the same as in Experiment 1, and we now added a sentence-initial adverb (ex.3). The adverbs were all single-word expressions in Bangla and were of three types: time adverbs (e.g., *yesterday, last_night*), manner adverbs (e.g., *immediately, quickly*), and place adverbs (e.g., *inside, there*). All of these can naturally and felicitously occur in sentence-initial position in

Bangla. A total of ten common adverbs were used, repeated twice, of which four were time adverbs, four manner adverbs, and three place adverbs. The adverb was also mentioned in the question, as shown in ex.(3), to maximize naturalness. The adverbs were single words and 2-3 syllables in length. The study also included 32 fillers, which were similar to those in Experiment 1 but some had adverbs in different positions – sentence-medial and sentence-initial – because adverbs were also added to the targets. The production phase was conducted in the same was as in Experiment 1; participants saw question-answer pairs and were instructed to say the answer aloud, and the second round of recordings was used for all analyses.

3.2 Perception phase

In this phase, similar to Experiment 1, we test whether participants can perceive differences between subjects and objects in new-information focus vs. Contrastive focus. The design of this perception study was the same as the perception study in Experiment 1. Twenty adult native speakers of Bangla from India participated (all had been in the U.S. for less than 5 years). None of them took part in the production phase of this experiment or in the perception phase of Experiment 1, but four of the participants had participated in the production phase of Experiment 1. This was primarily due to the lack of native Bangla speakers in and around University of Southern California. However, an average of 24 months passed between the time when these four participants completed the production phase of Experiment 1 and the perception phase of Experiment 2. Thus, we do not expect their participation to distort the data.

As in Experiment 1, the sound files from the production phase were used as the stimuli for the perception phase. There were 20 targets and 32 fillers, in a Latin-Square design. The lists were created such that every list had an equal distribution of target sentences spoken by all five speakers in all four conditions. The procedure was the same as the perception phase of Experiment 1.

3.2.1 Results for perception phase

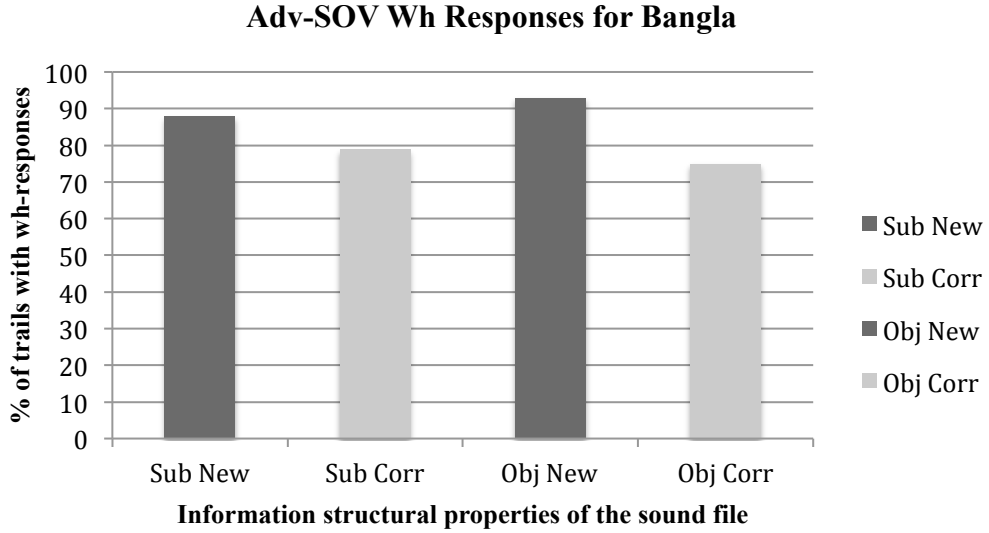
Similar to Experiment 1, we plot the results in terms of the percentage of 'wh-responses', shown in Figure 5. The rate of wh-responses can thus be thought of as the rate of new-information focus interpretations.

As a whole, Experiment 2 replicates the outcomes of Experiment 1. In object-focus conditions, the rate of wh-question choices is 93% for sound files elicited in a new-information focus context (wh-question), but only 75% for sound files been elicited in a corrective focus context (yes/no question). Similarly, in subject-focus conditions, the rate of wh-question choices for new-information focus sound files is 88%, and 79% for wh-question choices with corrective focus sound files. In other words, the overall wh-preference is again weakened for sound files elicited in a corrective context.

Linear mixed-effects regressions confirm this asymmetry: The difference in the rate of wh-question responses is significant in the case of object focus ($p<.05$) but not significant in the case of subject focus ($p=.11$). These results are consistent with the perception phase of Experiment 1. Thus, even with an adverb at the sentence-initial position, Bangla speakers are unable to perceive the prosodic difference between focus

types on subjects, although they can do so on objects. This suggests that the difference between subjects and objects is not an artifact of the subject's sentence-initial position.

Figure 5. Experiment 2: Rate of wh-question choices (indicating new-information focus)



Because the focus-types results for subjects are approaching marginal significance (commonly defined as $0.1 \leq p > 0.05$), we wanted to check whether we might be overlooking a meaningful result due to lack of statistical power. To test this, we decided to conduct a combined analysis on Experiments 1 and 2. However, before combining the perception data of Experiments 1 and 2, we first compared the experiments with repeated measures ANOVA with 'experiment' as a between-subjects factor, and found no significant effects of experiment ($p=0.419$). We then combined the perception data from Experiments 1 and 2 to see if effects that are almost marginal in the individual studies would reach significance with more data. However, mixed-effects logistic regression analyses yield the same results as we obtained for each experiment individually: With focused objects, we find significantly more wh-question choices when the object had been elicited with new-information focus than with corrective focus ($p<0.05$). However, for focused subjects there is no significant difference in the rate of selection of wh questions ($p=0.931$). These results corroborate our conclusion that native Bangla listeners can distinguish focus type on the (immediately preverbal) object, but not on the subject.

3.3. Acoustic analyses of production phase

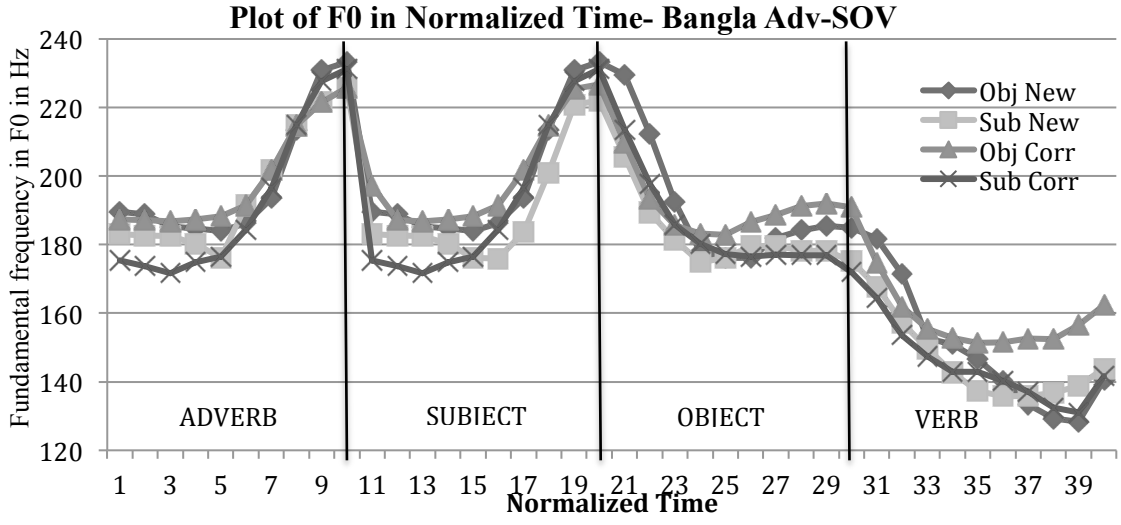
The acoustic data was analyzed in the same way as in Experiment 1, except that because the sentences contained an adverb, for the time-normalized analyses we now used 40 segments (segments 1-10: adverb, segments 11-20: subject, segments 21-30: object, segments 31-40: verb). (Similar to Experiment 1, we also looked at intensity, normalized over time, but found no significant effects of focus type, so we will not discuss the intensity data further in this paper.)

3.4 Results of production phase

Figure 6 shows the **fundamental frequency (*F0)*** patterns for Experiment 2, on Adverb-SOV sentences in Bangla. As can be seen in the figure, it looks like even with the presence of a sentence initial adverb, the subject still has a prosodic prominence, which may be related to prosodic resetting.[3] However, similar to SOV sentences in Experiment 1, we see that numerically, correctively focused objects (triangles) have a higher *F0* than new-information focused objects (diamonds). Statistical analyses (two-tailed paired t-tests) confirm that focused objects have significantly higher *F0* when they are contrastively focused than when they are in new-information focus ($t(4)=-4.850$; $P<0.05$), but no such difference is found on focused subjects ($t(4)= -1.946$; $P=0.124$). In sum, even when the subject is not at the sentence-initial position, the prosodic distinction between new-information focus and contrastive focus that we see on objects is still not reliably present on subjects.

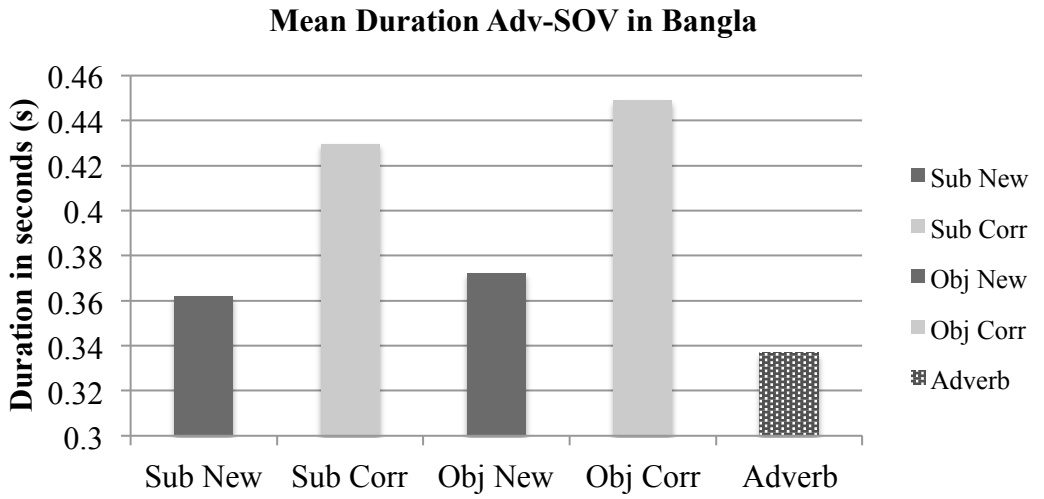
Figure 7 shows the **mean relative duration** of the focused constituent (subject or object) in all four conditions in Experiment 2. Similar to SOV sentences in Experiment 1, correctively-focused constituents appear to be considerably longer than constituents in new-information focus. Statistical analyses (two-tailed paired t-tests) confirm that the mean durations are significantly higher for elements in corrective focus than for elements in new-information focus (subjects: $t(4)=-3.111$, $p<0.05$; objects: ($t(4)=-3.420$ $p<0.05$). Thus, correctively-focused constituents are indeed significantly longer than constituents in new-information focus, regardless of grammatical role.

Figure 6. Experiment 2: Time-normalized F0 contours of the Adv-SOV target sentences



[3] A question that comes up regarding Figure 6 concerns the potential occurrence of an *F0* reset after the adverb and before the subject. It may be that the SOV part of the sentence forms a separate phonological phrase. This would explain why the subject (the first constituent of the hypothesized phonological phrase) has high *F0* and the downstep begins after the subject. We leave this as a question for future work.

Figure 7: Experiment 2: Mean relative duration of the focused constituent (and adverb)



3.5 Discussion of Experiment 2 (Bangla Adv-SOV)

Experiment 2 addressed a question left open by Experiment 1, namely whether the lack of significant focus type effects on the F0 measurements of subjects could be due to the sentence-initial position of the subject in Experiment 1. To test this, in Experiment 2 we used adverb-initial sentences. The results show that even when the subject is no longer sentence-initial, (i) Bangla speakers use *F0* to encode a difference between focus types on objects but *not* on subjects, and (ii) listeners are able to distinguish the two focus types reliably on objects but *not* on subjects. In sum, the results are very much in line with what we found in Experiment 1. Furthermore, it is also important to point out that in both Experiments 1 and 2, we find effects of focus type on the relative duration of both subjects and objects (correctively-focused elements are longer than elements in new-information focus). In light of this, it is intriguing that Bangla listeners do not seem to be sensitive to these duration cues. We return to this in the General Discussion section.

4. Experiment 3: Adv-SOV word order in Hindi

We conducted a study parallel to Experiment 2 in Hindi. As we will see, even two closely related languages do not pattern alike in terms of the prosodic encoding of focus types.

4.1 Production phase

The production phase was parallel to Experiments 1 and 2, except that the study was now conducted in Hindi. Five adult native Hindi speakers (3 female, 2 male; all living in Delhi, India at the time of testing and had lived there for most of their lives) participated. The participants did not speak any other Indian language apart from Hindi. The design was the same as in Experiments 1 and 2, with four conditions (Sub New, Obj New, Sub Corr, Obj Corr). Also, similar to Experiment 2, sentence-initial adverbs were used (4).

The design, methods and procedure were the same as Experiment 2, except for the language of the stimuli.

(4a) Subject wh-question to elicit new-information focus (Sub-New)
Q: parso kisne gari kharidi? A: parso [bhaiyya-ne]$_{newinfo}$ gari kharidi
day-before-yesterday who car bought day-before-yesterday brother car bought
'Who bought car day before yesterday? 'Brother bought car day before ystday'

(4b) Object wh-question to elicit new-information focus (Obj-New)
Q: pasro bhaiyya-ne kya kharida? A: parso bhaiyya-ne [gari]$_{newinfo}$ kharidi
day-before-yesterday brother what bought day-before-yesterday brother car bought
'What did brother buy day before yesterday? 'Brother bought car day before ystday'

(4c) Subject yes/no question to elicit corrective focus (Sub-Corr)
Q: parso papa-ne gari kharida kya? A: parso [bhaiyya-ne]$_{Corr}$ gari kharidi
day-before-yesterday father car bought Q day-before-yesterday brother car bought
'Did father buy car day before yesterday? 'Brother bought a car day before ystday'

(4d) Object yes/no question to elicit corrective focus (Obj-Corr)
Q: parso bhaiyya-ne skutar kharida kya? A: parso bhaiyya-ne [gari]$_{Corr}$ kharidi
day-before-yesterday brother scooter bought Q day-before-yesterday brother car bought
'Did brother buy scooter day before ystday?' 'Day before ystday brother bought car'

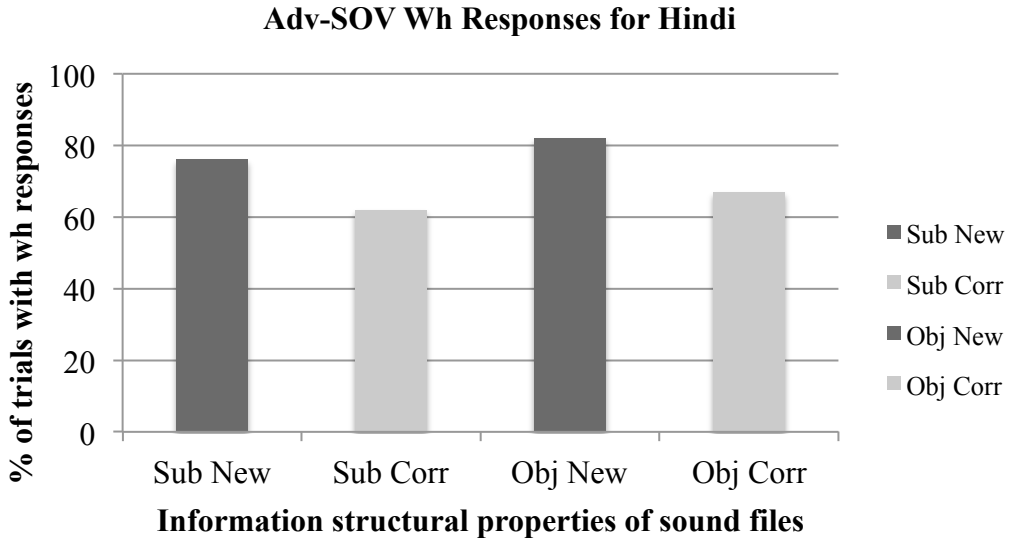
4.2 Perception Phase

The perception phase was again parallel to Experiment 2. Twenty adult native speakers of Hindi (all living in Delhi at the time testing, and had lived there for most of their lives) participated in the study. The participants did not speak any other Indian language. None of these participants took part in the production phase of this experiment. The design and procedure was he same as the perception phase of Experiment 2, except that the stimuli were now in Hindi.

4.2.1 Results for perception phase

We again present the results in terms of the percentage of wh-responses, i.e., the percentage of trials where the participants chose a wh-question (indicative of a new-information focus interpretation). In object-focus conditions, the rate of wh-question choices is 82% for sound files elicited in a new-information focus context (preceded by a wh-question), but only 67% for sound files elicited in a corrective focus context (preceded by a yes/no question). Similarly, in subject-focus conditions, the rate of wh-question choices for new-information focus sound files is 76%, and 62% for wh-question choices with corrective focus sound files. Linear mixed-effects regressions confirm that there is a main effect of focus type ($p<.05$), no effect of grammatical role ($p=.2$) and no interaction ($p=.76$). Thus, participants are able to perceive the difference between the two focus types equally well for subjects and objects. This is different from Bangla, where the participants were only able to distinguish the focus types for objects, and not subjects.

Figure 8. Experiment 3: Rate of wh-question choices (indicating new-information focus)



4.3. Acoustic analyses of production phase

For the acoustic analyses, the data was analyzed as in Experiment 2. (Similar to Experiments 1 and 2, we also looked at intensity, normalized over time, but found no significant effects of focus type, so we do not discuss the intensity data further here.)

4.4 Results of production phase

Figure 9 shows the **fundamental frequency (*F0)*** patterns for each of the four conditions in Experiment 3. We clearly see a general down-step pattern after the subject. However, the adverbs (segments 1-10) and subjects (segments 10-20) both have almost equally high (peak) *F0*'s. The *F0* starts to gradually fall after subjects such that the objects (segments 21-30) have a much lower (peak) *F0* than the subjects followed by the verbs (segments 31-40), which have the lowest *F0*. Statistical analyses show that there is no effect of focus type on either subjects or objects. Thus, unlike Bangla, where we found an effect of focus type on *F0*s for the object, in Hindi we do not find any *F0* differences in the focus types for either subjects or objects.

Figure 10 shows the **mean relative duration** of the focused constituent (subject or object) in all four conditions in Hindi. Similar to Bangla in Experiments 1 and 2, correctively-focused constituents appear to be considerably longer than constituents in new-information focus. Statistical analyses (two-tailed paired t-tests) confirm that the mean durations are significantly longer for elements in corrective focus than for elements in new-information focus (subjects: $t(4)=-3.679$ $p<0.05$; objects: $t(4)=-3.302$ $p<0.05$). Thus, contrastively-focused subjects and objects are produced with reliably longer duration than subjects and objects in new-information focus in Hindi. Thus, similar to Experiments 1 and 2 in Bangla, Hindi speakers encode the difference between the focus types using duration, and these duration cues are equally strong for subjects and objects.

Figure 9: Experiment 3: Time-normalized F0 contours of the Adv-SOV target sentences

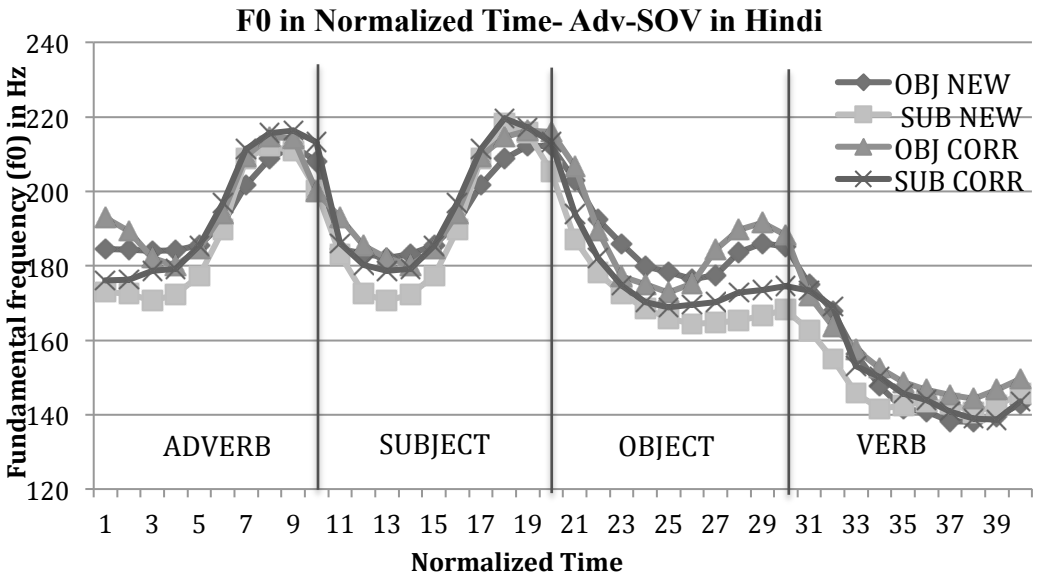
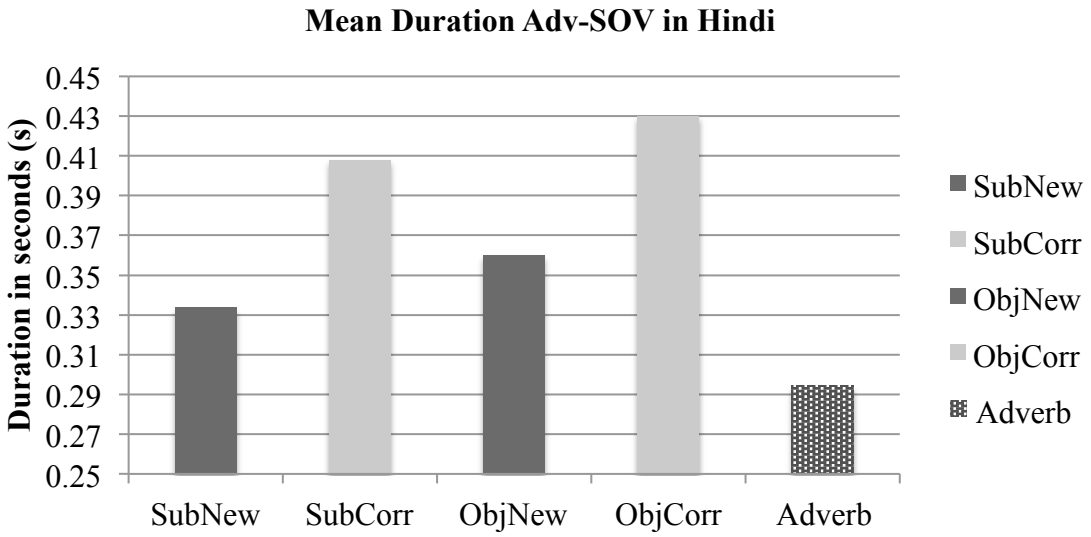


Figure 10: Experiment 3: Mean relative duration of the focused constituent (and adverb)



4.5 Discussion of Experiment 3 (Hindi Adv-SOV)

The acoustic analyses for Hindi indicate that speakers (i) are not using *F0* to encode focus types, but (ii) are using duration, because words in corrective focus are significantly longer than words in new-information focus. The results of the perception study show that Hindi listeners can reliably differentiate between the two focus types on both subjects and objects. In light of the acoustic analyses, this leads us to conclude that Hindi listeners are sensitive to duration cues but not to *F0* cues.

5. General Discussion

We conducted three production-perception studies to investigate whether speakers of Bangla and Hindi encode the distinction between new-information focus and corrective focus prosodically in sentences with canonical SOV order, and if so, what prosodic dimensions (e.g. *F0*, duration, intensity) are used, whether they are the same in both languages. We also tested whether listeners of these two languages are able to perceive the prosodic cues used to signal focus types (in the absence of any word order cues). These studies provide new empirical information about Hindi and Bangla and also contribute to the longstanding debate regarding the status of the focus types, i.e. whether new-information and corrective focus should be regarded as distinct notions, or as two subtypes that do not differ in their basic semantics.

Our results for Bangla show that Bangla speakers use *F0* to encode the distinction between new-information focus and corrective focus on objects but not on subjects. Interestingly, however, Bangla speakers use duration to distinguish focus types on both subjects and objects. Hindi speakers, on the other hand, do not seem to be using *F0* at all to encode distinctions in focus type on either subjects or objects. Like Bangla speakers, however, they produce correctively-focused elements with reliably longer duration than elements in new-information focus.

The finding that speakers in these two languages reliably distinguish new-information focus and corrective focus provides support for the idea that these two focus types are distinct categories. As discussed in section 1, there has been a long-standing debate regarding whether different focus types are information-structurally distinct. On the one hand, linguists like Kiss (1998), Vallduví &Vilkuna (1998), Hartmann & Zimmermann (2006) and Zimmermann & Onea (2011), distinguish between contrastive and new-information focus. On the other hand, others like Rooth (1992), Schwarzschild (1999) and Krifka (1993) argue that contrastive/corrective focus is not fundamentally different from information focus, as far as its underlying semantics are concerned. We find that speakers of both Bangla and Hindi produce elements in new-information focus and in contrastive focus with reliable prosodic differences, which is compatible with the idea that these two focus types are information-structurally distinct.

What about the perception side? Duration provides a reliable cue for focus types in both Bangla and Hindi on both subjects and objects, so it seems reasonable to expect listeners to be able to distinguish new-information focus from corrective focus on both subjects and objects in both languages. However, as we saw, whereas Hindi listeners are indeed able to do so, Bangla listeners are only able to distinguish focus types reliably when the object is in focus, not when the subject is in focus. (The data for the subject are going in the right direction, but do not reach significance. Thus, even if we were to say, optimistically, that Bangla listeners are somewhat able to distinguish focus types on the subject, they are *much* better at doing so when the focused element is the object.)

When we combine the production and perception data, we can identify some intriguing asymmetries between Hindi and Bangla. Although both languages encode focus type distinctions using duration, it appears that Hindi listeners rely on duration cues during perception more than Bangla listeners. Bangla listeners, in contrast, seem to rely more on *F0* cues (and are thus not able to reliably detect differences in focus type on the subject). Thus, our work provides the first crosslinguistic psycholinguistic

evidence that that Bangla and Hindi, two closely-related modern Indo-Aryan languages which share a lot of common syntactic properties, differ somewhat in the specifics of how focus types are encoded by speakers, but differ more strikingly in how sensitive listeners are to different kinds of prosodic cues. The deeper reasons for these differences are an intriguing direction for future work.

References

Boersma, Paul, and David Weenik. 2009. Praat: Doing phonetics by computer, www.praat.org

Bolinger, Dwight. 1961. Contrastive Accent and Contrastive stress. *Language* 37, 83–96.

Breen Mara, Evelina Fedorenko, Michael Wagner, and Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes*, 25, 7, 1044-1098

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. *Subject and Topic*, ed. Charles. Li. London, New York: Press. 25-56.

Chen, Ying., Yi Xu, and Susan Guion-Anderson, S. 2014. Prosodic realization of focus in bilingual production of Southern Min and Mandarin. *Phonetica* 71, 249-270.

Choudhury, Arunima. 2015. Interaction between Prosody and Information structure: Experimental evidence from Hindi and Bangla, Doctoral Dissertation. USC

Choudhury, Arunima. 2010. A Study of Focus in Bangla, M.A. Paper, USC, Los Angeles

Dik, Simon. 1997. The Theory of Functional Grammar. *Part 1: The Structure of the Clause.* Kees Hengeveld (ed.) Berlin, New York: Mouton de Gruyter.

Féry, Caroline, and Manfred Krifka. 2008. Information structure: Notional distinctions, ways of expression. In *Unity and diversity of languages*, 123-136. P. van Sterkenburg (eds.) Amsterdam: John Benjamins.

Féry, Caroline. 2010. The intonation of Indian languages: an areal phenomenon. *Problematizing Language Studies*. (ed.) Imtiaz Hasnain and Shreesh Chaudhury. Festschrift for Rama Agnihotri. Delhi: Aakar Books, 288-312.

Gussenhoven, Carlos. 2007. Types of focus in English. *Topic and Focus*: *Cross-linguistic Perspectives on Meaning and Intonation* (eds.) Chungmin Lee, Matthew Gordon, and Daniel Büring, 83–101.

Gussenhoven, Carlos. 2008. Notions and subnotions in information structure. *Acta Linguistica Hungarica* 55, 381-395.

Harnsberger, James D. 1994. Towards an intonational phonology of Hindi. Ms., University of Florida.

Hartmann, Katharina, and Matle Zimmermann. 2006. *Interdisciplinary Studies on Information Structure.* (eds.) Ishihara, S., M. Schmitz, and A. Schwarz. 5, 61-105

Halliday, Michael. 1967. *Intonation and grammar in British English*. The Hague: Mouton.

Hayes, Bruce, and Aditi Lahiri. 1991. Bengali Intonational Phonology. Natural Language and Linguistic Theory 9, 7-96

Ito, A., Nagasawa, T., Ogasawara, H., Suzuki, M., & Makino, S. (2006). Automatic Detection of English Mispronunciation Using Adaptation and Automatic Assessment of English Intonation and Speaker Rhythm. Educational Technology Research, 29:13-23. Japan Society of Educational Technology

Jackendoff, Ray S. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Kanerva, Jonni M. 1990. *Focus and Phrasing in Chichewa Phonology*. NY: Garland.

Katz, Jonah, and Elisabeth Selkirk. 2011. Contrastive focus vs. discourse-new: evidence from prosodic prominence in English. *Language*, 87(4), 771-816.

Khan, Sameer. D. 2008. The Intonational Phonology and focus prosody of Bengali. Doctoral Dissertation. UCLA, Los Angeles.

Kidwai, Ayesha. 2000. *XP-Adjunction in Universal Grammar: Scrambling and Binding in Hindi-Urdu.* Oxford: Oxford University Press.

Kiss, Katalin É. 1998. Identificational Focus versus Information Focus. *Language* 74: 245–273.

Krifka, Manfred. 1993. Focus and presupposition in dynamic interpretation. *Journal of Semantics* 10, 269-300.

Molnár, Valéria. 2002. Information structure in a cross-linguistic perspective. *Language and computers.* H. Hasselgård, S. Johansson, B. Behrens, & C. Fabricius-Hansen (eds.), 39, 147-161. Amsterdam, The Netherlands: Rodopi.

Moore, Robert R. 1965. A study of Hindi intonation. Doctoral Dissertation, University of Michigan.

Patil, Umesh, Gerrit Kentner, Anja Gollard, Frank Kugler, Caroline Féry, and Shravan Vasishth. 2008. Focus, word order and intonation in Hindi. *Journal of South Asian Linguistics* 1. 53-70

Pierrehumbert, Janet. 1980 The phonology and phonetics of English intonation. Doctoral Dissertation, MIT. Distributed 1988, Indiana University Linguistics Club.

Rochemont, M. S. (1986). Focus in generative grammar. Amsterdam, Philadelphia: John Benjamins.

Rooth, Mats. 1992. A Theory of Focus Interpretations. *Natural Language Semantics,* 1, 75–116.

Schwarzschild, Roger. 1999. Givenness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics*, 7, 144-177.

Vallduví, Enric and Maria Vilkuna. 1998. On Rheme and Kontrast. *In The Limits of Syntax,* (eds.) Peter Culicover & Louise McNally. New York: Press, 79–106.

Watson, Duane G., Michael K Tanenhaus, and Christine A Gunlogson. 2008. Interpreting Pitch Accents in Online Comprehension: H* vs. L+H*. *Cognitive Science* 32, 1232–1244

Wu, Wing Li and Yi Xu. 2010. Prosodic Focus in Hong Kong Cantonese without Post-focus Compression. In *Proceedings of Speech Prosody 2010*, Chicago

Xu, Yi. 2005-2012. ProsodyPro.praat: http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27. 55-105

Zimmermann, Malte and Edgar Onea. 2011. Focus Marking and Focus Interpretation *Lingua* 121: 1651-1670

Zubizarreta, Maria Louisa. 1998. *Prosody, Focus, and Word Order*, Linguistic Inquiry Monograph 33, MIT Press.

Arunima Choudhury & Elsi Kaiser
arunimac@usc.edu & emkaiser@usc.edu

Verb Agreement in Hindi and its Acquisition[1]

Benu Pareek, Ayesha Kidwai & Sonja Eisenbeiss

Jawaharlal Nehru University, University of Essex

1. Introduction

The paper explores the theoretical implications of an elicited production study of acquisition of Case marking in Hindi spoken in the Delhi region.[2] The results of this study indicate that while children as young as 3 years and 5 months[3] make no errors in verb agreement with an ergative case-marked subject, verb agreement errors in the context of nominative subjects yielded surprising results. We found that these errors may be split, with the lower auxiliary verbs showing agreement with the case-marked object and the higher with the nominative subjects. We describe this children's divergent use both in terms of adherence to a universal condition and partial or incomplete adherence to a language specific condition whereby agreement with an overtly case-marked nominal is not allowed on the verbal constituents.

The paper is organized as follows: Section 2 provides a brief description of the psycholinguistic experiment and the methodology employed for the study of the acquisition of Hindi case and agreement by children growing up in standard Hindi speaking households, followed by the findings for the use of case by children. Section 3 discusses the verb/auxiliary phi-feature agreement phenomenon as seen in the adult grammar of Hindi, and the combination of conditions of transitivity and temporal aspect that govern it. Section 4 discusses in some detail the findings for verb agreement in the experiment based on observed patterns in children's utterances. In Section 5, we suggest an analysis to account for the weaknesses in the developing grammar keeping the structural implications of the adult grammar as background, and in Section 6, we conclude the paper.

[1] We thank the anonymous reviewer for comments that helped strengthen the arguments in the paper. We also thank Raj Laxmi Singh for her insights on the complexities of aspectual/verbal structure for understanding the children's developing grammar.

[2] The study was funded by a JNU/Essex Development Fund grant to study 'The Acquisition of Hindi Case Marking' conducted in Delhi-NCR (2013)

[3] Henceforth age is represented as Y;MM. For example, 3 years 5 months as 3;5.

2. The Study: Acquisition of Hindi Case Marking[4]

The study consisted of an elicited production task with pairs of pictures for 15 (+2 practice items) verbs, with the intention of eliciting full definite noun phrases, with nominative, ergative, accusative/dative, instrumental and genitive cases. Conducted in the Delhi region with 21 children aged between 3;5 and 5;11 (10 boys and 11 girls), the study comprised of pictures for Transitive and Ditransitive predicates, with easily recognizable human characters and inanimate items for Subjects and Objects. All the children in the study were from Hindi speaking households where the primary language spoken by parents/caregivers was Hindi.

(1) *List of Verbs for the study*

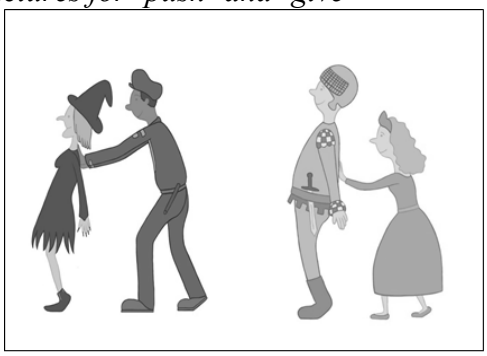
		Simple Predicates	**Complex Predicates**
V_t	**Nom-Acc**	1. *khiiNc* (pull) 2. *nehlaa* (bathe)	1. *dhanyavaad keh* (thank) 2. bye *kar* (wave) 3. *dhakkaa de* (push) 4. *gale lagaa* (hug) 5. kiss *kar* (kiss) 6. *gudgudi kar* (tickle)
V_t	**Nom-Gen**	-	1. *madad kar* (help) 2. *piichaa kar* (follow)
V_t	**Nom-Soc**	-	1. *haath milaa* (shake-hand)
V_{dt}	**Nom-Dat-Acc(Null)**	1. *likh* (write) 2. *dikhaa* (show) 3. *de* (give) 4. *bhej* (send)	

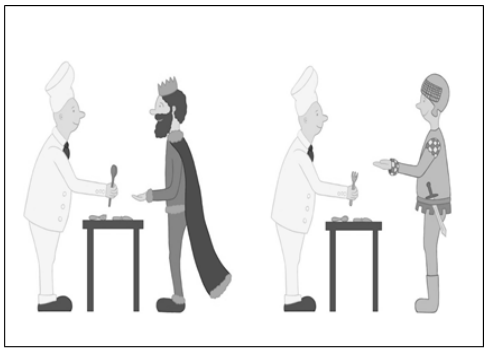
The table above lists the verbs included, alongside the predicate structures for the respective verbs. For each verb, two pictures with the same action but different characters were shown to the children and they were encouraged to talk about the events in both the pictures in complete sentences, using full definite noun phrases. The use of different characters was necessary to avoid argument omissions in the children's responses. Each of these target sentences were elicited in two rounds: in one round the children described the ongoing event in the progressive aspect; the second round comprised of the children talking to a puppet, telling him/her about what had earlier happened in the pictures. If children were unable to respond in the required complete sentence, which could have been due to incomprehension of the event in the picture, or a novel complex predicate, cues were provided to prompt the appropriate response. Much care was taken that these cues did not prime the case forms that the study aimed to elicit by omitting the arguments and or verb forms as they would appear in the context[5]. Below are pictures from the task for the verbs 'push' and 'give'.

[4] This study was part of a cross linguistic research, the design/tools for which were based on the work done in COST Action IS0804 "Language Impairment in a Multilingual Society: Linguistic Patterns and the road to Assessment" (www.bi-sli.org)

[5] For instance, the following was the cue provided to the child for pictures of the verb 'push':
ye tasviir dhakkaa dene ke baare me hE

(2) *Pictures for 'push' and 'give'*





Since all the images used for the task were of an ongoing event, as can be seen in pictures for 'push' and 'give' in (2), children naturally tended to describe the event in the progressive aspect repeatedly. The second round of elicitation was intended to encourage the children to use the perfective aspect, by narration to a puppet, who was asleep, and who now insisted on being told 'what happened'. Repeated attempts were be made to switch the children's narration to the perfective aspect, which sometimes also included prompting with the perfective form of the verb. The objective was to obtain utterances with an ergative marked subject, which appear in the perfective aspect of transitive clauses in Hindi.

2.1 The use of Morphological Case by children

Looking at the children's use of morphological case markers on nominal phrases in the targeted predicate structures, we observed a very high accuracy in the use of all case forms. We briefly discuss below these findings for the use of different case markers that this study aimed to capture.

With the children's preference to use the progressive aspect, it was predicted during the course of the experiment that they would display a large number of ergative marking omissions in their use of perfective aspect. This however was not the case, as children were seen to use the ergative marker on the subject of a perfective aspect construction with ease and accuracy. Of the 580 perfective responses with overt subjects, there were only 15 instances (approximately 2.5%) by 6 children where the ergative marker on the subject was omitted. A perusal of the conversation around these omissions leads us to attribute each of these to performance, owing to a difficulty in switching the narrative from the progressive to the perfective aspect. The following utterances instantiate the correct use of ergative marked subjects in the perfective aspect for a simple predicate (3) and a complex predicate (4).

(3)	laRke	ko	aunty	**ne**	khiiNcaa	(AG1:3;7)
	boy-OBL. M.SG.	ACC.	aunty F.SG.	ERG.	pull-PERF. M.SG.	
	'The aunty pulled the boy'					

---

this picture push give-INFI GEN. about LOC. Be-PRES.
'This picture is about pushing'

(4)	cook	**ne**	ek laRkii	kaa	piichaa	kiyaa thaa	(TA:4;7)
	cook	ERG.	ne girl	GEN.	follow	do-PERF.	be-PST.M.SG
	M.SG.		F.SG.	M.SG.		M.SG.	3P.M.SG.
	'The queen followed the soldier'						

This finding confirms Narasimhan's (2005) results in which children were found to be sensitive to the role of agency and aspectual conditions in their use of split-ergative case-marking. Moreover, her finding that children do not over-extend the *–ne* ergative marker to agents of transitive constructions in non-perfective contexts is corroborated by our study as well. There were 596 utterances in the non-perfective aspect with a non-omitted subject, and each of these carried a null case marking on them, as required in the adult grammar of Hindi.

Similarly, there was very high accuracy with respect to the use of overt *-ko* marking on indirect objects (5), the null accusative direct objects (5) and the bare use of the nominal in complex predicates (6).

(5)	raajkumaar	ne	doctor	**ko**	ghaRii	dii	(HA:4;7)
	prince	ERG.	doctor	ACC.	watch	give-PERF.	
	M.SG.		M.SG.		F.SG.	F.SG.	
	'The prince gave a watch to the doctor'						

(6)	cook	ne	raanii	**ko**	kiss	kiyaa	(HA:4;7)
	cook	ERG.	queen	ACC.	kiss	do-PERF.	
	M.SG.		F.SG.		M.SG.	M.SG.	
	'The cook kissed the queen'						

There was approximately 94% accuracy in the use of *–ko* marking on indirect objects (out of 404 overt indirect objects in the children's responses), almost a 99% accuracy for null accusative direct objects (out of 413 overt direct objects), and 93% accuracy in *–ko* marking on objects in transitive predicates (out of 560 overt objects). Each of the errors in these categories is safely presumed to be the result of reasons other than a shortfall in the children's application of the rules governing case in their grammar.[6] Limiting the discussion of the children's use of these case markers to these bare facts, we emphasize that children have attained the full adult Hindi competency by the age of 3, for the use of morphological case markers and assignment. We discuss this adult grammar of Hindi in the next section.

3. Verb Agreement in the adult grammar of Hindi

The noun class system of Hindi assigns a masculine or feminine gender to each noun in the lexicon and has a binary number distinction of singular and plural features in the

[6] The errors in the use of the dative *–ko* on indirect objects in ditransitives are all in the predicates for 'write' and 'send', in each of which there is an alternate assignment of Beneficiary theta role to the argument, instead of Theme role. The former of these takes a different morphological case than the latter. The errors in the –ko marking on objects in transitive predicates can be attributed to either of the following: wrong theta role assignment, or the use of an unfamiliar complex predicate for the first time by the child.

grammar. These features manifest in the form of an inflectional agreement system that shows up on various grammatical categories such as adjectives, the possessive marker and verb categories, the last of which is what this study focuses on. The main verb participles in their future, imperfective and perfective forms, progressive auxiliaries, mood auxiliaries and the verbal head of an N+V complex predicate show inflection for agreement with a nominal in the clause. This agreement is subject to a combination of transitivity and aspect of the clause in Hindi, which is more specifically an aspectual-split in the agreement paradigm, resulting in agreement with the highest non-overtly case-marked nominal.

In the imperfective aspect,[7] as seen in (7) and (8), the main verb is in its bare form, the progressive auxiliary agrees with the subject in gender and number, and the present auxiliary as well agrees in person and number with the subject.

(7)	Ram	kitaab	paRh	rah**aa**	hE
	Ram	book	read	PROG.	be-PRES.
	M.SG.	F.SG.		M.SG.	3P.SG.
	'Ram is reading a book'				

(8)	mEN	axbaar	paRh	rah**ii**	huuN
	I	newspaper	read	PROG.	Be-PRES.
	F.SG.	M.SG.		F.SG.	1P.SG.
	'I am reading the newspaper'				

The same tense and aspect conditions will trigger the same agreement pattern in complex predicates and ditransitive predicates.

In the Perfective aspect in Hindi, on the other hand, gender and number agreement on the perfective participle is with the object, but the present tense auxiliary in this case does not seem to agree in person at all with the present nominals in the clause, as can be seen in (9) and (10).

(9)	Ram	ne	kitaab	paRh**ii**	hE
	Ram	ERG.	book	read-PERF.	be-PRES.
	M.SG.		F.SG.	F.SG.	3P.SG.
	'Ram has read the book.'				

(10)	mE	ne	axbaar	paRh**aa**	hE
	I	ERG.	Newspaper	read-PERF.	be-PRES.
	F.SG.		M.SG.	M.SG.	3P.SG.
	'I have read the newspaper'				

There is, however, an additional difference in the agreement system in the past tense. In the progressive past, the tense auxiliary will agree in number and gender only. The progressive auxiliary and the main verb behave the same way as in the progressive present.

[7] We limit the use of the term 'imperfective' to 'progressive' in this paper.

(11) Ram kitaab paRh rah**aa** th**aa**
Ram book read PROG. be-PAST.
M.SG. F.SG. M.SG. 3P.M.SG.
'Ram was reading the book'

(12) tum axbaar paRh rah**ii** th**ii**
you newspaper read PROG. be-PAST
2P. M.SG. F. F.SG.
'You (F) were reading the newspaper'

In the perfective past, the perfective participle agrees with the object for gender and number.

(13) Ram ne kitaab paRh**ii** th**ii**
Ram ERG. book read-PERF. be-PAST
M.SG. F.SG. F.SG. 3P.F.SG.
'Ram had read the book'

(14) tum ne kitaab paRh**ii** th**ii**
you ERG. book read-PERF. be-PAST.
2P. F.SG. F.SG. 3P.F.SG.
'You had read the book'

An important rule that is seen to operate in the adult grammar of Hindi is that which limits the scope of person agreement to the subject. Across all tenses and aspects, it is only the tense auxiliary of Hindi that shows agreement for person (and number), whereas perfective participles and the progressive *rah-* auxiliary agrees only for gender and number. We resort to the putative universal of Baker (2008) Structural Condition on Person Agreement (SCOPA) to discuss and explain the operation of this limitation on person agreement.

(15) Structural Condition on Person Agreement
A functional category F can bear the features +1 or +2 if and only if a projection of F merges with an NP that has that feature, and F is taken as the label for the resulting phrase.

The central intuition underlying SCOPA is that the person agreement must always involve a SPEC-head configuration outside the vP, and where such configurations are not achieved agreement is always partial, based on number and gender alone. Baker (2008, 2011) discusses these restrictions on person agreement with cross linguistic evidence for partial agreement on predicate adjectives (Swahili, Hindi, Spanish, Arabic, Mayali and Tariana), partial agreement with *wh*-expressions in [Spec,CP] (nonstandard English), the impossibility of number and gender/animacy agreement with the first person theme argument in double-object constructions in structures like 'She sent me to him' (Nahuatl, Southern Tiwa and Shambala), the impossible long distance agreement with a first

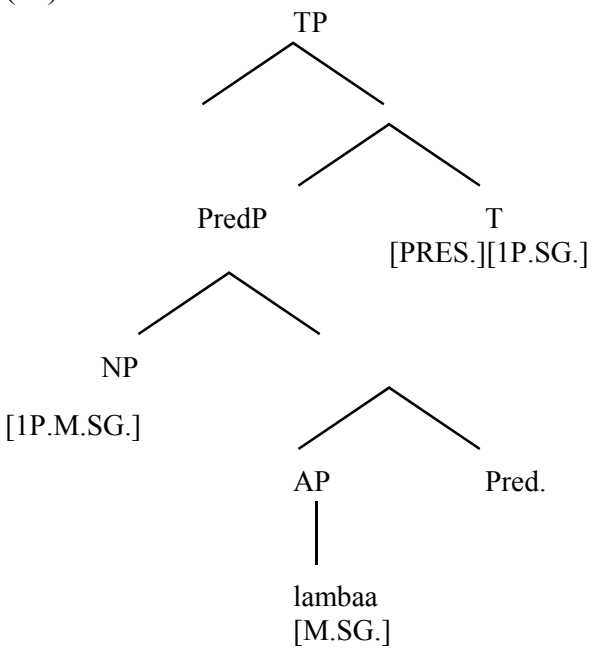
person object in a gerund construction (Lokaa), and Dative subject constructions with first or second person objects (Chicasaw and Icelandic).

In Hindi, just as SCOPA predicts, there is partial agreement on predicate adjectives, as seen in (16) and (17), where the *–aa* ending adjective agrees in gender and number with the subject, whereas the copula agrees in gender, number as well as person with the subject. This partial agreement follows from Baker's (2008) analysis of adjectival predication, whereby the subject is merged to [Spec,PredP], which disallows licencing of person agreement. The derivation of (17) would then look like that in (18).

(16)	tum	sab	laRke	lambe	**ho**
	you	all	boys-OBL.	tall	be-PRES.
			M.PL.	M.PL.	2P.PL.
	'All you boys are tall'				

(17)	mEN	lambaa	**huuN**
	I	tall	be-PRES.
	(M.SG.)	M.SG.	1P.SG.
	'I (M) am tall'		

(18) *mEN lambaa huuN*



Looking back to the examples of agreement with non-subject nominal phrases that we have described in this section, it is obvious that in the adult grammar of Hindi speakers, SCOPA would always be respected as direct/indirect objects or the nominal parts of a NV complex predicate can never raise to [Spec,TP].[8] The other probable probe v lacks an

[8] Double object constructions are not exactly the ideal contexts to show that SCOPA holds for Hindi, for the simple reason that either of a theme direct object or a goal indirect object will necessarily be marked for morphological case if it is a first or second person argument. Such differential object marking, that is,

EPP feature, so 1P and 2P agreement will never be allowed in this part of the structure. As a consequence, when the internal arguments (IA) become accessible for agreement, they only trigger agreement for number and gender and never person on the verbal perfective/imperfective participle forms of the verb[9].

These restrictions on person agreement then lead the diagnosis that the PROG auxiliary originates quite low in the structure and never raises to T. Consequently we never expect it to show agreement for person. The tense auxiliary though must raise to at least the vicinity of T, as there is always a person agreement on this auxiliary for person in this context. In the perfective too the aspectual head must never move to the vicinity of T, as we saw that in this context, there is never a person agreement on the perfective aspect.

The implications for this general analysis of the adult Hindi grammar predict that in non-perfective aspects all Hindi subjects should trigger subject agreement conforming to SCOPA. Secondly, it is expected that in the perfective aspect all Hindi direct objects should trigger SCOPA respecting object agreement. However, we see in the following examples that these expectations are not fully applicable in all contexts.

(19)	tum	ko /	ram	ko	bhuukh	lag	rah**ii**	hE
	you	DAT./	ram	DAT.	hunger	feel	PROG.	be-PRES.
	2P.		M.SG.		F.SG.		F.SG.	3P.SG.
	'You are/ Ram is feeling hunger'							

(20)	mujh	se /	ram	se	hii	ye khiRkii	khult**ii**	hE
	I-OBL.	INST.	Ram	INST.	EMPH.	this window	open-IMPERF	be-PRS.
	1P.SG.		M.SG.			F.SG.	F.SG.	3P.SG.
	'Only I am/ Ram is able to open this window'							

In the experiencer dative structure in Hindi in (19), the subject does not trigger agreement, as is also the case in the abilitative construction in (20). It is the nominal in the N+V complex that controls the agreement in the former and in the latter, it is the object.

---

the accusative *–ko* marking is obligatory for these objects owing to their high prominence on the Animacy and Definiteness scale (Aissen 2003).

(i)	Ravi ne	mujh	ko	dilli	bhejaa	hE	
	Ravi ERG.	I-OBL.	ACC.	delhi	send-PERF.	be-PRES.	
	M.SG.	DO (F.SG.)			IO-F.SG.	M.SG.	3P.SG.
	'Ravi has sent me to Delhi'						
(ii)	Ravii ne	mujh	ko	billii	dii	hE	
	Ravi ERG.	I-OBL.	ACC.	cat	give-PERF.	be-PRES.	
	M.SG.	IO-(F.SG)			DO-F.SG.	F.SG.	3P.SG.
	'Ravi has given me a cat'						

[9] A full discussion of the range of independent evidence that indicates that Hindi subjects do raise to [Spec,TP] would take us too far afield. See Kidwai (2000) and the references to earlier work cited therein for evidence from the subject orientation of reflexives and the anti-subject orientation of pronouns as evidence for a raising to [Spec,TP] analysis of Hindi external arguments.

(21) Ram ne almaarii dekh**ii** th**ii**
Ram ERG. cupboard see-PERF. be-PAST.
M.SG. F.SG. F.SG. 3P.F.SG.
'Ram had seen the cupboard'

(22) Ram ne almaarii ko dekh**aa** th**aa**
Ram ERG. cupboard ACC. see-PERF. be-PAST.
M.SG. F.SG. M.SG. 3P.M.SG.
'Ram had seen the cupboard'

In the simple transitive structure in the perfective aspect in (21), the object controls agreement as predicted by the analysis so far, but if the same object is *–ko* marked for differential object marking, this agreement is not allowed. Instead, there is a default no agreement, that is, a masculine singular agreement appears as a default form.

The evidence provided in the discussion of structures like those in (19) to (22) have led Hindi grammarians to conclude that the object agreement at play in the ergative is only a sub-case of a larger condition on verb agreement in Hindi, by which morphological Case-marking renders noun phrases invisible to agreement processes, which must then look to the next available unmarked argument. We frame the larger condition in more contemporary terms as stated in (23).

(23) The Hindi Case Blocks Agreement Condition
Overt Case marking renders the phi-features of nominal phrases invisible for agreement (i.e. to T and v probes)

The CBAC forces object agreement where the subject is overtly case-marked, or a default agreement in case the object too is overtly case marked. Kashmiri, Marathi amd Punjabi are other ergative Indo-Aryan languages known to show this property of overt case-marking rendering the nominal phrase invisible for agreement[10]. This entails that the Hindi verb agreement is with the highest bare (not morphologically case-marked) argument.

Drawing all the strands of the discussion thus far together, the adult grammar of Hindi as far as verb agreement is concerned is one in which SCOPA and the CBAC work in tandem. While SCOPA regulates which features can be expressed by the agreement relation based on the structural configuration of the agreeing heads, the CBAC determines which noun phrases can count as agreement controllers in the first place. In effect, SCOPA forces non-subject verb agreement to be insensitive to person.

4. The Acquisition of Verb Agreement

With SCOPA and CBAC restricting and governing verb agreement in the adult Hindi grammar, we try to look for evidence if the same are to be found functional in the

[10]Nepali and Gujarati, on the other hand, show agreement with ergative marked subjects and with overtly marked accusative objects respectively. Bhatt (2005) suggests that this difference in the treatment of case-markers and their ability to block agreement may be a matter of parametric variation in the languages.

children's developing grammar, that is, if either or both are hard-wired, or if one is and the other in not. If CBAC were active in the developing grammar, but SCOPA is not, then we expect children to do person agreement on the PROG and the TENSE auxiliaries in non-perfective contexts. If, however, SCOPA is active, but CBAC is not, then we expect children to show gender and number agreement on the PROG auxiliary and person agreement on the TENSE auxiliary with an overtly case marked argument. With these questions in mind, we look at the data for verb agreement in the corpus and talk about the observations in detail in this section.

Unlike the high accuracy in our findings with respect to Case which we discussed in Section 2, we found that children made more errors in verb agreement both in the progressive and perfective aspects. The numbers per child for total utterances in the progressive and the perfective along with the rate of accuracy in verb agreement are presented in the table below. Boldfaced percentages represent more than one instance of an error[11].

(24) *Summary of Agreement Errors in the Progressive and Perfective Aspect*

S.NO.	CHILD	AGE	TOTAL UTTERANCES			ACCURACY IN VERB AGR. %[12]	
			TOTAL	PROG.	PERF.	PROG.	PERF.
1.	CP	3;5	56	29	27	**86.20**	96.29
2.	ST2	3;7	59	30	29	96.66	**75.86**
3.	RG1	3;7	49	26	23	**80.76**	100
4.	AG1	3;7	53	26	27	**88.46**	100
5.	SS1	3;9	63	32	31	100	**87.09**
6.	RG2	4;3	53	23	30	**86.95**	**83.33**
7.	SR	4;4	59	30	29	**80**	**75.86**
8.	TA	4;7	60	33	27	100	100
9.	HA	4;7	59	31	28	100	**92.85**
10.	RM	4;7	61	32	29	96.87	**82.75**
11.	AS	4;7	65	33	32	100	**93.75**
12.	TB	4;9	64	31	33	100	96.96
13.	KR	5;0	57	28	29	100	96.55
14.	NK	5;1	61	31	30	100	96.66
15.	HSS	5;3	64	32	32	**90.62**	**87.50**

[11] The entire corpus comprised of over 8000 utterances. The 17 pairs of pictures, elicited twice for each of the 21 children, gave a target number of 1428 responses. For each of the target sentences, the response closest in similarity to the target structure was chosen as representative. Eliminating the irrelevant responses, null responses, gave us a total of 1238 responses for purposes of the statistical analysis of the corpus.

[12] All percentages have been approximated to two decimal places.

16.	HP	5;5	64	33	31	96.96	**90.32**
17.	SS3	5;6	59	33	26	100	96.15
18.	IS	5;6	49	33	16	**75.75**	**62.50**
19.	AG2	5;8	54	31	23	96.77	100
20.	SM	5;9	65	33	32	100	100
21.	AKS	5;11	64	33	31	96.96	**80.64**
TOTAL			**1238**	**643**	**595**	**93.95**	**90.24**

In the sub-sections that follow, we discuss these errors in the progressive and perfective aspects in some detail, analysing the possible causes for these errors based on the pattern of their occurrence, if any.

4.1 Acquisition of Verb Agreement in the Perfective Aspect

In the perfective aspect with an ergative marked subject, agreement on the perfective participle and the TENSE auxiliary should be with the non-case marked direct object, or there is a default agreement of third person, masculine, singular if the object is *–ko* marked for DOM, or in the case of complex predicates agreement is with the nominal part of the N+V. We see instances of these in the children's utterances, as seen in (25), (26) and (27).

(25) watchman ko ghaRii dikhaaii mere papa ne (CP: 3;5)
watchman DAT.. watch show-PERF. my father ERG.
M.SG. F.SG. F.SG. M.SG.
'My father showed the watch tothe watchman'

(26) aunty ne laRkii ko nehlaayaa (TA:4;7)
aunty ERG. girl ACC. bathe-PERF.
F.SG. F.SG. M.SG.
'The aunty bathed the girl'

(27) doctor ne nurse ko gudgudii kii (SM:5;9)
doctor ERG. nurse ACC. tickle do-PERF.
M.SG. F.SG. F.SG. F.SG.
'The doctor tickled the nurse'

The table below quantifies the findings of the study for children's use of verb agreement in the perfective aspect.

(28) *Children's Verb Agreement in the Perfective Aspect*

No. Of perfective responses	595
No. Of perfective responses with overt subject	580
No. Of correct agreement in perfective	539
No. Of incorrect agreement in perfective	56

Incorrect agreement (not with DO/N of N+V)	56

Ditransitives	36	M agreement instead of F agreement F agreement instead of M agreement	32 4

Complex predicates	19	M agreement instead of F agreement F agreement instead of M agreement	12 7

Transitive	1	F agreement instead of M agreement	1

Agreement with Ergative Subjects: 0

Of the 595 perfective responses, there were 56 responses with incorrect agreement on the perfective participle, which is less than 10 percent. These errors were in 36 responses for ditransitive predicates ('give', 'send', 'write', 'show'), 19 of complex predicates ('tickle', 'kiss', 'hug', 'push', 'help') and 1 simple predicate ('bathe'). In either predicate type, the only discernible pattern in these errors that may give an insight into the children's application of the rules of Hindi agreement is that in 44 of the 56 errors, which make up close to 80% of these, masculine agreement appears when the gender value of the controller of agreement is feminine. Because this masculine agreement occurs irrespective of the grammatical gender of either of the arguments of the sentence, we analyse this to be an overgeneralization of the default system of no agreement in these responses.

In the ditransitive constructions, gender and number agreement on the perfective participle was required to be with the theme DO with non-overt case marking, but there is ambiguity in these instances between the children using a default masculine agreement, as it seems to be in (29) (all the referents to the nominals are singular in number), or assigning a masculine gender to the borrowed English counterpart for the nominal instead of the feminine gender value of the Hindi counterpart, as it seems to be in (30), or the children not assigning the correct gender value to particular nominal, as in (31).

(29) uncle ne aunty ko ciTThii bhejaa (ST2:3;7)
uncle ERG. aunty ACC. letter send-PERF.
M.SG. F.SG. F.SG. M.SG.
'The uncle sent a letter to aunty'

(30) aadmii ne laRkii ko letter bhejaa (SS1:3;9)
man ERG. girl ACC. letter send-PERF.
M.SG. F.SG. F.SG. M.SG.
'The man wrote a letter to the girl'

(31) cook ne raajkumaar ko cammac dii (HA:4;7)
cook ERG. prince ACC. spoon give-PERF.
M.SG M.SG. M.SG. F.SG.
'The cook gave the spoon to the prince'

In the complex predicate structures, the errors in verb agreement on the perfective participle made by children appear to be caused by the use of a previously unfamiliar predicate, or the result of an incorrect gender value assigned to the nominal part of the N+V complex predicate in the children's vocabulary.

(32) nurse ko doctor ne gudgudi karaa
nurse ACC. doctor ERG. tickle do-PERF.
F.SG. M.SG. F.SG. M.SG.
'The doctor tickled the nurse'

An important point to note here though is that there is absolutely no systematic evidence to suggest an agreement controlled by ergative marked subjects in these erroneous instances of perfective constructions.[13] Besides, in spite of all nominals in the study being third person referents, we saw absolutely no violation of person agreement on the perfective participle. These facts very strongly indicate the presence of both SCOPA and CBAC in the children's grammar.

4.2 Acquisition of Verb Agreement in the Progressive Aspect

Agreement in the imperfective aspect in adult Hindi is with the nominative (null case-marked) subject for gender and number on the PROG auxiliary, and for number and person on the TENSE auxiliary. While 9 children have a hundred percent accuracy in their use of verbal agreement (for gender and number[14]) in the progressive aspect, 5 children had only one error, which we may attribute to a performance error, as a non-recurring phenomenon. Even though the average percentage of verb/auxiliary agreement errors in the represented sample set is less than 7% of the total number of responses in the progressive aspect, these are mostly concentrated in those of 7 children.

[13] The only non-nominative subjects in the study were ergative subjects, but we predict that these findings hold for other non-nominative subjects as well, such as the instrumental marked abilitative constructions and the dative marked experience subject constructions.

[14] Since all the referents to the target nominal phrases in the pictures were singular in number, but varied in gender, we are assuming that gender and number agreement occurs simultaneously.

(33) Children's Verb Agreement in the Progressive Aspect

No. of progressive responses	643
No. of correct agreement in progressive responses	606
No. of incorrect agreement in progressive	37

Category of error	**No. of errors**	**% of errors**
Agreement with F object/IO instead of M SU on PROG-AUX.	12	**32.43**
Agreement with M object/IO instead of F SU on PROG-AUX.	22	**59.45**
Ambiguous[15]	3	8.11
Total	37	100

Agreement with F object/IO on PROG AUX and M SU on TENSE AUX.[16]	**3**

The 12 errors in which feminine agreement appears instead of masculine are those in which agreement on the *rah-* progressive auxiliary is clearly controlled by a morphologically case marked object instead of the non-case-marked subject, as seen in (34), (35) and (36).

(34) laRkii laRke ko uThaa rahaa hE (AG1:3;7)
girl boy ACC. lift PROG. be-PRS.
F.SG. M.SG. M.SG. 3P.SG.
'The girl is lifting up the boy'

(35) raajaa is laRkii ko khiiNc rahii hE (SR:4;4)
king this girl ACC. pull PROG. be-PRES.
M.SG. F.SG. F.SG. 3P.SG.
'The king is pulling this girl'

(36) queen cook ko tickly kar rahaa hE (IS:5;6)
queen cook ACC. tickle do PROG. Be-PRS.
F.SG. M.SG. F.SG. M.SG. 3P.SG.
'The queen is tickling the cook'

There are 22 instances of ungrammatical masculine agreement instead of feminine agreement, as in (34) and (36) above and (37)[17], (38) below. It could be argued that some of these are a result of a default, no agreement system allowing a masculine singular agreement, but we see that each of these utterances has a masculine gender object with overt case-marking.

[15] The ambiguities in these 3 errors arise, as it appears the grammatical gender value assigned to the nominal is uncertain in the grammar of these children.

[16] These 3 utterances were not part of the representative sample set, but are part of the children's attempts to utter the target response. The reasons to especially include them in discussion here will become clear as the analysis proceeds in this section.

[17] The utterances in (36) and (37) are by two different children, hence not the same.

(37)	queen	cook	ko	tickly	kar	rahaa	hE	(RG2:4;3)
	queen	cook	ACC.	tickle	do	PROG.	be-PRS.	
	F.SG.	M.SG.		F.SG.		M.SG.	3P.SG.	
	'The queen is tickling the cook'							

(38)	ek fairy	magician	ko	kiss	kar	rahaa	hE	(IS:5;6)
	one fairy	magician	ACC.	kiss	do	PROG.	be-PRS.	
	F.SG.	M.SG.		F.SG.		M.SG.	3P.SG.	
	'A fairy is kissing the magician'							

Data such as that in (34) to (38), suggests a violation of CBAC in these utterances, allowing a morphologically case marked nominal to control the agreement on PROG auxiliary. It is not clear, however, if the TENSE auxiliary is in its default form, or it is agreeing with the subject or the object. The following three utterances, two of which are by the same child[18], not only support our observation of agreement with a case marked object, but also provide additional clues to the malfunctioning system in place in the developing grammar.[19]

(39)	aur	raajaa	laRkii	ko	gift	de	rahii	thaa	(SR:4;4)
	and	king	girl	ACC.	gift	give	PROG.	be-PAST	
		M.SG.	F.SG.		M.SG.		F.SG.	M.SG.	
	'and the king was giving a gift to the girl'								

(40)	ek laRkaa...	us	kaa	madad	kar	rahii	thaa	(RM:4;7)
	one boy	that	GEN.	help	do	PROG.	be-PAST	
	M.SG.	SG.	M.SG.	F.SG.		F.SG.	M.SG.	
	'A boy was helping her'							

(41)	ek laRkaa...	laRkii	kii	madad	kar	rahii	thaa	(RM:4;7)
	one boy	girl	GEN.	help	do	PROG.	be-PAST	
	M.SG.	F.SG.	F.SG.	F.SG.		F.SG.	M.SG.	
	'A boy was helping a girl'							

These instances of a simultaneous agreement, as seen in (39), (40) and (41), show more than one layer of agreement, wherein the aspectual head agrees in gender with the case-marked feminine object, and the past TENSE auxiliary agrees with the null case-marked masculine subject.

Even though this phenomenon of simultaneous agreement with more than one argument is rare in the corpus, it is strongly indicative that agreement on the TENSE auxiliary is restricted by SCOPA to be with the subject only. The CBAC, however,

[18]Refer to footnote 16.

[19] As noted by the anonymous reviewer, all of the examples of CBAC violation show progressive auxiliary agreement with the linearly closer but case marked object, which may reflect memory constraints in the children's grammar or effects of decay of the subject representation. This is an issue that we are aware of and are planning to address in future work. As yet, however, we are not in a position to address this, as our corpus does not contain the relevant information.

appears to be fallible in these children's grammar by the apparent tension in its use in the progressive and perfective aspects. We suggest that this is a grammar in transition for two reasons: first, the rate of the occurrence of these errors is very low; and second, younger children seem to make more of these errors than the older ones. Three out of five 3 year old children (60%), two out of seven 4 year old children (29%), and two out of nine 5 year old children (23%) had recurring instances of these errors.[20] It appears that the Hindi language specific CBAC takes a while to stabilize in the children's grammar, more so in the progressive aspect than in the perfective aspect. In the next section, we analyse in some detail this developing grammar and the derivational system employed by them, to account for this observable pattern in errors.

5. An analysis of Agreement in a Child's Developing Grammar

In order to analyse the derivational system in the developing grammar of children, we turn to Chomsky's (2000, 2001) operation of Agree to first discuss the adult system that governs phi-feature agreement on the verbal constituents in Hindi. We take recourse to Heck and Richards (2007) version of Agree as given below:

(42) Agree:

α can agree with β with respect to a feature bundle Γ iff a.-d. hold:

a. α bears at least one unvalued probe feature in Γ and thereby seeks the β-value of a matching goal feature β in Γ.
b. α c-commands β.
c. β is the closest goal to α.
d. β bears an unvalued case feature.

Where features are central to the minimalist framework, the uninterpretable features enter into the derivation unvalued and must be valued and deleted from the derivation by spell-out. This valuation takes place in the operation AGREE, whereby a Probe, for instance T or a v, looks in its domain for a set of interpretable phi-features on a Goal, to match its own set of corresponding uninterpretable features. With the v probe object agreement in a transitive clause of Hindi then occurs presumably alongside accusative licensing. Subject agreement, on the other hand, as well as nominative licensing occurs when T is the probe. The ergative case, however, does not relate to either of these, as it is license by the perfective aspect. Although it is often assumed that $ASP_{PERF.}$ and $ASP_{IMPERF.}$ are merged in the same position, we suggest the two have different merge positions in the structure.[21] We consider the implications of these structurally different configurations for the perfective and imperfective aspects, we propose that for some children CBAC is not firmly in place for $vAsp_{IMPERF.}$, which is a biclausal phrase comprising of two monadic predicates. The biclausal structure creates an additional layer

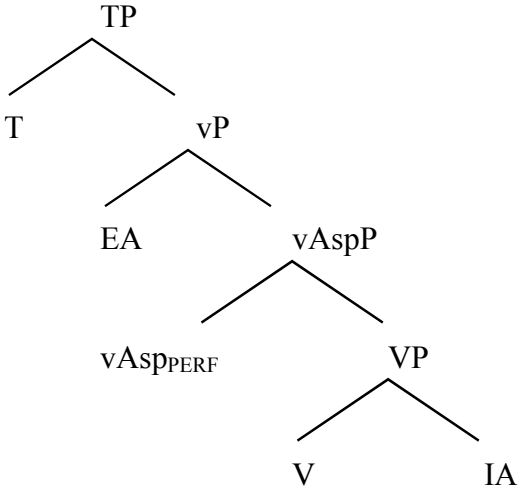
[20] Refer to the table in (24).

[21] A theory of merge of the two aspects in different positions has been explored in recent research, such as that by Coon (2010), which proposes a difference in the clausal structure on the basis of an aspectual split. This analysis argues for a monoclausal structure for perfective aspect and a biclausal structure for the imperfective aspect. While this paper does not dwell into details of this analysis, we think it worthwhile to explore and consider its ramifications in future work.

of predication where the PROG auxiliary takes as complement a lower phrase headed by the main verb.

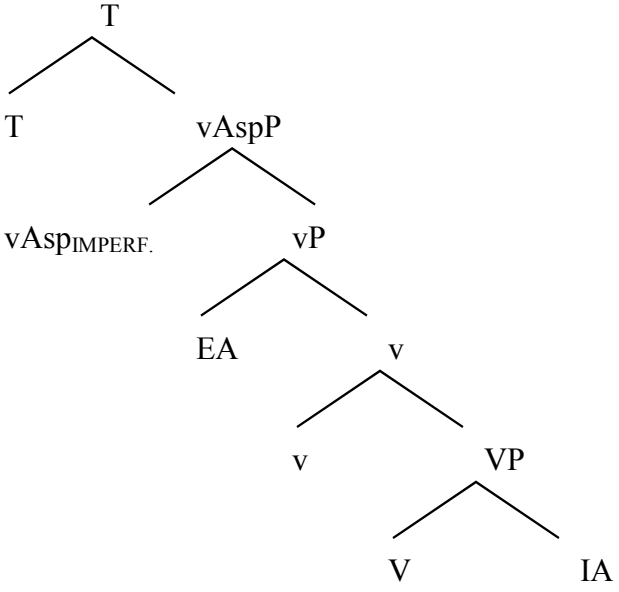
The perfective structure, for which we adopt Anand and Nevins' (2006) approach, is a simpler monoclausal structure that licenses the ergative, which considers the Hindi ergative to be assigned by a perfective v. As can be seen in the structure in (43), the location of the $vAsp_{PERF}$, the only goals available for it to seek for Agree are in the internal argument(s). The subject, therefore, cannot possibly provide the goal for the Asp probe in this configuration.

(43) *Perfective aspect is low in the structure*



The imperfective aspect merges higher in the structure as in (44) below. The external argument (EA) is merged in the [Spec,vP] and the internal argument (IA) in the complement of V, both merged lower than Asp. The Asp is able to then seek goals in both the external and the internal argument(s).

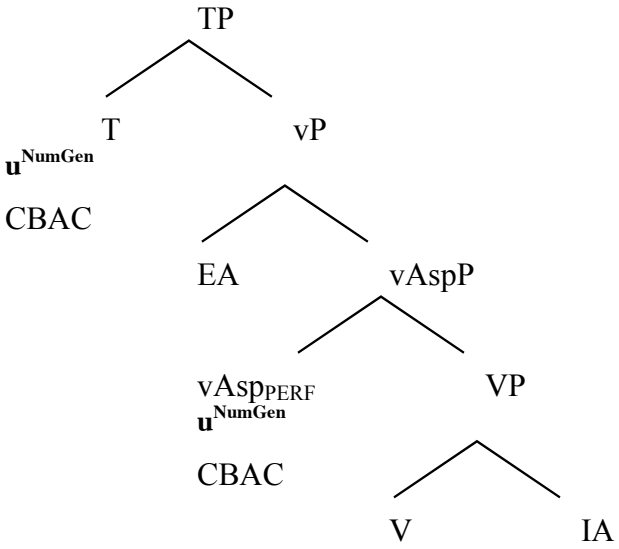
(44) *Imperfective aspect is higher in the structure*



A common characteristic of both the Asp heads in both the perfective and progressive aspects that comes to the forefront here is that they both lack an EPP feature preventing the movement of an argument to its specifier. By SCOPA this accounts for the absence of person agreement in the language on either Asp.

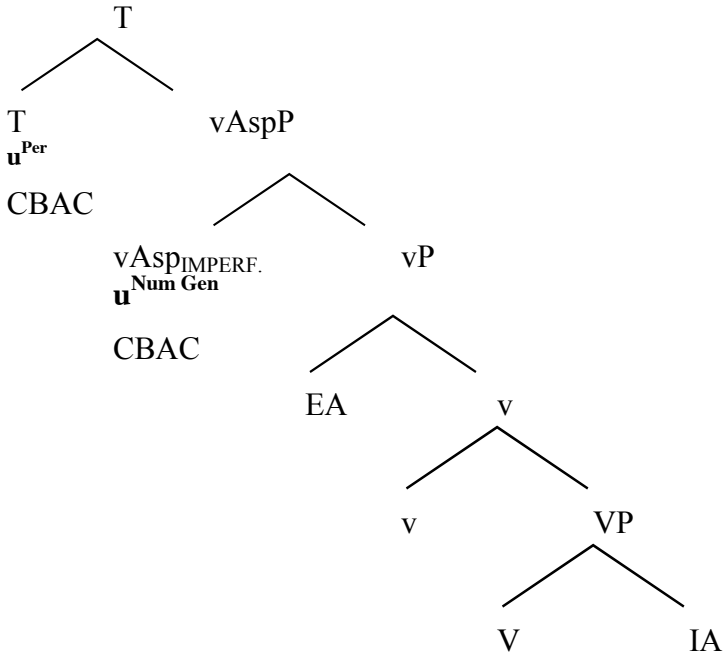
Turning to children's grammar of the perfective aspect to begin with, which we discussed in Section 4.1, we found evidence for the presence of CBAC, similar to its presence in the adult grammar as well. More specifically, for $vAsp_{PERF}$ children seem to have CBAC operational, as a result of which agreement takes place only with objects or in the default. The CBAC for T is also operational in the grammar which prevents an Agree relation with ergative subjects. This structure can be seen in (45) below.

(45) *Adult/children's grammar in the Perfective Aspect*



In nominative subject predicates in the progressive aspect, the Asp finds that the EA is the closest position to probe for Agree. The EA then raises to (Spec,TP] to meet its EPP requirement, where agreement for person takes place. The structure for this adult grammar version of such a clause in the progressive aspect would like that in (46).

(46) *Adult grammar in the imperfective aspect*



In some of the children's version of the ungrammatical derivation of the progressive aspect constructions, we predict that both the EA and the IA are visible to the progressive auxiliary, even though the IA is case marked in these utterances for DOM or the accusative. And since the children are seen to agree with this case marked IA, we suggest that for these children, CBAC is as yet fallible on the $vAsp_{IMPERF.}$ This lack of CBAC on the aspectual head then lets the case marked nominal control the agreement on the *–rah* auxiliary. Agreement on the TENSE auxiliary does not at all come under the purview of this violation as T agrees with the subject only after subject raising. We know this from appearance of a double layer of agreement wherein the aspectual head agrees in gender and number with the case-marked object and the tense marking 'be' auxiliary agrees with the null case-marked subject, as was the case in (39), (40) and (41). This suggests the presence of an additional subdomain for agreement in the children's developing grammar at the vP level that generates such a dual agreement.

6. Conclusion

The observations made in our psycholinguistic study have led us to conclude that the children's grammar lacks in a consistent application of CBAC in the progressive aspect, while it is not so in the perfective aspect. By our analysis SCOPA is available to children's grammar as a hard-wired primitive, whereas the CBAC does not stabilize in the developing grammar until after the age of 5. Surprisingly, his instability seems to have no bearing on the children's learning of case. The relative accuracy in the children's use of different morphological cases in the language indicates that the adult system for case may have been acquired by the children relatively earlier than the agreement system.

References

Aissen, J. 2003. Differential Object Marking: Iconicity vs. economy. *Natural Language and Linguistic Theory 17*. 673-711.

Anand, Pranav & Andrew Nevins. 2006. The locus of Ergative Case Assignment: Evidence from Scope. In *ERGATIVITY*. Springer Netherlands, 3-25.

Baker, Mark. 2008. *The Syntax of Agreement and Concord*. Cambridge, New York: Cambridge University Press.

Baker, Mark. 2011. When Agreement is for Number and Gender but not Person. In *NLLT Springer Science+Business Media*, 29:875-915.Stanford: CSLI Dissertations in Linguisitics.

Bhatt, Rajesh. 2005. Long Distance Agreement in Hindi-Urdu. In *Natural Language and Linguistic Theory* 23: 757-807.

Bhatt, Rajesh, and Vincent Homer. 2014. The Interpretation of Infinitival Negation in Hindi-Urdu. *FASAL-4*, Rutgers University.

Bhatt, Rajesh. 2014. Agreement features. *LISSIM handout*

Chomsky, Noam. 2000. Minimalist Inquiries: The Framework. In *Step by Step : Essays on Minimalist Syntax in Honour of Howard Lasnik*, ed. by R. Martin, D. Michaels and J. Uriagereka, 89-155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by Phase. In *Ken Hale: A Life in Language*, ed. by M. Kenstowicz, 1-52. Cambridge, MA: MIT Press.

Coon, Jessica. 2010. *Complementation in Chol (Mayan): A Theory of Split Ergativity*. Doctoral thesis, MIT.

Heck, Fabian & Marc Richards. 2010. A Probe-Goal Approach to Agreement and Incorporation Restrictions in Southern Tiwa. In *Natural Language and Lintuistic Theory* 28.3: 681-721.

Kidwai, Ayesha. 2000. *XP Adjunction in Universal Grammar: Scrambling and Binding in Hindi-Urdu*. New York: OUP.

Kidwai, Ayesha 2010. Morphological Case is Structural Case: The Case of Hindi Accusative –ko. GLOW in Asia VIII, Beijing.

Legate, Julie Anne. 2008. Morphological and Abstrct Case. In *Linguisitc Inquiry*, 39.1: 55-101.

Mahajan, Anoop Kumar. 1990. *The A/A-Bar Distinction and Movement Theory*. Doctoral thesis. MIT

Mohanan, Tara. 1994. *Argument Structure in Hindi*. CSLI Dissertations in Linguisitics.

Narasimhan, Bhuvana. 2005. Spliting the notion of 'agent': Case marking in early child Hindi. In *Journal of Child Language*, 32(4): 787-803

Ruigendijk, Esther. Contrastive Elicitation Task for Testing Case Marking. Ms., Carl von Ossietzky University, Oldenburg.

Spencer, Andrew. 2005. Case in Hindi. In *Proceedings of the LFG05 Conference*, ed by Mirium Butt and Tracy Holloway King. CSLI Publications.

Benu Pareek, Ayesha Kidwai & Sonja Eisenbeiss
benusharan@gmail.com, ayesha.kidwai@gmail.com, seisen@essex.ac.uk