JSAL

Volume 6 December 2013

Journal of South Asian Linguistics

CSLI PUBLICATIONS

Contents

Editorial Statement

We are very happy to present to you the sixth volume of the Journal of South Asian Linguistics. It comes to you several months later than the date it bears; we hope the excellence of the papers that make up this volume will make up in a small way for this delay. The volume consists of two papers and a book review.

The first paper is by Emily Manetta and it investigates sluicing in Hindi-Urdu. Sluicing has been a topic of much interest for the syntax-semantics interface at least since the publication of a seminal paper by Chung, Ladusaw, and McCloskey in 1995. A commonly assumed treatment of sluicing analyzes it as clausal ellipsis that leaves out the *wh*-phrase. Given this assumption, the very existence of sluicing in wh-in-situ languages becomes surprising. This is where Manettas paper comes in. It argues that even though Hindi-Urdu is in general descriptively wh-in-situ, in sluicing contexts a higher copy of the *wh*-phrase is pronounced. The notion of 'wh-in-situ' is not a primitive property of the syntax of Hindi-Urdu; instead it is contextually determined.

The second paper is by Pritty Patel-Grosz and it provides a solution for a very curious puzzle concerning the binding of anaphors in English, Greek, and Kutchi Gujarati. It is well known that reflexive pronouns like *himself* need to be locally bound. Patel-Grosz shows that in the languages under investigation, the local binding restriction only holds as long as the anaphors are not modified (e.g., 'himself' vs. 'his pathetic self'). She proposes that the local binding restriction arises from the incorporation of self into the predicate. If the reflexive is modified, the incorporation is blocked and non-local readings become available.

The volume concludes with a book review of an edited volume (Ibero-Asian Creoles: Comparative Perspectives. Edited by Hugo C. Cardoso, Alan N. Baxter, and Mário Pinharanda Nunes. Amsterdam: John Benjamins Publishing Company. 2012.) by Adam Roth Singerman.

Finally, we would like to thank Sebastian Sulger for continuing to maintain the journal website. As ever we also thank Dikran Karagueuzian of CSLI Publications for his role as a continually outstanding and supportive publisher.

Rajesh Bhatt, University of Massachusetts, Amherst
Miriam Butt, University of Konstanz
Tracy Holloway King, eBay Inc.

Copy theory in *wh*-in-situ languages: Sluicing in Hindi-Urdu

Emily Manetta, *University of Vermont*

Received August 2012; Revised December 2013

Abstract

Hindi-Urdu is known to be one of the *wh*-in-situ languages exhibiting a sluicing-like construction. Although many have proposed alternative accounts of such strings in *wh*-in-situ languages (e.g. Kizu 1997, Toosarvandani 2009, Gribanova 2011, Hankamer 2010), I argue that apparent sluicing in Hindi-Urdu can be analyzed in a manner consistent with the notion that the syntax of a sluice is the syntax of a regular wh-question (Ross 1969, Merchant 2001). Assuming the copy theory of movement (Chomsky & Lasnik 1993, Chomsky 1993, i.a.), we can understand sluicing in Hindi-Urdu as an exceptional instance of the pronunciation of the top copy in a *wh*-chain, correctly predicting that Hindi-Urdu sluiced structures have properties similar to genuine sluices in languages like English. This article pursues a continued refinement in the implementation of copy theory in *wh*-in-situ languages and importantly, contributes to the current line of work investigating intra-linguistic variation among *wh*-in-situ languages and the ways in which constellations of properties of *wh*-dependencies and ellipsis processes in these languages are best understood.

1 Introduction

As has been widely reported in the literature, Hindi-Urdu, traditionally understood to be a *wh*-in-situ language, features a construction that appears to be sluicing:

(1) mãĩ=ne yahãã kisi=ko dekh-aa
1SG=ERG here someone.OBL=ACC see-PFV.M.SG
par mujhe nahĩĩ pataa kis=ko.
but 1SG.DAT not know who.OBL=ACC
'I saw someone there, but I don't know who.'

Many have suggested that apparent sluicing in *wh*-in-situ languages challenges the influential approach to sluicing which posits that the syntax of a sluice is the syntax of an ordinary *wh*-question (Ross 1969, Merchant 2001), and have proposed alternative strategies for deriving the sluicing-like string. One family of accounts suggests that what looks like sluicing in these languages is in fact an instance of a reduced copular clause (RCC) (Kizu 1997, 2000, Merchant 1998, Gribanova 2011, i.a.). In these accounts the apparent sluice should exhibit properties similar to that of copular clauses or clefts. Another family of accounts claims that the movement feeding the deletion operation in otherwise *wh*-in-situ languages is somehow exceptional — that it is not ordinary *wh*-question formation in the language. For instance, Toosarvandani (2009) and Malhotra (2009) suggest that focus fronting to a dedicated focus projection located between CP and TP is the displacement feeding sluicing-like ellipsis in Farsi and Hindi-Urdu respectively. Manetta (2011) and Bhattacharya & Simpson (2012)

JSAL Volume 6, December 2013.

similarly claim that the movement of the *wh*-remnant is exceptional, but they posit that it targets the CP domain. Finally, in earlier work, Manetta (2006) proposed that what appears to be sluicing in Hindi-Urdu might be the elision of a constituent smaller than a TP (vP-ellipsis).

In this article I show that none of the approaches above adequately account for the properties of the apparent sluicing construction in Hindi-Urdu. I propose instead that we can maintain Merchant's (2001) core understanding of sluicing — that it is a non-pronunciation of a clause-sized constituent following ordinary *wh*-movement — if we adopt a novel account of sluicing in Hindi-Urdu. The account I develop here assumes the copy theory of movement (Chomsky & Lasnik 1993, Chomsky 1993, i.a.) and posits that so-called "*wh*-in-situ" in Hindi-Urdu is in fact a preference for pronunciation of the lower copy in a *wh*-movement chain (Groat & O'Neil 1996, Reintges, Lesourd, & Chung 2006, Reintges 2007; see also Bošković & Nunes 2007, Bošković 2011). Sluicing is then an exceptional instance of the pronunciation of the higher copy in the *wh*-chain (located in Spec, CP) under pressure from P-RECOVERABILITY, which requires that a member of a chain associated with phonetic content be pronounced (Landau 2006). The schematic in (2) illustrates the derivation of a regular *wh*-question and a sluice in Hindi-Urdu, with strikethrough representing non-pronunciation.

(2) I saw someone there, but I don't know...

a. ... ~~kis=ko~~ mãĩ=ne yahãã kis=ko dekh-aa. **Regular *wh***

b. ... kis=ko ~~mãĩ=ne~~ ~~yahãã~~ ~~kis=ko~~ ~~dekh-aa~~.
... who.OBL=ACC 1SG=ERG here who.OBL=ACC see-PFV.M.SG
'...who (I saw there)'. **Sluice**

Section 2 of this article presents detailed evidence that apparent sluicing in Hindi-Urdu is in fact the elision of a clause-sized constituent and has the characteristics of genuine sluicing in *wh*-movement languages. In section 3, I argue that we might best understand regular *wh*-in-situ constructions in Hindi-Urdu as an instance of pronunciation of the lower copy in a *wh*-chain. Section 4 details the analysis of sluicing in Hindi-Urdu as the exceptional pronunciation of the top copy in a *wh*-chain (what I will call top-copy sluicing here). In section 5, I conclude by exploring the intra-linguistic variation in the constellation of properties referred to as "*wh*-in-situ" and the properties of apparent sluicing structures.

This account seeks to extend the explanatory reach of copy theory in two ways. First, data from Hindi-Urdu supports the claim made elsewhere (Reintges, Lesourd, & Chung 2006, Reintges 2007) that there are apparent *wh*-in-situ languages in which *wh*-dependencies exhibit similarities to overt movement chains, and that these languages can be analyzed as exhibiting a preference for lower-copy pronunciation. Second, the approach to sluicing pursued here is harmonious with a line of work suggesting that various constraints on phonetic output might force the overt realization of a copy other than the preferred copy under certain circumstances (Bošković & Nunes 2007, Bošković 2011).

Finally, and perhaps most importantly, this account is part of the larger effort to advance a more nuanced approach to the phenomenon imprecisely labeled "*wh*-in-situ" and to pursue a better understanding of intra-linguistic variation in the properties of *wh*-dependencies in *wh*-in-situ languages.

2 Sluicing in Hindi-Urdu

The empirical goals of this section are twofold. First, I present new data from Hindi-Urdu arguing against several existing approaches to sluicing-like structures in *wh*-in-situ languages. Second, I bring together a range of facts concerning apparent sluicing in Hindi-Urdu that have been reported piecemeal elsewhere in the literature (Manetta 2006, 2010, Malhotra 2009, Simpson & Bhattacharya 2012). By pulling this body of evidence together, I seek to develop an analysis that accounts for all of the properties of the apparent sluicing construction discussed here.

2.1 Characteristics of putative Hindi-Urdu sluicing structures

There is significant evidence that apparent sluicing in Hindi-Urdu has the properties of sluices from more familiar languages. Displaced *wh*-phrases in Hindi-Urdu must be marked with the case

morphology they would have been assigned in-situ, as in (3). In apparent sluicing structures, Hindi-Urdu exhibits full case connectivity; the *wh*-remnant must be marked with the same case it would exhibit in the non-elided structure, as in (4):

(3) a. sita=ne kis=ko/*kis=ne/*kaun soc-aa
Sita.F=ERG who.OBL=ACC/who.OBL=ERG/who.NOM think-PFV.M.SG
ki ravii=ne ___ dekh-aa?
that Ravi.M=ERG see-PFV.M.SG
'Who did Sita think that Ravi saw?'

b. tum kaun/*kis=ne/*kis=ko soc-te ho ki ___ aa-yegaa.
2SG who.NOM/who=ERG/who=ACC think-HAB be.PRS.2SG that come-FUT.M.3SG
'Who do you think will come?' (Srivstav 1991)

(4) a. mãĩ=ne yahãã kisi=ko dekh-aa, par mujhe nahĩĩ pataa
1SG=ERG here someone.OBL=ACC see-PFV.M.SG but 1SG.DAT not know
kis=ko/*kis=ne/*kaun.
who.OBL=ACC/*who.OBL=ERG/*who.NOM
'I saw someone there, but I don't know who.'

b. kisi=ne aisha=ko dekh-aa par mujhe nahĩĩ pa-taa
someone.OBL=ERG Aisha.F=ACC see-PFV.M.SG but 1SG.DAT not know-HAB.M.SG
kis=ne/*kaun/*kis=ko.
who.OBL=ERG/*who.NOM/* who.OBL=ACC
'Someone saw Aisha, but I don't know who.'

Similarly, Hindi-Urdu requires that post-positions be pied-piped in general (as in (5a)), and they must also be pied-piped in a sluiced structure as in (5b):

(5) a. kis=ke saath aap kaam kar-te hãĩ /
who=GEN.OBL with 2PL work.M.SG do-HAB.M.OBL be.PRS.3.PL /
*kis aap=ke saath kaam kar-te hãĩ?
who.OBL 2PL=GEN.OBL with work.M.SG do-HAB be.PRS.3.PL
'Who do you work with?'

b. sita khaana pak-aa rah-ii hai, par ali=ko nahĩĩ pa-taa
Sita food.M.SG cook-Caus PROG-F.SG be.PRS.3SG but Ali.M=DAT NEG know-HAB.M.SG
kis=ke liye/*kis/*kaun.
who=GEN.OBL for/who.OBL/who.NOM
'Sita is cooking but Ali doesn't know for whom.'

2.2 Sluicing in Hindi-Urdu is not the elision of a phrase smaller than a TP

There are at least three types of evidence that suggest that the elided constituent in a sluice in Hindi-Urdu is indeed clause-sized, or a TP. Each of these involves material positioned in or below TP that must be interpreted within the ellipsis site and/or cannot be present alongside the *wh*-remnant of a sluice.

Though it has been claimed elsewhere (Manetta 2006) that sluicing in Hindi-Urdu could be the elision of a projection of vP, there is evidence to suggest that a larger (that is, TP-sized) constituent is elided. The tense auxiliary *hai* (third person singular present tense form of *ho* 'be'), is elided in an apparent sluicing structure (6):

(6) ali koi kitaab khariid-naa caah-taa hai. ham-ẽ nahĩĩ pa-taa
Ali.M some book.F.SG buy-INF.M.SG want-HAB.M.SG be.PRS.3.SG 1PL-DAT NEG know-HAB.M.SG
kaunsii ~~Ali~~ ___ ~~khariid naa~~ ~~caah taa~~ ~~hai~~.
which.F Ali.M buy-INF.M.SG want-HAB.M.SG be.PRS.3.SG
'Ali wants to buy a book. We don't know which one.'

It is widely assumed by researchers working on the language that the auxiliary *ho* is the overt realization of finite T (Mahajan 1990, Bhatt 2005, Kumar 2006; see also the argumentation in Davison 2002, Kush 2011).[1] If indeed apparent sluicing structures were the elision of a constituent smaller than TP in Hindi-Urdu, we would expect the auxiliary to grammatically appear in (6) above. Further, as is exemplified below in (15)–(16), Hindi-Urdu has no independent process of copula drop. Therefore the elided constituent in (6) is likely to be at least a TP.

The second type of evidence that the sluiced constituent is indeed as large as a TP comes from sentential negation in sluicing structures in Hindi-Urdu. As is clear from (7), sentential negation is interpreted to be within the ellipsis site. Further, negation cannot remain alongside the remnant in a sluicing structure, as in (8)–(9).

(7) a. arjun kisi=se is daftar=mẽ nahĩĩ mil sak-aa, par mujhe
Arjun.M someone=with this office.M.SG=in NEG meet can-PRF.M.SG but 1SG.DAT
nahĩĩ pataa kis=ko.[2]
NEG know who.OBL=ACC
'Arjun couldn't meet with someone in that office, but I don't know who.'

= b. Arjun couldn't meet with someone in that office, but I don't know who *Arjun couldn't meet with in that office.*

≠ c. Arjun couldn't meet with someone in that office, but I don't know who *Arjun could meet with in that office.*

(8) ?*arjun kisi=se is daftar=mẽ nahĩĩ mil sak-aa, par mujhe
Arjun.M someone.OBL=with this office.M.SG=in NEG meet can-PRF.M.SG but 1SG.DAT
nahĩĩ pataa kis=ko nahĩĩ.
NEG know who.OBL=ACC NEG

(9) A: koi arjun=se nahĩĩ mil sak-aa.
someone Arjun.M=with NEG meet can-PRF.M.SG
'Someone couldn't meet Arjun.'

B: kaun? / *?kaun nahĩĩ?
who / who NEG
'Who?'

Sentential negation must either immediately precede or follow the main verb in Hindi-Urdu in linear order. Kumar (2006) argues that negation heads a projection beneath TP, and that Neg-V order is derived via a V-to-T movement that picks up Neg along the way, while in V-Neg order the V remains within VP (Mahajan 1990 and Dwivedi 1991 argue for a similar basic syntactic position for negation).

Assuming that negation falls between TP and the vP-layer in Hindi-Urdu, if the *wh*-remnant in (7) were in Spec, vP (or, indeed, the specifier of a Focus phrase dominated by TP), we would not expect the interpretation of (7) to include negation (as in (7a)). Further we might expect negation in either of these positions higher in the clause to survive sluicing and to be able to follow the *wh*-remnant. However, this is strongly dispreferred. The evidence here suggests that the *wh*-remnant is higher than sentential negation, and that the elided clause is TP-sized.

[1] Cf. Bhattacharya et al. (2000), who present an antisymmetric account of auxiliaries as light verbs in Hindi-Urdu and other South Asian languages.

[2] These must be contexts in which the indefinite can scope over negation; otherwise they are ungrammatical as in English (i):

(i) *She didn't talk to any student, but I don't know who.

Also, for reasons unrelated to this discussion, sluices with *why* can appear with negation in both English and Hindi-Urdu (Horn 1978, Merchant 2000).

(ii) She didn't go to the dance, but we don't know why not.

The third piece of evidence for this claim comes from testing whether TP-adjoined adverbials are present in the interpretation of sluiced structures and can accompany the remnant in a sluice. Bell (2012) points out that left-adjoined adjuncts to TP cannot be stranded under sluicing in English.[3]

(10) a. *One of the employees opted out of the pay raise but I don't know who$_i$ [TP **unbelievably** [TP ~~t_i opted out of the pay raise~~]].

b. Speaker A: One of the employees opted out of the pay raise.
Speaker B: *Who$_i$ [TP **unbelievably** [TP $< t_i >$ ~~opted out of the pay raise~~]]?
(Bell 2012:14)

Bhatia (2006) argues that adverbs evaluating mood, such as *bhaagya=se* 'luckily/fortunately', are adjoined quite high in the Hindi-Urdu clause structure. Though adverbs are known to be challenging to use for diagnosing clausal positioning in Hindi-Urdu, in this case they pattern quite consistently with the rest of the data presented. In short, TP-adjoined adverbials are interpreted to be within the ellipsis site, and cannot comfortably precede a *wh*-remnant in a sluicing structure.[4]

(11) bhaagya=se kisi=ne gaaṛii=ko dekh-aa,
fortune=with someone.OBL=ERG car.F.SG=ACC saw-PRF.M.SG
par mujhe nahĩĩ pataa kis=ne.
but 1SG.DAT NEG know who.OBL=ERG
'Fortunately someone saw the car, but I don't know who.'
= a. Fortunately someone saw the car, but I don't know who fortunately saw the car.

(12) ?*bhaagya=se kisi=ne gaaṛii=ko dekh-aa, par mujhe nahĩĩ pataa bhaagya=se kis=ne.[5]

This is certainly unexpected in an analysis in which the *wh*-remnant is located in Spec, vP or in a specifier of a Focus phrase below TP since the TP-adjoined adverbial should not necessarily be interpreted within the ellipsis site and should preferentially remain, preceding the *wh*-remnant.

A fourth piece of evidence might potentially come from the types of constituent ellipsis available in Hindi-Urdu. Hindi-Urdu has been claimed (like French and Spanish) to not permit verb-phrase ellipsis (VPE) in general (Sailor 2014). Toosarvandani (2009) has investigated a type of VPE available in Farsi that strands light verbs (eliding their VP complements). This so-called v-stranding VPE is entirely unavailable for Hindi-Urdu:

(13) A: aap=ne kitaab=ko phẽk di-yaa?
2PL=ERG bookF.SG=ACC throw give-PRF.M.SG
'Did you throw the book away?'

B: hãã, phẽk diyaa. / *hãã di-yaa.
yes throw give-PRF / yes give-PRF.M.SG
'Yes, threw.' (='Yes, I did.')

If Hindi-Urdu does not permit elision of VP-sized constituents, then an analysis of sluicing in which the *wh*-phrase is in Spec, vP and a VP is elided seems even more unlikely. That said, there is some lack of clarity as to whether VPE is completely impossible and more careful work (along the lines of Goldberg 2005 and Gribanova 2013) is certainly needed. For instance, as Gribanova (2013) points out, if the antecedent of VPE contains a disjunction of two VP-sized constituents containing multiple parts, this cannot be explained away as an instance of argument ellipsis (argued to exist in Hindi-Urdu by Simpson, Choudhury, & Menon 2012). In other words, the ellipsis in (14B) below

[3]Notice that sluicing is grammatical in English if the TP host of the left-adjoined adverbial serves as the antecedent exclusive of the adjunct:

(i) [TP **Unbelievably**, [TP one of the employees opted out of a pay raise]], but I don't know who$_i$ [TP $< t_i >$ ~~opted out of a pay raise~~].

[4]For a similar line of argumentation concerning relative clause ellipsis in Hungarian, see van Craenenbroeck & Lipták (2013).

[5]The judgments exclude the irrelevant (and somewhat odd) interpretation in which *bhaagya=se* 'fortunately' is evaluating the embedding predicate *pataa* 'know' as in 'I don't know who it was, fortunately'.

must be understood as VPE if it is available. Rajesh Bhatt (p.c.) suggests that (14B) may be possible (though not perfect) on the VPE reading:

(14) A: mujhe lag-taa hai ki ram=ne sita=ko santaraa
1SG.DAT seem-HAB.M.SG be.PRS.3SG that Ram.M=ERG Sita.F=DAT orange
yaa mina=ko amruud di-yaa hogaa.
or Mina.F=DAT guava give-PRF.M.SG be.FUT.M.3SG
'It seems to me that Ram gave an orange to Sita or a guava to Mina.'

B: nahĩĩ, ram=ne nahĩĩ di-yaa hogaa.
NEG, Ram.M=ERG NEG give-PRF.M.SG be.FUT.M.3SG
'No, Ram didn't (give an orange to Sita or a guava to Mina).'

Although these facts certainly merit further investigation, I will argue for no particular account of them here. Instead I will simply point out that if indeed the elision of VP-sized constituents is blocked in Hindi-Urdu this might prove challenging to a proposal in which the *wh*-remnant in sluicing structures is understood to be in Spec, vP (Manetta 2006), since it would be mysterious why, just in this special case, a type of VP-elision is allowed. We would then need to explain the mysterious division between sluicing-like cases of VPE and V/v-stranding VPE. On the other hand, if sluicing indeed involves TP ellipsis, this concern does not arise.

2.3 Sluicing in Hindi-Urdu is not a reduced copular clause

Sluicing-like structures in a number of *wh*-in-situ languages seem amenable to analysis as a reduced copular clause with a *wh*-remnant. A reduced copular clause (RCC) is derived via dropping of at least a subject and copula (schematized in English in (15b)), in contrast to a sluice in (15a).

(15) a. I know you saw someone, but I don't know [who (you saw ___)] **sluice**
b. I know you saw someone, but I don't know [who (it) (was)] **RCC**

Hindi-Urdu does in fact have a limited cleft strategy, and as in English the pivot of the cleft can be a *wh*-phrase (in (16)).[6]

(16) kyaa hai jo mez=kii daayii taraf hai?
what be.PRS.3SG REL table.F=GEN.F. right side.F be.PRS.3SG
'What is it that is to the right of the table?'

That said, Hindi-Urdu does not generally permit the copula to be dropped — an operation that we might expect to exist independently if apparent Hindi-Urdu sluices were actually RCCs. As (17)–(19) show, the copula is required except in the presence of negation.

(17) siitaa mer-ii dost *(hai/thii/hogii).
Sita.F 1GEN-F.SG friend be.PRS.3SG/be.PAST.F.SG/be.FUT.3SG.F
'Sita is/was/will be my friend.'

(18) tum kis=ke saath *(ho)?
you who.OBL=GEN.OBL with be.PRS.2SG
'Who are you with?' (Koul 2008)

(19) siitaa mer-ii dost nahĩĩ (he).
Sita.F 1GEN-F.SG friend NEG be.PRS.3SG
'Sita is not my friend.'

[6] Though some in the literature have claimed that Hindi-Urdu lacks clefts (Malhotra 2009) their properties have been addressed in eye-tracking studies (see Vasishth et al. 2012), and naturally occurring examples with *wh*-pivots are relatively easy to find/overhear. For instance:

(i) kaun hai jo aap=ke dil=kii awaaz sun-taa hai?
who be.PRS.3SG REL 2PL=GEN.M.OBL heart.M=GEN.F.SG voice.F hear-HAB.M.SG be.PRS.3SG
'Who is it that hears the voice of your heart?'
(http://qna.rediff.com/questions-and-answers/kaun-he-jo-apke-dil-ke-awaz-sunta-hai/18247630/answers accessed 4/25/12)

Given these facts, the basic operations necessary to form an RCC are not independently present in Hindi-Urdu.

Further, properties of apparent sluices in Hindi-Urdu and properties of RCCs diverge. In RCCs the *wh*-pivot is typically nominative or unmarked (Merchant 2001, van Craenenbroeck 2009, Lasnik 2007), but as we have seen above apparent sluices in Hindi-Urdu require full case connectivity (in (3)–(4)). Sluicing with adjunct *wh*-phrases is grammatical in Hindi-Urdu, but clefting with adjunct *wh*-pivots is not (unlike with arguments, as in (22)).

(20) us=ne gaaṛii=ko fiks ki-yaa,
3SG.OBL=ERG car.F.SG=ACC fix do-PRF.M.SG
magar mujhe nahĩĩ pa-taa kese (*thaa).
but 1SG.DAT NEG know-HAB.M.SG how (be.PST.M.SG)
'He fixed the car, but I don't know how (*it was).' (e.g. with what tool)

(21) subhan ali aa-yaa, magar mujhe nahĩĩ pa-taa kyõõ (*thaa).
Subhan Ali come-PRF.M.SG, but 1SG.DAT NEG know-HAB.M.SG why (be.PST.M.SG)
'Subhan Ali came, but I don't know why (*it was)'

(22) us=ne koi gaaṛii fiks kii,
3SG.OBL=ERG some car.F.SG fix do.PRF.F.SG
magar mujhe nahĩĩ pa-taa kaunsii (thii).
but 1SG.DAT NEG know-HAB.M.SG which (be.PST.F.SG)
'He fixed some car, but I don't know which one (it was).'

For clefts with *wh*-pivots, only an exhaustive reading is available (23B). On the other hand, sluices are compatible with a 'mention-some' non-exhaustive interpretation (23B′).

(23) A: aap=ko kisi ofisar=se baat kar-nii caahiiye.
2PL=DAT some.OBL officer=with talk do-INF.F.SG want
'You should speak with an officer.'

B: #kaun hai, masail=ke tor par?
who be.PRS.3.SG example=GEN.OBL manner on
'Who is it, for example?'

B′: masail=ke tor par, kis=se?
example=GEN.OBL manner on who.OBL=with
'For example, who?'

If RCCs are an instance of what Hankamer & Sag (1976) call deep anaphora, then material that seems to be missing should be recoverable pragmatically, not necessarily under linguistic identity. In Hindi-Urdu, as in English, sluicing appears to be surface anaphora, requiring a linguistic antecedent.

(24) [Shown a picture of an unknown woman]

a. #I don't know who.

b. I don't know who she is.

c. #mujhe nahĩĩ pa-taa kaun.
1SG.DAT NEG know-HAB.M.SG who
'I don't know who.'

d. mujhe nahĩĩ pa-taa kaun hai.
1SG.DAT NEG know-HAB.M.SG who be.PRS.3SG
'I don't know who it is.'

The data in (17)–(24) suggest that apparent sluicing structures in Hindi-Urdu are not reduced copular clauses or clefts of any kind, but instead have some other derivation.

2.4 Sluicing in Hindi-Urdu is not stripping (Hankamer 1979)

Hankamer (2010) proposes that putative instances of sluicing in Turkish can be analyzed as stripping, an ellipsis in which all constituents but one of a second conjunct go missing (Hankamer 1979, Merchant 2003), as in the English example in (25).

(25) Amit left for Delhi, and Jamal too.

First, stripping is not possible in embedded contexts (unless the antecedent clause too is embedded) as in the English example in (26), but apparent sluicing in Hindi-Urdu can be embedded, as in (27).

(26) *Amit left for Delhi, and I know Jamal too.

(27) amit kahĩĩ ga-yaa, aur mujhe lag-taa hai
Amit somewhere go-PFV.M.SG and 1SG.OBL strike-HAB.M.SG be.PRS.3SG
ki mãĩ jaan-tii hũ kahãã.
that 1SG.NOM know-HAB.F.SG be.PRS.1SG where
'Amit went somewhere, and it seems to me that I know where.'

Second, stripping cannot precede its antecedent (backward anaphora), as in (28). Sluicing in Hindi-Urdu, on the other hand, can.

(28) *Jamal too, and Amit left for Delhi.

(29) mujhe nahĩĩ pataa kahãã, lekin mãĩ jaan-tii hũ ki amit
1SG.OBL NEG know where but 1SG.NOM know-PFV.F.SG be.PRES.1SG that Amit.M
kahĩĩ ga-yaa hai.
somewhere go-PFV.M.SG be.PRS.3SG
'I don't know where, but I know Amit went somewhere'.

Therefore it seems that sluicing-like structures in Hindi-Urdu are not likely to be instances of stripping.

2.5 Sluicing in Hindi-Urdu is not fed by movement to a high focus projection

Toosarvandani (2009) claims that sluicing in Persian is fed by movement to a high focus projection (above TP). There is evidence that this position is independently active in Persian for contrastive focus (Karimi 1999, 2003):

(30) [faghat be kimea]$_i$ man t_i se tâ ketab dâd-am PERSIAN
only to Kimea 1SG three PART book gave-1SG
'It was only to Kimea that I gave three books.'

Similarly, we see high positional focus in languages like Italian and Gungbe, below (see Cinque 1990, Zubizarreta 2010).

(31) Qualcosa, di sicuro, io farò. (Cinque 1990:15) ITALIAN
SOMETHING, surely, I do.FUT
'Surely, I do something'.

(32) Mótò w? Dòsú kù wá. (Aboh 2007:84) GUNGBE
car FOC Dosu drive come
'Dosu came by CAR'.

However, previous work (Butt & King 1996, Kidwai 1999, 2000) suggests that the unmarked position for both interrogative and non-interrogative focus in Hindi-Urdu is low, immediately preceding the clause-final verb.

(33) mãĩ=ne kamre=mẽ [in=hii tiin larkõ=ko] bhej-aa.
1SG=ERG room.M.OBL=in [these=only three boy.PL.OBL=ACC] sent-PFV.M.SG
'I sent *these* three boys to the room.' (Butt & King 1996)

(34) kitabẽ kal mãĩ laa-yaa thaa.
book.F.PL yesterday 1SG bring-PFV.M.SG be.PST.M.SG
'*I* brought the books yesterday (It is *I* who brought the books yesterday)' (Kidwai 2000)

(35) kitaab kis=ne dekh-ii
book.F.SG who.OBL=ERG see-PFV.F.SG
'Who saw the book?'

Since we have established above that apparent sluicing in Hindi-Urdu cannot be the ellipsis of a constituent smaller than TP, then the movement that feeds sluicing is unlikely to be typical movement for focus.

Could it instead be scrambling which feeds sluicing-like ellipsis in Hindi-Urdu?[7] Though the term scrambling can refer to a range of optional displacements in Hindi-Urdu with differing characteristics (Mahajan 1990, 1994, Kidwai 2000), we can show that the movement that precedes apparent sluicing is not scrambling either. The *wh*-word *kyaa* 'what', resists scrambling and in general is most felicitous in the preverbal position (as in (36)).

(36) a. aap abhii kyaa kar-te hãĩ?
2PL now what do-HAB.M.PL be.PRS.PL
Now what are you doing?

b. #kyaa aap abhi kar-te hãĩ?

In apparent sluices, however, *kyaa* is a completely felicitous remnant *wh*-word (in (37)).

(37) mãĩ=ne yahãã kuch dekh-aa par mujhe nahĩĩ pataa kyaa
1SG=ERG here something see-PFV.M.SG but 1SG.DAT NEG know what
'I saw something there, but I don't know what.'

If the movement feeding apparent sluicing was scrambling, we might expect *kyaa* to be a less acceptable *wh*-remnant, contrary to fact. While there certainly appears to be some kind of displacement from the base position in Hindi-Urdu sluicing-like structures, it is unlikely that this displacement is either movement to a focus position or scrambling.

2.6 Summary

The data in this section collectively suggests that putative sluices in Hindi-Urdu feature a displaced *wh*-remnant at the clause edge and involve elision of a clause-sized constituent.[8] The simplest conclusion is that apparent sluicing structures in Hindi-Urdu are just that: sluicing structures. But there must be something exceptional about them, because there is no (visible) regular process of *wh*-movement to the clause edge in the language. In what follows, I propose that sentences like (1) do indeed feature genuine sluicing, and that what is exceptional in their derivation is the copy of the *wh*-element that is pronounced.

[7]Thanks to Veneeta Dayal and Rajesh Bhatt for bringing this question and this data to my attention.

[8]A reviewer points out that it is important to mention the role of the finite clause subordinator *ki* in this proposal, as *ki* can optionally precede a sluice as in (i):

(i) mãĩ=ne yahãã kisi=ko dekh-aa par mujhe nahĩĩ pataa ki kis=ko.
1SG=ERG there someone.OBL=ACC see-PFV.M.SG but 1SG.DAT NEG know that who.OBL=ACC
'I saw someone there, but I don't know who.'

This would be surprising if *ki* were a complementizer located in C, and might seem to indicate that sluicing is not fed by *wh*-movement to Spec, CP in Hindi-Urdu. However, I have pursued elsewhere (Manetta 2006, 2011) that the invariant particle *ki* is not in fact a complementizer in the true sense but instead an optional marker of the phase boundary, not located in C. It is not selected for by the verb (appearing optionally before any type of embedded clause including interrogative complements), is transparent to selection, and has no particular semantic content. Further, it cannot appear before a preposed clause. Also, *ki* can co-occur in an embedded clause with the optional yes/no question marker *kyaa* also argued to be in C. For these reasons I will not consider *ki* to be located in C and will not address it further here.

3 "*wh*-in-situ" is lower copy pronunciation

3.1 Support for *wh*-movement in Hindi-Urdu

Among languages that exhibit *wh*-in-situ characteristics, recent work has explored the possibility that the copy privileged for phonological realization might be the lowest copy in a *wh*-chain (Groat & O'Neil 1996, Reintges, Lesourd, & Chung 2006, Reintges 2007, i.a.).

There are two types of support for the claim that an apparent *wh*-in-situ language actually has syntactic *wh*-movement that is concealed by lower copy pronunciation. In general, we should look for such a language to exhibit some features of "overt" *wh*-movement. One type of support comes from the morphological evidence known as *wh*-agreement, in which the morphology of complementizers or verb forms indicates that *wh*-movement has taken place (Reintges, LeSourd, & Chung 2006, Reintges 2007). Unfortunately, Hindi-Urdu does not feature this kind of morphology (though see the speculative discussion of *wh*-expletives below).

Another type of support for this analysis comes from the most basic diagnostics for movement. If a seemingly in-situ *wh*-construction actually exhibits properties associated with movement, we might hypothesize that movement has in fact taken place, but the higher copy of the *wh*-chain goes unpronounced. Among traditional tests for A-bar movement is the presence of Weak Crossover (WCO) effects, arising when a *wh*-chain and a pronoun are co-indexed and the tail of the *wh*-chain fails to c-command the pronoun. The WCO properties of Hindi-Urdu are well known (Mahajan 1990, Dayal 1994, Kidwai 2000), as they are of particular interest in accounts of scrambling in the language, so I will provide only a sketch of the basic facts here.

A pronominal coindexed with a lower *wh*-phrase is strongly dispreferred, as in (38). If that *wh*-phrase is displaced to a position preceding the subject, we see an obviation of WCO effects (compare the grammaticality of (39) with the unacceptability of its English translation). Mahajan (1990) takes this to mean that displacement (scrambling) of this type must therefore be A-movement (though see Dayal 1994, Kidwai 2000 for complexities). When a *wh*-phrase is displaced into a higher clause as in (40) (unambiguously A-bar movement in Hindi-Urdu), the expected WCO effects appear.

(38) ???[us=kii$_i$ behin]=ne [kis laṛke]=ko$_i$ dekh-aa.
3SG.OBL=GEN.F.SG sister.F.SG=ERG which.OBL boy.M.SG.OBL=ACC see-PFV.M.SG
'Which boy$_i$ did his$_i$ sister see?'

(39) [kis laṛke]=ko$_i$ [us=kii$_i$ behin]=ne t_i dekh-aa.
who.OBL boy.M.SG.OBL=ACC 3SG.OBL=GEN.F.SG sister.F.SG=ERG see-PFV.M.SG
'Which boy$_i$ did his$_i$ sister see?'

(40) *[kaunsaa/har aadmii]$_i$ [[us=kii$_i$ behin]=ne soc-aa
which/every man.M 3SG.OBL=GEN.F.SG sister.F.sg=ERG think-PFV.M.SG
ki ram=ne t_i dekh-aa]?
that Ram.M=ERG see-PFV.M.SG
'Which/every man$_i$ did his$_i$ sister think that Ram saw?' (Bhatt 2003:11–19)

Unfortunately this picture has little to tell us about the approach to *wh*-movement in Hindi-Urdu proposed here. If we assume that (38) involves true *wh*-movement in the narrow syntax to Spec, CP, followed by lower copy pronunciation, we would expect WCO effects to arise as a result of the configuration schematized in (41), giving rise to the ungrammaticality of (38).

(41) [CP ~~which boy~~$_i$ [his$_i$ sister [vP which boy$_i$ see]]]

Then again, if we assume that Hindi-Urdu is a *wh*-in-situ language, or a language with A-bar movement to a position below the subject, then we would also anticipate the ungrammaticality of (38), since the binder does not c-command the bound variable. In general, Hindi-Urdu is a language in which linear precedence determines binding possibilities (Bhatt & Dayal 2007, Manetta 2012).

(39) also fails to provide straightforward information about the present account. This configuration cannot simply represent the pronunciation of the higher copy in the *wh*-chain, which would be

unexpected under our approach, given that there is nothing to force pronunciation of the dispreferred copy here. Instead, we would need to understand (39) as a context in which the copy pronounced is the result of further scrambling subsequent to *wh*-movement. Under a theory of scrambling and binding like that articulated in Kidwai (2000), scrambling as XP-adjunction renders the copy of 'which boy' at the top of the *wh*-chain immediately dominating the subject pronominal ineligible to serve as a local binder (as it itself is a bound variable).[9] As Kidwai points out, the same facts hold in the case in which the direct object is a quantificational phrase (e.g. *har laṛke=ko* 'each boy'), suggesting that the acceptability of (39) tells us nothing in particular about *wh*-movement. For these reasons, WCO facts cannot really inform the analysis proposed here.

Let us now turn to island effects, which prove to be a more useful diagnostic. Hindi-Urdu exhibits the full range of island sensitivities, as in languages with overt movement (Malhotra 2011:6, 86, Bhattacharya & Simpson 2012).

Complex NP Island

(42) *[kyaa$_i$ ravii=ko [$_{DP}$ yeh baat [$_{CP}$ ki miiraa t_i khaa-yegii]] pataa hai]?
what Ravi.M=DAT this fact that Mira.F eat-FUT.F.3SG know be.PRS.3SG
'What does Ravi know the fact that Mira will eat?' (Malhotra 2009:35) ***wh*-extraction**

(43) *[raam=ne kyaa kah-aa [ki ravii=ko [yeh baat [ki miiraa kyaa khaa-yegii] pataa hai]]]?
Ram.M=ERG EXPL say-PFV.M.SG that Ravi.M=DAT this fact that Mira.F what eat-FUT.F.3SG know be.PRS.3SG
'What did Ram say that Ravi knows the fact that Mira will eat?'
(Malhotra 2009:32–33) ***wh*-expletive construction**

(44) *raam=ko ye baat [ki siitaa kis=se mil-ii] pataa hai?
Ram.M=DAT that fact that Sita.F who.OBL=with meet-PFV.F.SG know be.PRS.3SG
'Who does Ram know the claim that Sita met? ***wh*-in-situ**

Adjunct Island

(45) *raam=ne kyaa$_i$ kah-aa [ki siitaa bazaar jaa-yegii [kyunki mohan t_i nahĩĩ laa-yaa]]?
Ram.M=ERG what say-PFV.M.SG that Sita.F market go-FUT.F.3SG because Mohan.M NEG bring-PFV.M.SG
'What did Ram say that Sita will go to the market because Mohan didn't bring?'
***wh*-extraction**

(46) *raam=ne kyaa kah-aa [ki siitaaa bazaar jaa-yegii [kyunki mohan kyaa nahĩĩ laa-yaa]]?
Ram.M=ERG EXPL say-PFV.M.SG that Sita.F market go-FUT.F.3SG because Mohan.M what NEG bring-PFV.M.SG
'What did Ram say that Sita will go to the market because Mohan didn't bring?'
(Malhotra 2009:32–33) ***wh*-expletive construction**

Relative Clause Island

(47) *raam=ko kyaa$_i$ [$_{DP}$ vo laṛkaa [$_{CP}$ jo t_i laa-yaa]] pasand hai
Ram.M=DAT what DEM.3SG boy.M.SG REL buy-PFV.M.SG liking be.PRS.3SG
'What does Ram like the boy that bought?' (Malhotra 2009:58) ***wh*-extraction**

[9]Space does not permit an in-depth exploration of the details of Kidwai's account of scrambling in these cases in which expected WCO does not arise (though see Kidwai 2000:124–138). The account of *wh*-movement presented here does not necessarily depend on the particulars of any single approach to scrambling in Hindi-Urdu.

***Wh*-island**

(48) *raam kaunsaa kamraa$_i$ pataa kar rahaa hai
Ram.M which room.M.SG know do PROG.M.SG be.PRS.3SG
ki kaunsii laṛkii t$_i$ kiraaye=par le-gii?
that which girl.F.SG rent.OBL=LOC take-FUT.F.3SG
'Which room will Ram find out which girl will rent?' ***wh*-extraction**

(49) *raam=ne kis=ko puch-aa ki kyaa miiraa=ne *t* dekh-aa?
Ram.M=ERG who.OBL=ACC ask-PFV.M.SG that what Mira.F=ERG see-PFV.M.SG
'Who did Ram ask whether Mira saw?' (Malhotra 2009:78) ***wh*-extraction**

(50) *raam jaan-naa caah-taa hai agar miiraa=ne kyaa kharid-aa?
Ram know-INF.M.SG want-HAB.M.SG be.PRS.3SG if Mira.F=ERG what buy-PFV.M.SG
'What does Ram want to know whether Mira bought?' ***wh*-in-situ**

Further, Hindi-Urdu seems to allow parasitic gaps (Mahajan 1994:317–323; see also Bošković 2002, Lin 2005). For instance, in (51) the parasitic gap (PG) is licensed by the *wh*-phrase *kaunsi kitaab* 'which book', that appears to be sitting in its base-generated pre-verbal position.

(51) ali=ne [paṛh-ne=se pehle] kaunsi kitaab phẽk d-ii?
Ali.M=ERG read-INF.M.OBL=with before which book.F.SG throw give-PFV.F.SG
'Which book did Ali throw away before reading?'

At first glance, the status of (51) as a PG is not completely clear, because as Davison (1999) and Bhatt (2003) have pointed out, these gaps seem to be possible in the absence of movement altogether, as in (52). The gap labeled *e* in (52) is best understood as *pro*.

(52) ram=ne [binaa *e*$_i$ paṛhe] [vo kitaab]$_i$ phẽk d-ii
Ram.M=ERG without reading DEM.3SG book.F.SG throw give-PFV.F.SG
'Ram threw that book away without reading (it).'

However, in Manetta (2013), I claim that Hindi-Urdu does indeed exhibit true PGs (I summarize that argumentation here, but for more detail see Manetta 2013). Following Abe & Nakao (2009) and Abe (2011) for Japanese, I suggest that the *pro* strategy is certainly available for Hindi-Urdu, but that real PGs are also present when the *pro* strategy is unavailable.

For instance, PGs into which reconstruction must apply cannot be easily understood as an instance of *pro*. In (53)–(54) the reflexive can be bound by *har laṛke=ne* 'each boy'.

(53) [kaunsi apn-ii$_i$ tasveer=ko]$_j$ har laṛke=ne [binaa *e*$_i$ dekhe] kah-aa ki
which self.F.SG picture.F.SG=ACC each boy.M.OBL=ERG without seeing say-PFV that
miriam=ne *t*$_j$ pasand ki-yaa
Miriam.F=ERG liking do-PFV.M.SG
'Which picture of himself did each boy, without seeing, say that Miriam liked?'

(54) [kaunsi apn-ii tasveer=ko$_i$ [har laṛke=ne [jis=ne *e*$_{?i}$ dekh-aa]]]
which self.F.SG picture.F.SG=ACC each boy.M.OBL=ERG REL.OBL=ERG see-PFV.M.SG
kah-aa ki miriam=ne *t*$_i$ pasand ki-yaa?
say-PFV.M.SG that Miriam.F=ERG liking do-PFV.M.SG
'Which picture of himself did each boy who saw say that Miriam liked?'

In fact, the reflexive cannot be bound by *Miriam* in (53)–(54). If this is so, the *wh*-phrase containing the reflexive must be interpreted as though it were reconstructed into the gap preceding the verb *dekhe* 'seeing', not into the position of the trace preceding *pasand kiyaa* 'like'. It seems that the alleged PG in (53)–(54) can then not be understood as a *pro*. In Manetta (2013) I also show that since configurations like those in (53) and (54) must contain a real PG, other properties of real PGs hold, such as case matching. Careful testing therefore demonstrates that *wh*-structures in Hindi-Urdu do license parasitic gaps.

The diagnostics in this section on the whole suggest that *wh*-movement has indeed taken place in apparent *wh*-in-situ constructions in Hindi-Urdu. Importantly, they do not yet allow us to determine that *wh*-movement targets Spec, CP in the narrow syntax, as the data above is consistent with an analysis in which *wh*-movement targets the edge of the vP layer (as in Manetta 2006, 2011). Theoretically there are then two viable alternatives: in line with my previous work, we could assume that *wh*-movement takes place to Spec, vP (or an equivalent position) regularly in Hindi-Urdu in the narrow syntax but that clause-wide scope is obtained by a mechanism other than syntactic movement (for instance, via an Agree relation valuing features on the interrogative C, or via additional movement operations designed to achieve scope at LF). In this view, the syntactic movement required to produce genuine sluicing in Hindi-Urdu must be understood as exceptional. On the other hand, we could pursue an account in which an additional mechanism to obtain clause-wide scope is not required, and sluicing is not fed by exceptional movement. In that analysis, which I lay out in the remainder of this paper, Hindi-Urdu has regular *wh*-movement to Spec, CP in the narrow syntax, but a lower copy in the *wh*-movement chain is typically pronounced. This second approach is not only consistent with the facts presented in this section, but is more analytically parsimonious, as it posits a single mechanism driving overt displacement within the grammar.

The analysis of a typical *wh*-question in Hindi-Urdu being proposed here is schematized in (55).

(55) [$_{CP}$ ~~kis=ko~~ [aap=ne yahãã kis=ko dekh-aa]]
who=ACC 2PL=ERG here who.OBL=ACC see-PFV.M.SG
'Who did you see here?'

In (55), the *wh*-phrase *kis=ko* 'who' undergoes syntactic *wh*-movement to the specifier of CP from its base-generated position as the complement to the verb *dekhaa* 'see'. However, the upper copy in Spec, CP goes unpronounced at PF (as indicated by the strikethrough). Instead it is the lower copy that is pronounced, giving rise to a structure that resembles *wh*-in-situ.

Importantly, this type of lower copy pronunciation is quite different from that proposed to exist in languages like Romanian or Bulgarian, in that is not exceptional or as a last resort (Bošković 2002, Nunes 2004). Instead, Hindi-Urdu exhibits a language-wide preference for lower-copy pronunciation in *wh*-chains, as in Coptic Egyptian (Reingtes 2007). As we will see below, what is exceptional in Hindi-Urdu is pronunciation of higher copies, as in the case of sluicing.

3.2 Which copy?

As (33)–(35) above illustrate, Hindi-Urdu is in fact a so-called "*wh*-focus" language, like Hungarian or Turkish, in the sense that there is a dedicated unmarked position for both interrogative and non-interrogative focus. This fact complicates the question of which copy is being pronounced.

Previous approaches to *wh*-dependencies in Hindi-Urdu have attempted to capture this property of Hindi-Urdu by claiming that there is indeed a regular process of *wh*-movement in the language, not to Spec, CP, but to a lower (preverbal) position, Spec, vP (Manetta 2006, 2010, Malhotra & Chandra 2007). This movement is sometimes concealed if the *wh*-word is the direct object (in which case *wh*-movement to Spec, vP is string vacuous), or if further scrambling of other constituents for information-structural purposes alters the surface order. If these accounts are on the right track and the criterial position for *wh*-material in Hindi-Urdu is Spec, vP, then under a copy theoretic approach it would not be the bottom-most copy that would be preferentially phonetically realized but an intermediate copy.

Others have claimed that there are languages in which intermediate copies can be pronounced. For instance, Fanselow and Ćavar (2001) analyze data from Bahasa Indonesia in which they claim that an intermediate copy may be realized in the specifier of an embedded CP:

(56) ~~Siapa~~ Bill tahu [siapa yang Tom cintai ~~siapa~~] BAHASA INDONESIA
~~who~~ Bill knows [who FOC Tom loves ~~who~~]
'Who does Bill know Tom loves?'

In this view, Hindi-Urdu would share with other *wh*-in-situ languages a dispreference for phonetic realization of the topmost copy in a *wh*-movement chain (the one in Spec, CP). It would instead favor pronunciation of the copy located in the Spec, vP of an interrogative clause as in (57).

(57) [CP ~~kis=ko~~ aap=ne yahãã [vP kis=ko ~~kis=ko~~ dekh-aa]]
~~who.OBL=ACC~~ 2PL=ERG here who.OBL=ACC ~~who.OBL=ACC~~ see-PFV.M.SG
'Who did you see here?'

We will see below that such an assumption may also be useful in understanding long-distance *wh*-dependencies in Hindi-Urdu, as matrix question interpretation can only arise if phonetically overt *wh*-material is present in the preverbal position in the matrix clause.

One could envision a system of optimized constraints at work to encode the language's preference for phonetic realization of copies in Spec, vP, which is overridden in exceptional scenarios such as in sluicing configurations under pressure from more highly ranked requirements. Just such a system is developed in Fanselow & Ćavar (2001). That said, I will leave a detailed elaboration of this system in this particular case to future work, and turn now to questions of long-distance *wh*-dependencies and top-copy sluicing.

3.3 Long-distance *wh*-dependencies

Scope in Hindi-Urdu is clause-bound. In order to take matrix scope out of an embedded clause, *wh*-phrases must either appear displaced into the clause over which they take scope (though not to its edge) as in (58a), or the *wh*-expletive *kyaa* must be used in the preverbal position as in (58b).

(58) a. sita=ne **kis=ko** soc-aa ki ravii=ne ___ dekh-aa?
Sita.F=ERG who.OBL=ACC think-PFV.M.SG that Ravi.M=ERG see-PFV.M.SG
'Who did Sita think that Ravi saw?'

b. sita=ne **kyaa** soc-aa ki ravii=ne **kis=ko** dekh-aa?
Sita.F=ERG EXPL think-PFV.M.SG that Ravi=ERG who.OBL=ACC see-PFV.M.SG
'Who did Sita think that Ravi saw?'

This data presents at least two questions for the theory of *wh*-in-situ in Hindi-Urdu as lower copy pronunciation. First, why should it be the case that in instances in which the *wh*-phrase takes scope in a clause higher than its own, either the *wh*-phrase or an expletive must appear in that higher clause? Second, what precisely is the *wh*-expletive construction? Is (58b) a sentence in which *wh*-movement into the higher clause has taken place, or not (independent of how the copies are phonetically realized)?

Manetta (2010) answers the first question by claiming that there is a syntactic requirement of v heads in interrogative clauses (that is, clauses at which an embedded *wh*-phrase will take scope) that overt interrogative material appear there. This is encoded syntactically in the form of an EPP feature on v in the scoping clause. Under the present proposal, this requirement would need to be stated in another way, and in fact in another component of the grammar. Since I have posited here an account in which *wh*-movement to the criterial position always takes place in Hindi-Urdu, it is the copy that is ultimately pronounced that is at issue. We would require a constraint favoring overt phonetic realization of *wh*-content in the Spec, vP of interrogative clauses. The formalization of RECOVERABILITY proposed in section 4 below suggests that interrogative Spec, vP is a position associated with special phonetic content and therefore requires that the member of a movement chain appearing there be pronounced. Such a constraint is easily satisfied by pronunciation of the *wh*-copy in Spec, vP in single interrogative clauses, as we discussed above. Further, an ungrammatical version of (58a) in which the matrix clause had interrogative features (in other words, in which the embedded *wh*-phrase should take matrix scope) but the lower copy of the *wh*-phrase (in the embedded clause) was pronounced would violate this constraint and would be ruled out.

Turning to (58b), the question is whether this surface form represents one in which any *wh*-movement into the matrix clause has taken place. In Manetta (2010), the answer is no; the *wh*-expletive *kyaa* serves to satisfy the EPP on the matrix v just in the case that the embedded *wh*-

phrase does not move. On the other hand, the present account gives us a set of tools to look at (58b) somewhat differently. Could *wh*-movement have taken place, as usual, in the narrow syntax, but resulting in the realization of multiple copies? If so, *kyaa* would need to be understood as an alternate pronunciation of a higher copy in the *wh*-chain, as Hindi-Urdu does not exhibit canonical multiple copy realization as in German. Following a particular proposal in Nunes (2004), this alternate pronunciation of the higher *wh*-phrase as *kyaa* could be the result of fusion of an interrogative head (v according to Manetta 2010) and the moved *wh*-word. Another analytical approach to the minimal *wh*-word that serves as a *wh*-expletive in Hindi-Urdu might be very similar to Fanselow's (2001) treatment of resumptive pronouns in movement chains.[10]

There is one piece of (as yet unexplained) evidence that this view of *kyaa* might be important to pursue further. Hindi-Urdu *wh*-expletive structures seem to exhibit island effects, as in (43)–(46) above, and here in (59) (Malhotra & Chandra 2007, Malhotra 2011).

(59) a. *[raam=ne kyaa kah-aa [ki ravii=ko [yeh baat [ki miiraa kyaa
Ram.M=ERG EXPL say-PFV.M.SG that Ravi.M=ACC this fact that Mira.F what
khaa-yegii] pataa hai]]]?
eat-FUT.F.3SG know be.PRES.3SG
'What did Ram say that Ravi knows the fact that Mira will eat?'

b. *raam=ne kyaa kahaa [ki siitaa bazaar ja-yegii [kyunki mohan
Ram.M=ERG EXPL say-PFV.M.SG that Sita.F market go-FUT.F.3SG because Mohan.M
kyaa nahĩĩ lay-aa]]?
what NEG bring-PFV.M.SG
'What did Ram say that Sita will go to the market because Mohan didn't bring?'

The ungrammaticality of the structures in (59) suggests that *wh*-movement must have taken place, triggering island violations. Under the tentative approach to *wh*-expletive constructions explored in this section, the full version of the lower copy is pronounced, while a modified/minimal version of the matrix clause copy is realized, in the form of the *wh*-expletive *kyaa*.[11]

While the issues explored in this section do not bear directly on the question of sluicing in Hindi-Urdu, they do seem to provide support for a copy-theoretic approach to *wh*-in-situ in the language

[10]Fanselow & Ćavar 2001 (footnote 8) do not consider *wh*-expletive constructions instances of "true partial *wh*-movement", instead reserving this term for configurations in languages like Bahasa Indonesia in which the *wh*-phrase is pronounced in an intermediate position that is neither its scope position nor its base-generated position. It seems from the discussion here that Hindi-Urdu may well be just such a language, however, with unmarked *wh*-material appearing in Spec, vP.

[11] Unexpected under the copy-theoretic account advanced here is the fact that *wh*-in-situ structures in Hindi-Urdu exhibit intervention effects ((ia) and (iia)), while structures with overt *wh*-displacement across the offending quantificational element do not ((ib) and (iib)) (Malhotra 2011:92–93).

(i) a. *raam=hii kis=ko dekh-egaa?
Ram.M=only who.OBL=ACC see-FUT.M.3.SG
'Who will only Ram see?'

b. kis=ko raam=hii t dekh-egaa?
who.OBL=ACC Ram.M=only see-FUT.M.3SG
'Who will only Ram see?'

(ii) a. *raam=ne kyaa kah-aa ki siitaa=ne=hii kis=ko maar-aa?
Ram.M=ERG EXPL say-PFV.M.SG that Sita.F=ERG=only who.OBL=ACC kill-PFV.M.SG
'Who did Ram say that only Sita killed?'

b. kis=ko raam=ne kah-aa ki siitaa=ne=hii t maar-aa?
who.OBL=ACC Ram.M=ERG say-PFV.M.SG that Sita.F=ERG=only t kill-PFV.M.SG
'Who did Ram say that only Sita killed?'

This contrast in *wh*-in-situ languages has historically been understood in terms of LF movement (Beck 1996, Pesetsky 2000); it is the required LF movement of the *wh*-phrase over the quantificational element that causes ungrammaticality. The issue of how intervention effects are best treated in a single cycle model in which there is no LF/covert movement is beyond the scope of this work. However we would want such an account to capture the empirical observation that a language may make a distinction between displacement with PF effects and without (in contrast to a *wh*-in-situ language like Coptic Egyptian, which does not exhibit any intervention effects (Reintges 2007)).

in general. It also seems clear that adopting this account would require a rethinking of various widely-accepted analyses of *wh*-phenomena in Hindi-Urdu, but that this rethinking might result in increased empirical coverage and solutions to unresolved puzzles. I leave the remainder this effort to future work and return now to the account of sluicing.

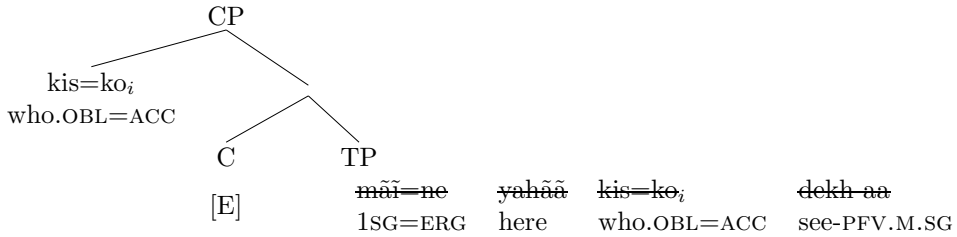
4 Top-copy sluicing

The core assumption of top-copy sluicing, following Franks (1998) (see also Bošković & Nunes 2007, Reintges 2007, Bošković 2011), is that in a given language the pronunciation of a particular copy in a *wh*-chain at PF is a matter of preference, which can be overridden if pronunciation in the preferred position leads to a PF violation.

Under this account, a sluicing structure in Hindi-Urdu is a marked instance in which the lower copy[12] cannot be pronounced, as it resides in a TP marked for non-pronunciation due to the [E] feature on C (Merchant 2001).

(60) a. I saw someone there, but I don't know...

b.



Given this scenario, if the top copy is also not pronounced, the sluiced structure will violate a constraint like RECOVERABILITY that requires that at least one copy of a lexical item to be pronounced.

The primary challenge remaining for the account is to clearly define the mechanisms governing RECOVERABILITY. That is, precisely how do we prevent a scenario in which no copy of a *wh*-chain is phonetically realized? While the reasoning behind recoverability is fairly intuitive, its formalization is not trivial. In particular formulations such as that in Pesetsky (1998) are problematic in that the realization of a copy at PF is contingent on information related to its interpretation — information presumably inaccessible in the PF component.

(61) RECOVERABILITY (Pesetsky 1998)
A syntactic unit with semantic content must be pronounced unless it has a sufficiently local antecedent.

Here I will pursue an alternative formalization in which chain resolution is an exclusively PF process (as in Franks 1999, Bošković 2001, Bobaljik 2002), using the principle of P-RECOVERABILITY developed in Landau (2006).

(62) P-RECOVERABILITY
In a chain $< X_1, \ldots X_i \ldots X_n >$, where some X_i is associated with phonetic content, X_i must be pronounced.

As Landau points out, P-RECOVERABILITY is a principle that places a lower bound on what must be pronounced in a chain (at least one copy). The upper bound is enforced by an economy condition, preventing all copies in a chain from being pronounced.

(63) Economy of Pronunciation
Delete all chain copies at PF up to P-RECOVERABILITY

[12] In what follows, for simplicity I will continue to refer to the preferred copy for pronunciation in unmarked interrogatives in Hindi-Urdu as the "lower" copy, even though it may in fact be an intermediate copy as discussed in section 3.2 above. As the intermediate copy in Spec, vP would be contained within any TP marked for non-pronunciation (sluiced TP), the distinction is not crucial for this portion of the analysis.

Certainly, P-RECOVERABILITY must always override Economy of Pronunciation, indicating that at least P-RECOVERABILITY is an overarching principle not subject to re-ranking as in an optimal theoretic system. Together, these two principles ensure that at least one copy of a *wh*-chain will be realized.

To complete the definition of P-RECOVERABILITY, Landau defines associated with phonetic content as follows:

(64) X is associated with phonetic content iff:
  a. X has phonetic content, or
  b. X is in a position specified with some phonological requirement.

Landau suggests that in the case of V(P)-topicalization in Hebrew, it is the spellout of tense and agreement features and the intonation required for topicalized VPs that require the pronunciation of the two realized copies of the verb. Sturgeon (2008) claims that the intermediate copy in Czech left dislocation is pronounced due to the need to realize associated phonological rise (see also Roberts 2010). Similarly, in the case of a typical *wh*-question in Hindi-Urdu, one could attribute the phonetic realization of the preverbal copy to the need to pronounce the focal stress that appears on immediately pre-verbal *wh*-phrase (Kidwai 2000). It is this requirement to pronounce the member of the chain associated with phonetic content that constitutes the preference in Hindi-Urdu for the pronunciation of the pre-verbal copy in the *wh*-chain, and it is the Economy Condition in (63) that is responsible for the simultaneous non-realization of the top copy.

On the other hand, in a sluicing structure in Hindi-Urdu (the schematic of which is repeated below in (65)), the preverbal copy will no longer be associated with additional phonetic content (focal stress) due to PF-deletion of TP. P-RECOVERABILITY now forces the realization of the phonetic content associated with the *wh*-word for at least one copy of the chain. The only copy now available is that in Spec, CP, resulting in a genuine sluice.

(65) I saw someone there, but I don't know...
  a. ...~~kis=ko mãĩ=ne yahãã kis=ko dekh aa~~ *P-RECOVERABLE
  b. ... kis=ko ~~mãĩ=ne~~ ~~yahãã~~ ~~kis=ko~~ ~~dekh aa~~ SLUICE
    who.OBL=ACC ~~1SG=ERG~~ ~~here~~ ~~who.OBL=ACC~~ ~~see-PFV.M.SG~~
    '...who ~~I saw here~~'.

The phonological deletion mechanism that results in a sluice (TP-ellipsis) is independent of the process of chain formation and pronunciation (Landau 2006). P-RECOVERABILITY and Economy of Pronunciation as they are formulated here operate only over chains and therefore cannot force the pronunciation of a segment otherwise designated for non-pronunciation (i.e. undo a sluice or VP-ellipsis). If, indeed, there were no members of the chain outside of the non-pronounced segment that could be realized, the result would be ungrammatical, violating the principle of P-RECOVERABILITY.

Under this account, then, sluicing structures in Hindi-Urdu are, in fact, genuine sluices like those familiar from languages like English. There is full *wh*-movement to the clause edge in the narrow syntax. The C head possesses a feature that calls for non-pronunciation of its TP complement. The only difference between English and Hindi-Urdu is then the manner by which the higher copy in the *wh*-chain comes to be pronounced. In English, this is a matter of course, since English prefers the highest copy in a *wh*-chain to be phonetically realized. In Hindi-Urdu, it is an exception, forced when the copy preferred for phonetic realization, the lower copy, is in a clause already marked for non-pronunciation. The higher copy must then be pronounced to avoid losing phonetic realization of the *wh*-chain altogether.

This analysis then correctly predicts that Hindi-Urdu sluiced structures have properties quite similar to genuine sluices in languages like English, in sharp contrast to other *wh*-in-situ languages which seem to employ other strategies to derive sluicing-like strings (see discussion below of e.g. Gribanova 2011 on the use of the RCC strategy in Uzbek). Properties such as full case connectivity and post-position pied-piping find explanation in the present account since real syntactic *wh*-movement

to Spec, CP does seem to take place. Similarly, it is unsurprising that material in the Tense head is elided in a sluice since a full TP goes unpronounced as in more familiar languages.

Finally, one slightly more controversial property of Hindi-Urdu — the potential for island violation repair — is explained under this approach since it posits full *wh*-movement in the syntax. This property has become emblematic of genuine sluicing and for that reason is important to discuss here. Insofar as island violations can be repaired under sluicing (Malhotra 2011:35; c.f. Bhattacharya & Simpson 2012),[13] we could suggest that the problematic copy, the one inside the island, goes unpronounced in a sluice. The sentence in (66a) shows extraction out of a complex NP island, resulting in ungrammaticality. However, the sluiced version in (66b), leaving behind the *wh*-remnant *kyaa* 'what', is acceptable.

(66) a. *kyaa ravii=ko [DP yeh baat ki miiraa t khaa-yegii] pataa hai
what Ravi.M=DAT this fact that Mira.F eat-FUT.F.3SG know be.PRS.3SG
'What does Ravi know the fact that Mira will eat?'

b. raviii=ko [DP yeh baat ki miiraa kuch khaa-yegii] pataa hai
Ravi.M=DAT this fact that Mira.F something eat-FUT.F.3G know be.PRS.3SG
par mãĩ nahĩĩ jaan-taa kyaa [...]
but 1SG NEG know-HAB.M.SG what
'Ravi knows the fact that Mira will eat something, but I don't know what.'

5 Conclusions

The primary goal of the analysis presented here is to capture the empirical particulars of sluicing-like constructions in Hindi-Urdu. Adopting the copy theory of movement provides the framework necessary to make two linked claims: (a) normal *wh*-questions in Hindi-Urdu are instances of lower copy pronunciation (a language-specific preference for phonetic realization of a *wh*-chain), and (b) sluices are exceptional instances of top-copy pronunciation. Taken together, these claims offer an explanation for a number of properties of sluicing, as well as other interesting puzzles concerning *wh*-dependencies in the language.

More broadly, this article pursues continued refinement in the implementation of copy theory in *wh*-in-situ languages. The line of research spurred by Groat & O'Neil (1996), Nunes (2004), and Reintges, LeSourd & Chung (2006) (among others) has made important gains in untangling the empirical puzzles presented by *wh*-in-situ languages and the realization of *wh*-chains. The preliminary account of Hindi-Urdu pursued here has suggested that Hindi-Urdu is of the family of *wh*-in-situ languages that typically prefers a lower copy in a *wh*-chain to be pronounced, but requires this preference to be overridden in cases in which scoping pressures or P-RECOVERABILITY demand it.

The present account also represents the early stages of a larger project investigating intra-linguistic variation among the (sometimes radically) different presentations of *wh*-in-situ (Gribanova & Manetta 2013). This project is concerned with the way that certain constellations of properties of *wh*-dependencies and ellipsis processes in *wh*-in-situ languages are best understood. Certainly research up to this point has demonstrated that status as a "*wh*-in-situ" language alone does not predict whether or what kind of sluicing is available in a language (e.g. Takahashi 1994, Ince 2006, Kizu 1997, 2000, Toosarvandani 2009, Gribanova 2011).[14] Even languages with many overlapping *wh*-in-situ properties may not behave precisely the same with respect to sluicing processes — as in, for instance, Hindi-Urdu and Bangla (Bhattacharya & Simpson 2012). Gribanova and Manetta ask if some Indic languages behave like Hindi-Urdu in allowing a structure that looks like genuine sluicing while some Turkic languages seem to use RCC-like strategies (see, for instance, Hankamer

[13]The potential for island violation repair is still controversial in Hindi-Urdu, and though some efforts are currently being made to perform the more delicate empirical research needed to make a clear case (see Malhotra 2010), further work needs to be done. The claim I make is that if indeed island violations are repaired under sluicing, the copy theoretic approach pursued here provides an explanation for these facts.

[14]Conversely, some languages exhibiting overt *wh*-movement have been claimed to exhibit RCC-like strategies to form sluice-like structures (Vincente 2008, van Craenenbroeck 2010)

2010, Gribanova 2011) — from what formal properties of these languages does this split follow? We hope that further investigation can reveal and provide a framework with which to analyze the interface between patterns of *wh*-dependency formation and ellipsis.

Acknowledgements

I am grateful to Vera Gribanova for many useful conversations developing the ideas this paper, and for comments on various versions. Thanks to Bob Freidin for the encouraging us to think about chains and copy realization. I also thank Rajesh Bhatt, Veneeta Dayal, Masha Polinsky, and Alice Davison for their comments, questions, and suggestions. Finally, I am grateful to the attentive audience at Formal Approaches to South Asian Languages 2 at MIT, where this paper was first presented.

References

Abe, Jun. 2011. Real parasitic gaps in Japanese. *Journal of East Asian Linguistics* 20(3):195–218.

Abe, Jun and Chizuro Nakao. 2009. On ATB-movement and parasitic gaps in Japanese. In S.-W. Kim, ed., *Visions of the Minimalist Program: Proceedings of the 11th Seoul International Conference on Generative Grammar*, pages 1–19. Hankuk Publishing.

Aboh, Enoch Oladé. 2007. Leftward focus versus rightward focus: the Kwa-Bantu conspiracy. In *SOAS Working Papers in Linguistics*, vol. 15, pages 81–104.

Beck, Sigrid. 1996. *Wh-constructions and Transparent Logical Form*. Ph.D. thesis, Universität Tübingen. Available at http://www.sfs.uni-tuebingen.de.

Bell, Taylor Serota. 2012. The (in)visibility of adjunction hosts in the syntax. Masters Thesis, University of California, Santa Cruz.

Bhatia, Archna. 2006. Testing Cinque's hierarchy: Adverb placement in Hindi. In *Proceedings of the Workshop in General Linguistics 2006*, vol. 6, pages 10–25. Linguistics Student Association, University of Wisconsin, Madison.

Bhatt, Rajesh. 2003. Topics in the syntax of modern Indo-Aryan languages. Course handouts.

Bhatt, Rajesh. 2005. Long distance agreement in Hindi-Urdu. *Natural Language and Linguistic Theory* 23:757–807.

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

Bhattacharya, Tanmoy, M. T. Hanybabu, and Kalyanamalini Sahoo. 2011. Auxiliaries as heavily grammaticalized light verbs: Evidence from four South Asian languages. In *Proceedings of Western Conference on Linguistics*, vol. 2, pages 113–125.

Bhattacharya, Tanmoy and Andrew Simpson. 2012. Sluicing in Indo-Aryan: An investigation of Bangla and Hindi. In J. Merchant and A. Simpson, eds., *Sluicing: Cross-Linguisitic Perspectives*, pages 183–219. Oxford: Oxford University Press.

Bobaljik, Jonathan David. 2002. A-chains at the PF-interface: Copies and 'covert' movement. *Natural Language and Linguistic Theory* 20(2):197–267.

Bošković, Željko. 2011. Rescue by PF deletion, traces as (non)interveners, and the that-trace effect. *Linguistic Inquiry* 42(1):1–44.

Bošković, Željko and Jairo Nunes. 2007. The copy theory of movement: A view from PF. In N. Corver and J. Nunes, eds., *The Copy Theory of Movement*, pages 13–74. Amsterdam: John Benjamins.

Butt, Miriam and Tracy H. King. 1996. Structural topic and focus without movement. In *Proceedings of the First LFG Conference*. CSLI Publications. XRCE, Grenoble.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In K. Hale and S. J. Keyser, eds., *The View from Building 20*, pages 1–52. Cambridge, MA: The MIT Press.

Chomsky, Noam and Howard Lasnik. 1993. The theory of principles and parameters. In A. von Stechow, J. Jacobs, W. Sternefeld, and T. Vennemann, eds., *Syntax: An International Handbook of Contemporary Research*, pages 506–569. Berlin: De Gruyter.

Chris, Reintges, Phillip LeSourd, and Sandra Chung. 2006. Movement, wh-agreement, and apparent *wh*-in-situ. In L. L.-S. Cheng and N. Corver, eds., *Wh-Movement Moving On*, pages 165–194. Cambridge, MA: The MIT Press.

Cinque, Guglielmo. 1990. *Types of A-bar-Dependencies*. Cambridge, MA: The MIT Press.

Davison, Alice. 1999. Empty categories: anaphoric vs. pronominal coindexing. University of Iowa.

Davison, Alice. 2002. Perfective aspect as underspecified past tense. Paper presented at the South Asian Language Analysis Conference, June 2002.

Dayal, Veneeta. 1994. Binding facts in Hindi and the scrambling phenomenon. In M. Butt, T. H. King, and G. Ramchand, eds., *Theoretical Perspectives on Word Order Issues in South Asian Languages*, pages 237–261. Stanford: CSLI Publications.

Dwivedi, Veena. 1991. Negation as a functional projection in Hindi. In K. Hunt, T. Perry, and V. Samiian, eds., *Proceedings of the Western Conference on Linguistics (WECOL 91)*, vol. 4, pages 88–100.

Fanselow, Gisbert. 2001. Features, θ-roles, and free constituent order. *Linguistic Inquiry* 32(3):405–437.

Fanselow, Gisbert and Damir Ćavar. 2001. Remarks on the economy of pronunciation. In G. Müller and W. Sternefeld, eds., *Competition in Syntax*, pages 107–150. Berlin: Mouton de Gruyter.

Franks, Stephen. 1998. Clitics in Slavic. Position paper for Comparative Slavic Morphosyntax workshop, held at McCormick's Creek State Park, 5–7 June.

Goldberg, Lotus. 2005. *Verb-stranding VP Ellipsis: a cross-linguistic study*. Ph.D. thesis, McGill University.

Gribanova, Vera. 2011. Copular clauses, clefts, and putative sluicing in Uzbek. Stanford University.

Gribanova, Vera. 2013. A new argument for verb-stranding verb phrase ellipsis. *Linguistic Inquiry* 44(1):145–157.

Gribanova, Vera and Emily Manetta. 2013. Ellipsis in *wh*-in-situ languages: Deriving apparent sluicing in Hindi-Urdu and Uzbek. Stanford University and University of Vermont.

Groat, Erich and John O'Neil. 1996. Spell-out at the LF interface. In W. Abraham, S. Epstein, H. Thrainsson, and C. J.-W. Zwart, eds., *Minimal Ideas*, pages 113–139. Amsterdam: John Benjamins.

Hankamer, Jorge. 1979. *Deletion in Coordinate Structures*. New York: Garland Publishing.

Hankamer, Jorge. 2010. Pseudosluicing in Turkish. Presented at the 29th West Coast Conference on Formal Linguistics (WCCFL).

Hankamer, Jorge and Ivan A. Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7(3):391–426.

Horn, Lawrence. 1978. Remarks on neg-raising. In P. Cole, ed., *Pragmatics*, pages 129–220. New York: Academic Press. Syntax and Semantics 9.

Ince, Atakan. 2006. Pseudo-sluicing in Turkish. In N. Kazanina, U. Minai, P. Monahan, and H. Taylor, eds., *University of Maryland Working Papers in Linguistics*, vol. 14, pages 111–126.

Karimi, Simin. 1999. A note on parasitic gaps and specificity. *Linguistic Inquiry* 30(4):704–713.

Karimi, Simin. 2003. Focus movement and uninterpretable features. In A. Carnie, H. Harley, and M. Willie, eds., *Formal Approaches to Function in Grammar. Festschrift for Eloise Jelinek*, pages 297–306. Amsterdam: John Benjamin.

Kidwai, Ayesha. 1999. Word order and focus positions in Universal Grammar. In G. Rebuschi and L. Tuller, eds., *The Grammar of Focus*, pages 213–245. Amsterdam: John Benjamins.

Kidwai, Ayesha. 2000. *XP-Adjunction in Universal Grammar: Scrambling and Binding in Hindi-Urdu*. Oxford: Oxford University Press.

Kizu, Mika. 1997. Sluicing in *wh*-in-situ languages. In K. Singer, R. Eggert, and G. Anderson, eds., *Proceedings of the Chicago Linguistic Society (CLS)*, vol. 33, pages 231–244.

Kizu, Mika. 2000. A note on sluicing in *wh*-in-situ languages. In *MIT Working Papers in Linguistics*, vol. 36, pages 143–159.

Koul, Omkar N. 2008. *Modern Hindi Grammar*. Hyattsville: Dunwoody Press.

Kumar, Rajesh. 2006. *Negation and licensing of negative polarity items in Hindi syntax*. New York: Taylor and Francis.

Kush, Dave. 2011. Height-relative determination of (non-root) modal flavor: Evidence from Hindi. In *Proceedings of Semantics and Linguistic Theory (SALT)*, vol. 21, pages 413–425.

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

Lasnik, Howard. 2007. On ellipsis: The Pf approach to missing constituents. In A. Conroy, C. Nakao, and E. Takahashi, eds., *Maryland Working Papers in Linguistics 15*, pages 143–153. University of Maryland.

Lin, Jonah. 2005. Does *wh*-in-situ license parasitic gaps? *Linguistic Inquiry* 36(2):298–302.

Mahajan, Anoop. 1990. *The A/A-Bar Distinction and Movement Theory*. Ph.D. thesis, MIT.

Mahajan, Anoop. 1994. The ergativity parameter: have-be alteration, word order and split ergativity. In M. Gonzalez, ed., *Proceedings of the North Eastern Linguistic Society (NELS)*, pages 317–331.

Malhotra, Shiti. 2009. Quantifer induced barriers and wh-movement. In *Proceedings of the 32nd Annual Penn Linguistics Colloquium*, vol. 15, pages 139–145.

Malhotra, Shiti. 2010. Sluicing is PF deletion: Evidence from Hindi/Urdu. Presented at West Coast Conference on Linguistics (WECOL), Fresno State, CA.

Malhotra, Shiti. 2011. *Movement and Intervention Effects: Evidence from Hindi/Urdu*. Ph.D. thesis, University of Maryland, College Park.

Malhotra, Shiti and Pritha Chandra. 2007. A short note on wh-scope marking. In *University of Maryland Working Papers*, vol. 16. University of Maryland.

Manetta, Emily. 2006. *Peripheries in Kashmiri and Hindi-Urdu*. Ph.D. thesis, University of California, Santa Cruz.

Manetta, Emily. 2010. Wh-expletives in Hind-Urdu: The vP phase. *Linguistic Inquiry* 41(1):43–74.

Manetta, Emily. 2011. *Peripheries in Kashmiri and Hindi-Urdu: The Syntax of Discourse-Driven Movement*. Amsterdam: John Benjamins.

Manetta, Emily. 2012. Reconsidering rightward scrambling: Postverbal constituents in Hindi-Urdu. *Linguistic Inquiry* 43(1):73–74.

Manetta, Emily. 2013. Signs of movement: Parasitic gaps in Hindi-Urdu. Presented at Formal Approaches to South Asian Languages III, University of Southern California, March 9, 2013.

Merchant, Jason. 1998. 'Pseudosluicing': Elliptical clefts in Japanese and English. In A. Alexiadou, N. Fuhrhop, P. Law, and U. Kleinhenz, eds., *ZAS Papers in Linguistics 10*, pages 88–112. Berlin: Zentrum für Allgemeine Sprachwissenschaft.

Merchant, Jason. 2001. *The Syntax of Silence: Sluicing, Islands, and the Theory of Ellipsis*. Oxford: Oxford University Press.

Merchant, Jason. 2003. Remarks on stripping. University of Chicago.

Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*. Cambridge, MA: The MIT Press.

Pesetsky, David. 1998. Some optimality principles of sentence pronunciation. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, and D. Pesetsky, eds., *Is the Best Good Enough?*, pages 337–383. Cambridge, MA: The MIT Press.

Pesetsky, David. 2000. *Phrasal Movement and its Kin*. Cambridge, MA: The MIT Press.

Reintges, Chris. 2007. Variable pronunciation sites and types of *wh*-in-situ. In J. Nunes and N. Corver, eds., *The Copy Theory of Movement*, pages 249–287. Amsterdam: John Benjamins.

Roberts, Ian. 2010. *Agreement and Head Movement: Clitics, Incorporation, and Defective Goals*. Cambridge, MA: The MIT Press.

Ross, John. 1969. Guess Who? In R. Binnick, A. Davidson, G. Green, and J. Morgan, eds., *Papers from the 5th Regional Meeting of the Chicago Linguistic Society (CLS)*, pages 252–286.

Sailor, Craig. 2014. VP ellipsis in Tag questions: A typological approach. In *Proceedings of the 46th Annual Meeting of the Chicago Linguistic Society (CLS 46)*. Chicago: University of Chicago.

Simpson, Andrew, Arunima Choudhury, and Mythili Menon. 2012. Argument ellipsis and the licensing of covert nominals in Bangla, Hindi and Malayalam. University of Southern California.

Srivastav (Dayal), Veneeta. 1991. *Wh-dependencies in Hindi and the Theory of Grammar*. Ph.D. thesis, Cornell University.

Sturgeon, Anne. 2008. *The Left Periphery: The Interaction of Syntax, Pragmatics and Prosody in Czech*. Amsterdam: John Benjamins.

Takahashi, Daiko. 1994. Sluicing in Japanese. *Journal of East Asian Linguistics* 3:265–300.

Toosarvandani, Maziar. 2009. Ellipsis in Farsi complex predicates. *Syntax* 12(1):60–92.

Van Craenenbroeck, Jeroen. 2009. What sluicing can do, what it cant and where. on the cross-linguistic syntax of ellipsis. In *Diagnosing syntax: perspectives, procedures and tools*. Leiden & Utrecht, January 2009.

van Craenenbroeck, Jereon. 2010. Invisible last resort: A note on clefts as the underlying source for sluicing. *Lingua* 120(7):1714–1726.

van Craenenbroeck, Jeroen and Anikó Lipták. 2013. What sluicing can do, what it can't, and in which language. In L. L.-S.Cheng and N. Corver, eds., *Diagnosing Syntax*, pages 502–536. Oxford: Oxford University Press.

Vasishth, Shravan, Rukshin Shaher, and Narayanan Srinivasan. 2012. The role of clefting, word order and given-new ordering in sentence comprehension: Evidence from Hindi. *Journal of South Asian Linguistics* 5:35–56.

Vicente, Luis. 2008. Syntactic isomorphism and non-isomorphism under ellipsis. http://ling.auf.net/-lingBuzz/000776.

Zubizarreta, Maria-Luisa. 2010. The syntax and prosody of focus: The Bantu-Syntax connection. *Iberia: An International Journal of Theoretical Linguistics* 2(1):131–168.

The Principle A Problem

Pritty Patel-Grosz, *University of Tübingen*

Received June 2013; Revised December 2013

Abstract

The presence or absence of Principle A of the Binding Theory can be explained by a requirement at the syntax semantics interface. Principle A effects surface in the presence of unmodified *self*, where there is a requirement for the subject and object to be identical; this requirement triggers *self*-incorporation into the predicate. In contrast, modified *self*-incorporation is blocked in the syntax, giving rise to an asymmetric part-of relation. Given that *self*-incorporation is absent, we predict that Principle A effects do not surface, and this is exactly what we find.

1 Introduction

1.1 The Puzzle

This paper explores the possibility of a uniform approach for the examples in (1) and (2). The English data in (1) illustrate that anaphors are sensitive to Principle A of the Binding Theory, Chomsky (1981), meaning that they require a local antecedent. It is debatable whether Principle A should be expected to apply to the examples in (2).[1] One could argue that in (2), *his pathetic self* is a noun phrase where *self* is a noun, therefore subject to Condition C; whereas in (1), *himself* is an anaphor and thus sensitive to Principle A.

(1) a. John saw **himself** (in the mirror) *Sensitive to Principle A*
b. *John knows that Maria saw **himself** (on the reality show)

(2) a. John saw **his** pathetic **self** (in the mirror) *Not sensitive to Principle A*
b. John knows that Maria saw **his** pathetic **self** (on the reality show)

Under such a view, *self* in English is lexically ambiguous,[2] i.e. there are two lexical entries for *self*. I will refer to the two constructions as UNMODIFIED SELF, (1), and MODIFIED SELF, (2). Assuming lexical ambiguity implies that unmodified *self* in (1) and the presence of Principle A can be explained by traditional Binding Theory, whereas in the modified *self* construction in (2), *self* refers to some abstract aspect of the referent of *his*. It appears however, that the two occurrences of *self* in (1) and

[1] Natural occurrences of 'his pathetic self' (from Google):
(i) Let's watch his pathetic self tread water for six hours.
(ii) Anyone who saw his pathetic self on his reality show knows that.
(iii) His ex wife im sure is so happy that she left his sorry pathetic loser self.

[2] Of course, treating English *self* as lexically ambiguous presupposes an analysis where *himself* is the spell-out of the DP *his self*. Alternatively, *himself* and *his ... self* may be classed as distinct items altogether.

JSAL Volume 6, December 2013.

(2) share a core meaning, namely some notion of IDENTITY; therefore, a uniform approach is worth pursuing.

Further support for a uniform approach comes from languages such as Greek (Greek) and Kutchi Gujarati (KGu) (an Indo-Aryan language). In English, lexical ambiguity is motivated by the fact that *himself* is morphologically distinct from *his self*, i.e. the form of the anaphor in (1) is not *his self*. This distinction however, does not carry over to other languages. Unlike English, anaphora in Greek and Kutchi Gujarati are morphologically complex. In these languages it is not immediately obvious that *self* is lexically ambiguous, because both the unmodified *self* and the modified *self* forms look identical apart from the presence or absence of the adjective (cf. (3b) and (4b) vs. (5b) and (6b)). The Greek and Kutchi Gujarati data in (3) and (4), respectively, parallel the English examples in (1). In all three languages the unmodified *self* must be locally bound and is sensitive to Principle A.

(3) a. O Costas$_i$ vlepi [**ton** **eafto$_i$** **tu**]. Greek
Costas$_i$ sees det.M.SG self$_i$.M.SG 3.M.SG.GEN
'Costas$_i$ sees himself$_i$.'

b. *O Costas$_i$ xeri oti Maria vlepi [**ton** **eafto$_i$** **tu**]. Greek
Costas$_i$ knows that Maria sees det.M.SG self.M.SG 3.M.SG.GEN
'Costas$_i$ knows that Maria sees himself$_i$.'

(4) a. john$_i$ [**e-na** **potha$_i$-ne**] jo-yo. KGu
John$_i$ 3.SG-GEN self$_i$-ACC see-PFV.M.SG
'John$_i$ saw himself$_i$.'

b. *john$_i$ kidthu ke maria [**e-na** **potha$_i$-ne**] jo-yo. KGu
John$_i$ said that Maria 3.SG-GEN self$_i$-ACC see-PFV.M.SG
'John$_i$ said that Maria saw himself$_i$.'

Identical to the English modified *self* example in (2), in Greek (5b) and Kutchi Gujarati (6b) the presence of an adjective alters the acceptability of the utterance. While (5a)+(6a) with unmodified *self* are ungrammatical, presumably due to Principle A, (5b)+(6b) with modified *self* are grammatical, and have a reading where the *mother* loves the *true part* of the referent (*Jannis* in (5b) and *Valji* in (6b)). Unlike English modified vs. unmodified *self*, the Greek and Kutchi Gujarati examples in (3a)+(5b) and (4a)+(6b), respectively, show that the forms of modified *self* and unmodified *self* are identical (*eafto* in Greek, and *potha-ne* in Kutchi Gujarati).

(5) a. *[I mitera tu Janni$_i$] agapai [ton eafto$_i$ tu]. Greek
3.F.SG.NOM mother Janni$_i$.GEN loves 3.M.SG.ACC self$_i$ 3.M.SG.GEN
'Jannis$_i$'s mother loves his self$_i$.'

b. [I mitera tu Janni$_i$] agapai [ton **alithino** eafto$_i$ tu]. Greek
3.F.SG.NOM mother Janni$_i$.GEN loves 3.M.SG.ACC true self$_i$ 3.M.SG.GEN
'Jannis$_i$'s mother loves his true self$_i$.'

(6) a. *valji$_i$-ni ma [e-na potha$_i$-ne] prem kar-e. KGu
Valji$_i$-GEN.F mother 3.SG-GEN self$_i$-ACC love do-3.SG
'Valji$_i$'s mother loves his self$_i$.'

b. valji$_i$-ni ma [e-na **sacha** potha$_i$-ne] prem kar-e. KGu
Valji$_i$-GEN.F mother 3.SG-GEN true self$_i$-ACC love do-3.SG
'Valji$_i$'s mother loves his true self$_i$.'

There is a question at this point whether modified *self* may be a type of logophor, i.e. a pronominal element that refers to the 'source of information' (Büring 2005:62) for the environment that it occurs in. It can be easily shown that this is not the case. In the Kutchi Gujarati example in (7), the embedded clause 'she saw his true self' is embedded under a verb of saying, *kidthu* 'said' with *Mary* as its subject. If *ena sacha pothane* 'his true self' was a logophor in this example, then its only

possible antecedent would be *Mary*, as the source of information. However, *eno sacho pothane* can refer to *John's true self*, which shows that it is not a logophor.

(7) john$_i$ avyo. mary kidthu ke i [**e-na sacha potha$_i$-ne**] jo-yo. KGu
John came Mary said that 3.SG 3.SG-GEN true self$_i$-ACC see-PFV.M.SG
'John$_i$ came. Mary said that she saw his true self$_i$.'

Given the facts discussed in this section, it seems worth pursuing a view where there is simply one, non-lexically ambiguous *self* (at least for Greek and Kutchi Gujarati). Such an approach raises the following question: if unmodified *self* in Greek and Kutchi Gujarati is sensitive to Principle A, why does this effect vanish in the case of modified *self*? I will argue that the Principle A effects do not arise directly from the semantics of *self* (in the sense that there are two lexical entries), but rather from the structural configuration and syntactic environment where *self* is situated, in combination with interface requirements.

1.2 Overview of the Solution

In order to explain the peculiar distribution of unmodified and modified *self* with respect to Principle A, I propose that the lexical entry for *self* in these languages denotes a relation R, which I assume is equivalent to something along the lines of PART-OF (defined in section 2.2.2). I argue that this part-of relation can pick out different aspects of an individual e.g., the *evil side* of someone's personality as well as the *good side* of that person's personality.

Furthermore, I argue that the difference between modified and unmodified *self* arises due to a requirement at the syntax-semantics interface; unmodified *self* would be too unconstrained if it simply expresses the part-of relation (because there are potentially infinitely many aspects of an individual[3]).

(8) *unconstrained meaning of self as* part-of *(to be discarded)*
John saw **himself** (in the mirror).
≈ John saw the unique salient part of John (in the mirror).

This requirement (which militates against (8)) is implemented by assuming that unmodified *self* differs from modified *self* due to predicate incorporation of *self* in the former, (9), but not in the latter, (10). While the part-of relation is asymmetric, *self*-incorporation gives rise to a symmetric relation, as *self* is interpreted both in its base position and in its landing position, thus reversing the arguments; this constrains the interpretation of unmodified *self*. In contrast, modified *self*-configurations are cases where the anaphor cannot incorporate (due to independent constraints in the syntax), allowing the relation expressed by *self* to remain asymmetric. Furthermore, I will show that *self*-incorporation not only yields a symmetric relation, but also requires identity of the subject and object, triggering Principle A effects. In contrast, when incorporation is blocked (the cases of modified *self*), the anaphor is exempt from Principle A.

(9) *unmodified* self
John saw **himself** (in the mirror).
⟹ John **self**-saw **himself** (in the mirror).
≈ John is a part of the unique salient part of John and John saw the unique salient part of John (in the mirror).
= John saw John (in the mirror).

(10) *modified* self
John saw **his pathetic self**.
≈ John saw the unique salient entity that is pathetic and a part of John.

[3]I will show that an approach where default interpretation of 'the unique salient part' as 'the part that corresponds to the whole' cannot be the correct one.

2 The Semantics of *self*

In this section, I propose a uniform semantics for modified and unmodified *self*. In section 2.1, I present a simplified version of the lexical entry for *self*, which reduces the relation between the arguments to identity. Although this explains unmodified *self*, section 2.2 shows that the lexical entry is too conservative and fails to account for the empirical scope regarding modified *self*. I propose an alternative lexical entry that relaxes the identity relation replacing it with the part-of relation in section 2.3. I illustrate that such an approach can uniformly account for both modified and unmodified *self*.

2.1 The Semantics of Unmodified *self*

The utterances in (11a) and (11b) intuitively correspond to a meaning similar to *Costas admires Costas* and *Valji saw Valji*, respectively. On a par with Iatridou (1988), and Anagnostopoulou & Everaert (1999), I assume that *eafto* 'self' (and *potha* 'self') is the head of a complex DP containing a true determiner *ton* 'the' (which takes the shape of the differential object marker *ne* in Kutchi Gujarati), as well as a bound genitive pronoun *tu* 'his' (and *ena* 'his'). To derive the correct truth conditions, I propose that the counterparts of *self* in Greek and Kutchi Gujarati are relational nouns, and that *self* denotes identity between its arguments.[4] Under such an analysis of *self* as expressing the identity relation, *eafto*'s first argument in (11a) is the bound genitive pronoun *tu*, and its second argument is bound by the iota operator introduced by the determiner *ton*. The resulting meaning is *Costas admires the unique individual identical to Costas*, which is trivially synonymous with *Costas admires Costas*.

(11) a. O Costas$_2$ thavmazi [$_{DP}$ **ton** [**eafto** **tu**$_2$]]. Greek
Costas admires det.M.SG self.M.SG 3.M.SG.GEN
'Costas admires himself.'

b. valji$_2$ [e-na$_2$ potha-ne] jo-yo. KGu
Valji 3.SG-GEN self-ACC see-PFV.M.SG
'Valji saw himself.'

Formally speaking, given that *self* in these languages seems to express identity, we can assume the preliminary lexical entry given in (12): A function that takes two arguments and yields the truth value 1 iff the two arguments are identical.

(12) *the semantics of 'eafto/potha (self)'* (preliminary)[5]
||eafto/potha|| = [λx.λy.y = x]

To derive the syntactic distribution of Principle A, I argue that unmodified *self* obligatorily incorporates into the matrix verb (see (14b) for an illustration), unless blocked by independent syntactic constraints.[6] For the purposes of the present section, it suffices to state the following. In order to derive the correct truth conditions for the utterance, I assume that both copies of *self* are interpreted. Fox (1999, 2002) proposes that higher copies are interpreted as operators and lower copies as variables. Given his explanation of Principle C effects in quantifier raising (QR), his findings suggest that we have reason to believe that copies are interpreted in every merge position modulo semantic convergence.[7] In order to interpret the higher copy of *self*, we need a new predicate modification rule, given in (13). In order for *eafto/potha* (being of type <e,<e,t>>) to combine with

[4]See Saxon (1984), Lubowicz (1999), and Gast (2006) for the idea that *self*-forms generally express identity functions.

[5]This analysis of Kutchi Gujarati *potha* and Greek *eafto* as relational nouns that denote the identity function was first proposed in Patel (2010:6); the adaptation of predicate modification for relational predicates was first given in Patel (2010:13). For Greek *eaftos*, the same proposal was made in parallel by Spathas (2010:161-162).

[6]See Anagnostopoulou & Everaert (1999), Reuland (2001, 2005, 2011) and Reuland & Winter (2009) for predecessors of this approach.

[7]Cf. For Fox (1999, 2002), interpreting both copies is more economical than interpreting one. There does not seem to be a principled reason why we should not be able to interpret both copies as they are, if this yields a well-formed interpretation.

a transitive verb (also of type <e,<e,t>>), I assume that predicate modification can be extended (from combining two functions of type <e,t> to combining two functions of type <e,<e,t>>). Put differently, predicate modification for relational predicates contains sets of ordered pairs, and is a relation that holds between two things.

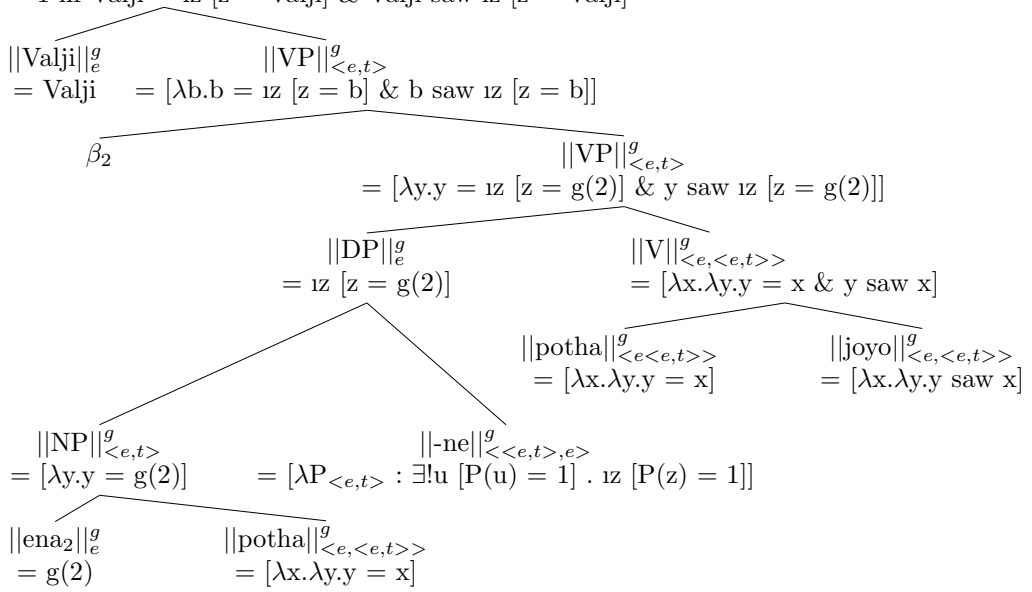
(13) *Predicate Modification for Relational Predicates*
For any β and γ, which are functions of type <e,<e,t>>, and assignment g,
$|| \beta \gamma ||^g = [\lambda x.\lambda y.\ || \beta ||^g(x)(y) = 1 \ \& \ || \gamma ||^g(x)(y) = 1]$.
(based on Heim & Kratzer 1998:95)

Compositionally, we can now derive the meaning of a construction containing a reflexive DP; this is illustrated in (14) for Kutchi Gujarati.[8] Note that for ease of exposition, I use Büring's (2005) β operator,[9] which does not require movement of the binder.

(14) a. valji$_2$ [e-na$_2$ potha-ne] jo-yo. KGu
Valji 3.SG-GEN self-ACC see-PFV.M.SG
'Valji saw himself.'

b. LF: Valji$_2$ [ena$_2$ potha-ne] *potha*-joyo.

c. $||S||^g_t$
= 1 iff Valji = ɪz [z = Valji] & Valji saw ɪz [z = Valji]



d. *In words*: Valji equals the unique individual (in the utterance context) that equals Valji and Valji saw the unique individual that is equal to Valji.

The syntax-semantic analysis sketched above is appealing for various reasons. First, having shown how to derive the truth conditions for sentences with *self*-incorporation that satisfy Principle A, we can now turn to examples that violate Principle A, illustrated in (15).

(15) a. *O Costas$_i$ xeri oti Maria vlepi [**ton** **eafto**$_i$ **tu**]. Greek
Costas$_i$ knows that Maria sees det.M.SG self.M.SG 3.M.SG.GEN
'Costas$_i$ knows that Maria sees himself$_i$.'

[8]The semantic analysis proposed here also accounts for the equivalent Greek data, and derives identical truth conditions to the Kutchi Gujarati case. In this example, I analyze the differential object marker *-ne* as an element that contributes the meaning of the definite determiner; however, this is a simplification, since *-ne* is known to interact with various factors, including animacy, definiteness and specificity, cf. Butt and Ahmed (2011), Mistry (1997).

[9]Büring's (2005:85) Binder Index Evaluation Rule (BIER) is defined as follows:

(i) For any natural number n, $|| \beta_n \ Y||^g = \lambda b.||Y||^{g[b/n]}(b)$.

b. *john$_i$ kidthu ke maria [**e-na** **potha$_i$-ne**] jo-yo. KGu
John$_i$ said that Maria 3.SG-GEN self$_i$-ACC see-PFV.M.SG
'John$_i$ said that Maria saw himself$_i$.'

The above analysis derives the following truth conditions given in (16) for these examples.

(16) a. LF: ||O Costas$_i$ xeri oti Maria **eafto**-vlepi [**ton eafto$_i$ tu**]|| Greek
'Costas$_i$ knows that Maria self-sees himself$_i$'
= 1 iff Costas knows that [Maria equals the unique individual that equals Costas and Maria sees the unique individual that is equal to Costas].

b. LF: || John$_i$ kidthu ke Maria [**e-na potha$_i$-ne**] **potha**-joyo || KGu
'John$_i$ said that Maria self-saw himself$_i$'
= 1 iff John said that [Maria equals the unique individual that equals John and Maria saw the unique individual that is equal to John].

Since it is part of the truth conditions that *Maria* is identical to whoever the reflexive refers to, these statements will be false whenever *Maria* is not identical to the referent of *himself* (given that the identity of *Maria* and *himself* is part of the truth conditions). The ill-formedness of these examples then arises from the fact that the intended reading is one where *Costas/John* is the referent of *himself*. *Costas/John* and *Maria* can only both be co-referent with *himself* if *Costas/John* and *Maria* refer to the same person (i.e. *Costas/John* = *Maria*).

One of the consequences of the analysis, which must be addressed, is why modified *self* is exempt from Principle A; this is the topic of section 2.2. We address unmodified *self* in section 2.3.

2.2 The Semantics of Modified *self*

2.2.1 Problems with Identity

It follows from the proposal in the previous section, that given identity, a subject and object must be identical whenever the object contains *self*, and *self* is incorporated into the predicate. However, it can be shown that once *self* is modified by an adjective, treating the anaphors in Greek and Kutchi Gujarati as identity relations will not do.

Consider the following scenario. I use English for simplicity, though the same argument applies to Kutchi Gujarati and Greek. Assume that there are two sides to John's personality. John has an *attractive self* and an *ugly self*. I can now say *John admires his attractive self and fears his ugly self*, and the use of this utterance in such a way avoids any form of contradiction. If the anaphors in question were to require the subject and object to be identical, it is not clear how this could be implemented for this scenario; how could *John's attractive self* be identical to *John*, when *John* also encompasses an *ugly self*?

Or, what is more worrying: If *John's attractive self* was identical to *John*, and *John's ugly self* was identical to *John*, then *John's attractive self* would be identical to (and indistinguishable from) *John's ugly self*. This is clearly an undesirable result. For instance, we can make a statement like (17a), to give an extreme case. Let us assume that a modified *self* construction does not involve *self*-incorporation, given that modified *self* is not subject to Principle A (I will derive this from independent syntactic constraints in section 3). Assuming (which might be a simplification here,[10] see also section 2.2.2) that *attractive* and *ugly* are intersective adjectives, we would derive truth conditions as in (17b). Given the nature of identity, the only individual that is equal to *John* is *John himself*. The truth conditions in (17b) thus wrongly predict that (17a) is equivalent to (18a), (18b) and (18c), which in turn should all be equivalent.

(17) a. [John's attractive self] fears [his ugly self]

b. ||(17a)|| = 1 iff [the unique individual that is attractive and equals John] fears [the unique individual that is ugly and equals John]

[10]This simplification should not have consequences for the point that I am making. If *attractive* and *ugly* are subsective, *John's attractive self* would still be a *self of John's*. Therefore, if *self* expresses identity, the same argument would hold.

(18) a. John fears himself.
b. [John's attractive self] fears itself.
c. [John's ugly self] fears itself.

As (18a), (18b), (18c) and (17a) do not have the same meaning, *self* in modified *self* constructions cannot denote the identity function.[11] Given these examples, it is appropriate to relax the meaning of *self*. I argue that in the cases where incorporation occurs and the subject and object are identical, the relation between the two arguments is not strict identity, but simply an illusion of identity, which I formalize more precisely in section 2.2.2.

2.2.2 Analyzing Modified *self*

Given that strict identity is problematic for the modified *self* cases, in this section I argue that a less conservative lexical entry for *self* is required. In the remainder of this paper, I argue for a single lexical entry for *self* given in (19), that consists of a part-of relation. I will show that the compositional semantics for modified and unmodified *self* are different, despite the fact that they make use of the same lexical entry. This is motivated by the fact that the semantic relations present in the modified and unmodified *self* cases (after computing the meaning of the entire clause) are not the same. (Strict identity in the case of unmodified *self* as opposed to a relaxed part-of relation in the case of modified *self*). I begin by laying out the foundations of my proposal, and I then illustrate its application to the modified *self* cases. I show that the proposed analysis accounts for the different classes of adjectives that combine with *self*. In the section that follows, I argue that crucially, the unmodified *self* examples can all be derived if we relax the semantics in this manner. Furthermore, I argue that the semantic unification of the two types of *self* gives rise to the differences present in the syntax (presence vs. absence of *self*-incorporation, which in turn give rise to the presence vs. absence of Principle A).

(19) *The meaning of 'self'* (first sketch)
a. ||self || = λx.λy.y bears R to x *Asymmetric*
b. ||self|| = λx.λy. R(y, x) *Symmetric*

In order to explain the modified *self* examples in Greek and Kutchi Gujarati, I argue that the meaning of the utterance can be derived if we assume that one argument, say x, is a part of the second argument, y, however the argument y is not a part of the argument x. By assuming such an asymmetric relation, we derive a meaning as sketched in (20).

(20) ||John saw his ... self|| = 1 iff John saw ɪz . R (z, John)
and it is possible that not R (John, z)

In (20), I assume that a definite determiner combines with *self*, namely *ton* in Greek and *–ne* in Kutchi Gujarati, which picks out the unique, salient individual that is in relation R to the possessor pronoun bound by *John*. If we add adjectival modifiers to (20), they serve to further specify this individual.

The Greek[12] and Kutchi Gujarati data in (21)–(24) show that there are three types of adjectives (intersective, non-intersective but subsective, and non-subsective but privative) that can modify *self*.[13]

[11] This problem is not restricted to a particular class of adjectives, but applies to all of the classes of adjectives that can combine with the anaphor.

[12] Greek speakers find these modified *self* cases acceptable but slightly marked. Naturally occurring examples appear in Google; thanks to Dimitris Michelioudakis (p.c.) for pointing this out to me.

[13] See Partee (2007) for diagnosing adjective classes. The only class of adjective that cannot modify *self* in Greek and Kutchi Gujarati is plain non-subsective. It is not immediately clear why this particular class of adjective cannot combine with anaphors; this is not predicted by this analysis.

Intersective

(21) a. O Costas idhe ton arosto eafto tu (ston kathrefti). Greek
Costas saw det.M.SG sick self.M.SG 3.M.SG.GEN in.the mirror
'Costas saw his sick self (in the mirror).'

b. valji e-no bimar potha-ne jo-yo. KGu
Valji 3.SG-GEN.M sick self-ACC see-PFV.M.SG
'Valji saw his sick self.'

Non-intersective, subsective

(22) a. O Costas idhe ton sinithi eafto tu. Greek
Costas saw det.M.SG usual/typical self.M.SG 3.M.SG.GEN
'Costas saw his usual/typical self.'

b. valji e-no thik thak potha-ne jo-yo. KGu
Valji 3.SG-GEN.M usual/typical self-ACC see-PFV.M.SG
'Valji saw his usual/typical self.'

Non-subsective, privative

(23) a. Costas idhe ton fandastiko eafto tu. Greek
Costas saw det.M.SG imaginary self.M.SG 3.M.SG.GEN
'Costas saw his imaginary self.'

b. valji e-na khota potha-ne jo-yo. KGu
Valji 3.SG-GEN imaginary self-ACC see-PFV.M.SG
'Valji saw his imaginary self.'

Plain non-subsective

(24) O Costas idhe ton ?*dhinitiko / *endhechomeno eafto tu. Greek
Costas saw the potential / potential-contingent self.M.SG 3.M.SG.GEN
'Costas saw his potential self.'[14]

In order to explain the first class, namely intersective adjectives, identity (as discussed in the previous section) alone would suffice, as long as we focus on a fixed point in time. For instance, if I am sick right now, and I utter the following: *I see my sick self in the mirror*, an analysis in terms of identity would yield the meaning that I see the unique salient individual in the mirror that is identical to *me*, and *sick*. This would be a true statement. However something more needs to be said in order to account for subsective and privative adjective examples. Clearly, *John's better self* (where *better* is subsective) would not be identical to *John* (even at a fixed point in time), and *John's former self* or *John's imaginary self* (former and imaginary being privative) would also not be identical to *John*.

The core of the idea that I have alluded to earlier in this section is that the relation between the two arguments is equivalent to an asymmetric notion meaning something along the lines of part-of;[15] a is a part of b; however, b is not necessarily a part of a. The specific definition of the relation between the two arguments in (25) is not trivial, and I will dedicate the reminder of this section to pinning it down. I will then illustrate the application of the proposal to the non-intersective but sub-sective adjective class.

(25) ||John saw his ... self|| = 1 iff John saw ɪz . R (z, John)
and it is possible that not R (John, z)

Certain classes of adjectives make reference to various points in the individual's life, e.g., *former self*, *current self*, whereas other adjectives may make references to co-existing aspects of an individual

[14]The Kutchi Gujarati equivalent is not possible here, for *possible* is not an adjective, but a complex verb which cannot combine with *self*.

[15]This relation picks out psychological aspects of the individual, for example if John painted himself blue, looked in the mirror and said 'I saw my blue self', the only possible reading in Greek and Kutchi Gujarati is that *John* saw his sad/depressed self.

at a set point in time (e.g., *cynical self*, *idealistic self*). We must first establish the concept of the individual that the denotation of *self* selects a part of (i.e. the internal argument of *self*), before turning to the relation between the two arguments of *self*. I will refer to this individual as the HOST INDIVIDUAL;[16] for every individual, there is a corresponding host individual. The host individual is the collection of all temporal stages[17] of an individual, and the collection of all physical and psychological aspects of that individual at any temporal stage. In other words, the host individual is equal to the individual in their entirety, temporally, physically and psychologically at each point throughout their existence. However, a host individual, encompassing all aspects of an individual, can be underspecified with respect to certain of its properties; for example, a host individual can be both *good* and *bad*, but (naturally) not simultaneously.

The host individual will allow the denotation of *self* to select a part of it, where A PART is defined as follows: In line with Carlson (1977) and Lewis (1983), I assume that at a fixed temporal point and in a fixed world, a synchronic part of a host individual is related to the host individual in the following way. First, the part of the host individual is specified more than the host individual for one or more properties. Second, the part of the host individual does not differ from the host individual in any other way. To illustrate, assume that the host individual δ_1 corresponds to a person, *Dan*, who can be good, but can also be bad (but not both simultaneously); i.e. δ_1 has the property of being *good or bad* — and δ_1 is not further specified. A part μ of δ_1 may make a selection in this regard. *Dan's good self* (or μ_g) might be *good and not bad*, whereas *Dan's bad self* (or μ_b) might be *bad and not good*. Naturally, these may entail further selections — for instance, *Dan's good self* might also be *idealistic and not cynical* whereas *Dan's bad self* might be *cynical and not idealistic*. Again, δ_1 would be underspecified, i.e. δ_1 would be *cynical or idealistic*. The crucial point is that apart from being specified more than δ_1, all of its parts (e.g., μ_g and μ_b) are identical to δ_1; this derives non-identity under near-identity (i.e. *my good self* feels identical to me in some loose sense even though it is not).

Every part of a host individual is a part of a host individual at some set point (or more) throughout the individual's existence, e.g., the part of an individual's personality that is *evil*, the part of their personality that is *fair* etc. all surface at separate points in time. This will be crucial for explaining the non-identity cases. Now that the PART OF relation, which I argue to hold between the two arguments of *self* has been specified, I return to the various classes of adjectives that may combine with *self*. We can first define the lexical entry in (26), based on the above discussion.

(26) $||\text{self}|| = \lambda x.\lambda y.$ y is a part of x

Naturally, there is an interaction between time and parts of a host individual (e.g., if John was innocent in the past, we might say *I admired John's innocent self*, but he may no longer be innocent and never become innocent again). To simplify, I focus on cases where the relevant part of the individual exists at the same time at which the event or state denoted by the predicate exists. For example, consider *John saw his good self*; the part of *John* that is *good*, must be present in the timeframe in which John carried out the act of seeing *this good part of him*, cf. Musan (1999). To illustrate my semantic analysis for the comparatively simple cases, I will first discuss the intersective and subsective class. At a fixed point in time t, the data in (27) can be treated as intersective if *John is sick*, given that at this point in time t John in his entirety is sick.

(27) a. john e-na bimar potha-ne jo-yo. KGu
John 3.SG-GEN sick self-ACC see-PFV.M.SG
'John saw his sick self (in the mirror).'

b. O Costas idhe ton arosto eafto tu ston kathrefti. Greek
Costas saw det.M.SG sick self.M.SG 3.M.SG.GEN in.the mirror
'Costas saw his sick self (in the mirror).'

[16] I inherit the term *host individual* from Musan (1999).

[17] Stages are temporal parts or slices of the individual, cf. Quine (1960), Carlson (1977). A stage can be any length that is included in its host individual's time of existence.

I assume that *bimar* 'sick' combines with *potha* 'self' by means of predicate modification, as is usually assumed for intersective adjectives (though the first argument slot of *potha* 'self' needs to be filled first, e.g., by *John* in (29)). Thus, based on the meaning of *potha* 'self' in (26), the relevant lexical entries and truth conditions associated with *bimar potha* 'sick self' are given in (28) and (29), respectively. The lexical entry shows that in the intersective case (29), the unique individual in object position (in this case a part of *John* that may be identical to the host individual *John*) is sick.

(28) a. ||potha/eafto/self|| = λx.λy. y is a part of x
b. ||bimar/arosto/sick|| = λx.x is sick

(29) a. ||John saw his self|| = 1 iff John saw ɪz . z is a part of John
b. ||John saw his sick self|| = 1 iff John saw ɪz, z is sick and z is a part of John

The privative class of adjectives, as illustrated in (30) and (31) is not as straightforward as the subsective and intersective classes, for we need to assume that the host individual can be segmented according to TEMPORAL SLICES (for (30)) and that we can talk about parts of the host individual that exist in worlds other than the real world (for (31)).

(30) a. john e-na pelano potha-ne nafrat kar-e. KGu
John 3.SG-GEN former self-ACC hate do-3.SG
'John hates his former self.'
b. O Costas misos ton proigoumenos eafto tu. Greek
Costas hates det.M.SG former self.M.SG 3.M.SG.GEN
'Costas hates his former self.'

(31) a. john e-na khota potha-ne prem kar-e. KGu
John 3.SG-GEN imaginary self-ACC love do-3.SG
'John loves his imaginary self.'
b. O Costas agapai ton fandastiko eafto tu. Greek
Costas loves det.M.SG imaginary self.M.SG 3.M.SG.GEN
'Costas loves his imaginary self.'

A temporal slice is a stage of a host individual at a fixed point in time. A host individual exists at different stages of time. For any stage of time, the temporal slice of a host individual at a particular stage of time is an individual that is identical to all of what the individual is at a particular stage of time. Temporal slices interact with parts of a host individual, for example, *John* at *age 7* may have a part that is innocent (but no part that is mature), whereas *John* at *age 28* may no longer have such an innocent part (and only a part that is mature). In this case, the host individual in its entirety has an innocent part as well as a mature part, but both are temporally bound; the innocent part to the earlier time slices of *John* and the mature part to the later time slices of *John*.

Turning to the meaning of *his former self*, we need to introduce a time argument for *self*, illustrated in (32a); *self* can then combine with *former*, as defined in (32b), adapted from von Fintel & Heim (2010:69). (I differ from von Fintel & Heim in assuming that t is of type l.) As shown in (32c), we derive the correct truth conditions for *John hates his former self*, namely that *John* hates some part (or all) of what he was at some point in the past, but what he is no longer.

(32) a. $||\text{potha/eafto/self}||^t = \lambda x.\lambda y.$ y is a part of x at t
b. $||\text{pelano/proigoumenos/former}||^t = \lambda f_{<l,<e,t>>}.\lambda x.[f(t)(x) = 0$ & ∃t' before t: $f(t')(x) = 1]$
c. $||\text{John hates his former self}||^t = 1$ iff John hates ɪz . z is not a part of John at t & ∃t' before t: z is a part of John at t'

Having accounted for (30), let us move on to (31). Although *khota potha* 'imaginary self' is a privative adjective, it is not enough to simply apply the analysis for *pelano* 'former'; something more is required. The problem is that *John's imaginary self* can refer to an individual that only exists in John's dreams and does not exist at any point in time in the real world. In this sense, *John's*

imaginary self may not be part of the host individual *John* in the real world. If, however, we relativize *self* to possible worlds,[18] we can analyze this on analogy to *pelano* 'former'. I provide a rough sketch in (33), assuming that the person whose imagination *khoto* 'imaginary' refers to, is provided by the context (as indicated in (33b)), and resolved towards *John* in (33c). The truth conditions in (33c) correctly capture the fact that *John loves his imaginary self* is true in a situation where *John* believes that he is a hero (even though he is not), and John loves the hero that he thinks he is.

(33) a. $||\text{potha/eafto/self}||^w = \lambda x.\lambda y.$ y is a part of x in w

b. $||\text{khoto/fandastiko/imaginary}||^w = \lambda f_{<s,<e,t>>}.\lambda x.[f(w)(x) = 0$ & ∀w' compatible with the day dreams of some salient individual y: $f(w')(x) = 1]$

c. $||\text{John loves his imaginary self}||^w = 1$ iff John loves ιz . z is not a part of John in w & ∀w' compatible with the day dreams of John: z is a part of John in w'

The question remains why *his alleged self* or *his putative self* (with plain non-subsective adjectives) seem to be impossible (crosslinguistically); the semantic analysis would predict these to be acceptable. At this stage, it is not clear how to account for this apparent gap.

2.3 Unifying Unmodified & Modified *self*

In the previous section, we saw that relaxing the identity relation allowed us to derive the correct truth conditions for modified *self*. In this section, we will see that the proposed semantics from the previous section carries over to the cases that appear to exhibit identity. In order to account for these cases, I assume that the relation between the arguments becomes symmetric due to the incorporation of *self*. Before presenting this analysis, it is worth discussing an alternative approach to deriving identity from the part-of relation. I argue that this alternative faces problems that my analysis does not.

Take the above analysis, supported by native speaker intuitions regarding modified *self* in Greek and Kutchi Gujarati (*self* picks out a particular part of the individual, rather then the individual in their entirety). It seems appropriate to assume that the maximal part of an individual in the absence of an adjective (i.e. in the case of unmodified *self*) is by default the individual in its entirety, which may be most salient. Thus, in both the modified and unmodified *self* cases, the part of the individual that is picked out could be explained in terms of saliency. In the example *John's sick self*, the most salient part of *John* is selected that counts as *sick*, whereas in the case of *himself*, the host individual of the referent (*him*) is selected in its entirety.

Although such a view is plausible, this line of enquiry seems empirically incorrect. Consider the following scenarios. If we assume that *himself* refers to a maximal, salient part of the individual selected by *him* (rather than necessarily the entire individual), one possible reading that should be available for (34) is one where *John likes his dark self* (which would be the most salient part of *John's*). However, (34) only seems to have the reading where *John likes John* (as an individual, in his entirety).

(34) After discovering John's dark self last weekend, I'm surprised that he likes himself.

Similarly, the anaphor *himself* should fail to refer to a particular part of *John* (including the maximal part that corresponds to *John* in his entirety) if several parts of *John* are made salient, as in (35a). We find such an effect in (35b), where the referent of *his dog* cannot be resolved easily (though resolution to the closest antecedent may be an option). However, in (35a), such an effect is absent. The final sentence clearly means *John nevertheless likes John (in his entirety)*, even though this is not the most salient part of *John* and in fact there is no unique most salient part of *John* in this utterance.

(35) a. There are many sides to John. He clearly has a good self, he cares about others, he wants

[18] Of course, once we relativize *potha* 'self' to worlds and times, the world and time parameter will always be present, but for simplicity, I only write them as and when they are needed.

to do things morally, etc. At the same time, his psychotic self always takes over when he's stressed, and his greedy self takes over when large amounts of money are involved. John is aware that he has issues, but he nevertheless likes himself.

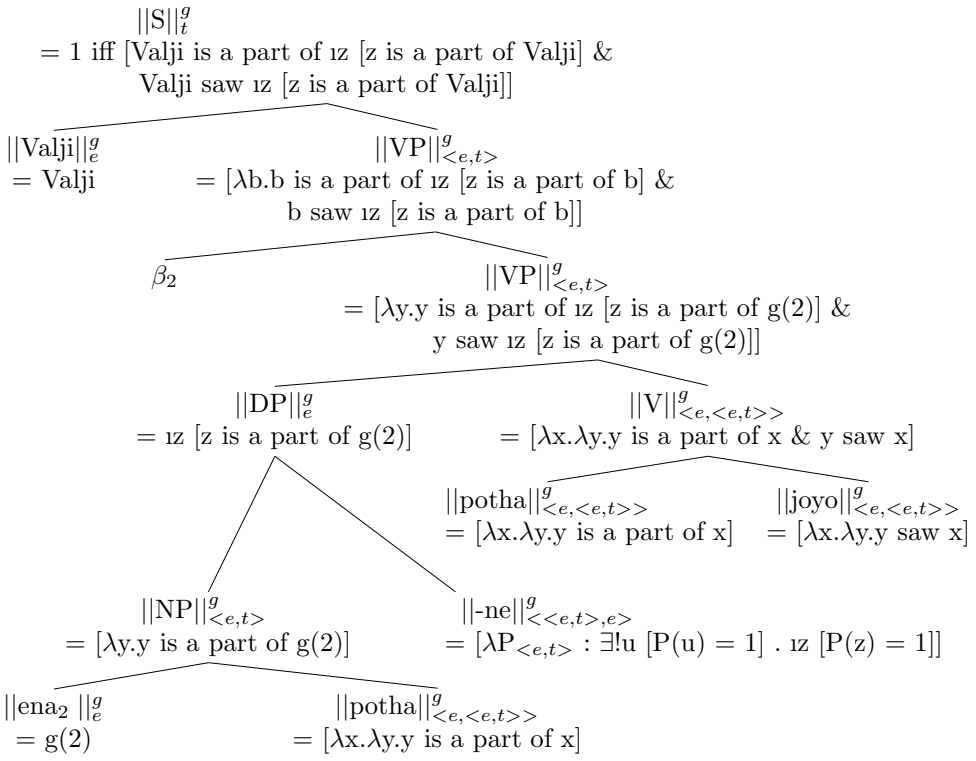
b. John owns a small kennel. He has a cute, well-behaved dog. He also has a psychotic dog that attacks people, and a timid dog that always hides. ?? Nevertheless, he's happy and he likes his dog.

To account for identity with unmodified *self*, I propose (as outlined already in section 2.1) that unmodified *self* undergoes incorporation of the anaphor into the predicate, and that the anaphor is interpreted twice, reversing the order of the predicates. While this was irrelevant in section 2.1, identity being symmetric, it has an impact in the present case, as it compositionally yields *x is a part of y* and *y is a part of x*, thereby turning the part-of relation into a symmetric relation. Put differently, in (36a), the meaning of the DP *ena potha-ne* is the unique individual that is a part of *Valji*. By interpreting *potha* twice and combining it with the meaning of the predicate, the meaning of the VP *ena potha-ne (potha-)joyo* is a conjunction between *being a part of the unique individual that is a part of Valji* and *seeing the unique individual that is a part of Valji*. The LF for an example such as *valji ena potha–ne joyo* 'Valji saw himself' is given in (36b), the complete derivation is illustrated in (36c).

(36) a. valji$_2$ [e-na$_2$ potha-ne] jo-yo. KGu
Valji 3.SG-GEN self-ACC see-PFV.M.SG
'Valji saw himself.'

b. LF: Valji$_2$ [ena$_2$ potha-ne] potha-joyo.

c.



d. *In words:* Valji is a part of the unique individual (in the utterance context) that is a part of Valji and Valji saw the unique individual that is a part of Valji.

Due to *self*-incorporation (which leads to assertion of 'Valji is a part of ɪz [z is a part of Valji]'),

identity between the subject and the object is established in the grammar. In other words, even though the part-of relation is not symmetric, *self*-incorporation creates a symmetric relation between the two arguments. This derives both identity and Principle A effects with unmodified *self*.

Conversely, in the case of modified *self*, *self*-incorporation would lead to an incorrect interpretation. Consider the following scenario: Costas has a pitiful side, but he also has an over-confident side, both of which he shows at different times. As a matter of fact, he admires not only his over-confident side, but also his pitiful side. In this scenario, (37) may be uttered. If *self*-incorporation took place here, we would get the meaning in (38a), which incorrectly entails (38b); the correct reading is derived without *self*-incorporation, this is given in (39). This is compatible with the assumption that there is no *self*-incorporation with modified *self*.

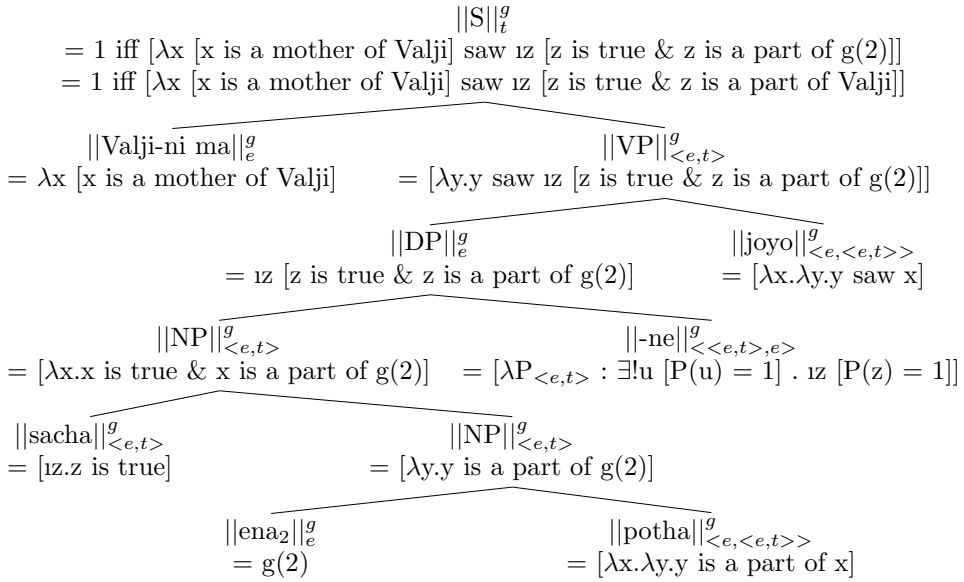
(37) O Costas thavmazi [ton **aksiolipito** eafto tu]. Greek
Costas admires det.M.SG pitiful self.M.SG 3.M.SG.GEN
'Costas admires his pitiful self.'

(38) ||O Costas thavmazi ton **aksiolipito** eafto tu]||
a. = 1 iff Costas is a part of ɪz [z is pitiful & z is a part of Costas]
& Costas admires ɪz [z is pitiful & z is a part of Costas]] *incorrect*
b. entails: Costas (in his entirety) is pitiful.

(39) ||O Costas thavmazi ton **aksiolipito** eafto tu]||
a. = 1 iff Costas admires ɪz [z is pitiful & z is a part of Costas]] *correct*
b. does not entail: Costas (in his entirety) is pitiful.

Since modified *self* does not involve *self*-incorporation, we now derive the absence of Principle A effects, e.g., in (40a), which has the LF in (40b) and derivation in (40c). (Note that in (40), the pronoun *ena* must reconstruct to a position adjacent to *potha* 'self', since it is interpreted as the first argument of the relational noun phrase *potha* 'self'.)

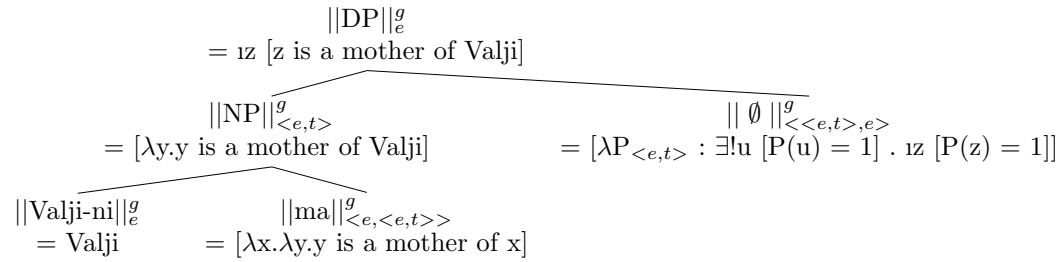
(40) a. valji$_i$-ni ma [e-na **sacha** potha$_i$-ne] jo-yo. KGu
Valji$_i$-GEN.F mother 3.SG-GEN true self$_i$-ACC see-PFV.M.SG
'Valji$_i$'s mother saw his true self$_i$.'
b. LF: [Valji$_2$-ni ma] [sacha ena$_2$ potha-ne] joyo.
c.



d. *In words:* Valji's mother saw the unique individual (in the utterance context) that is true and that is a part of Valji.

The structure and derivation of *valji-ni ma* 'Valji's mother' is given separately in (41) for reasons of space. Here, *ma* 'mother' is assumed to be a relational noun, which takes *Valji* as its first argument. In subject position, definiteness is not overtly marked, but by analogy to the definiteness marking in object position (due to the differential object marker *–ne*), I assume that definiteness is introduced by a null marker, as indicated.

(41)



We can now make a stronger claim to motivate *self*-incorporation in the case of unmodified *self*, but not in the case of modified *self*. Given that *self* selects a part of a host individual and the individual in its entirety does not seem to be available as a default maximal part (cf. (34) and (35)), *himself* (without *self*-incorporation) would be radically underspecified at any given point in time.

Even more problematic, as we can see in (42a-c), the part-of relation can pick out something that existed in the past and something that will exist in the future. Finally, it is possible to modify the part-of relation explicitly to pick out individuals in their entirety, (42d).

(42) a. In the moral sphere, I make decisions and count on my future self to carry them out. (www.ucs.mun.ca/ davidt/Intuition.htm)
⟹ *my future self* refers to an individual that does not yet exist

b. Although GD is now 'hardcore', I still miss his innocent self. When he was with Big Bang, he was so cute.
(http://sookyeong.wordpress.com/2009/12/07/gdragons-concert-pinned-with-controversies/)
⟹ *his innocent self* refers to an individual that no longer exists

c. Schumacher kept smiling, kept giving non-committal answers ... and kept driving like a shadow of his former self.
(www.itv-f1.com/Controller.aspx?PO_ID=49236)
⟹ *his former self* refers to an individual that no longer exists

d. If he did not receive a glowing review from the boss over his last presentation, the Perfectionist sees it as a failure of his entire self.
(therapyinphiladelphia.com/selfhelp/tips/is_it_low_self_esteem/)
⟹ *his entire self* possibly refers to a complete host individual

The question would thus arise, which part of a host individual unmodified *self* refers to, if it still expresses the part-of relation. As indicated in (43), *himself* typically picks out an entire host individual at the point in time that the predicate holds at. *self*-incorporation is a means to grammatically encode this connection, deriving identity from the part-of relation. *self*-incorporation achieves the same results (structurally) that we would get from modifying *self* by means of the adjective *entire* (as in *his entire self*, cf. (42d)).

(43) John admires himself.
≈ At a point t John admires the entire time slice of John at t.
≠ John admires his former self (only).
≠ John admires his good self (only). (*but he may be unaware that he has another side*)

Based on this discussion, I propose the LF requirement in (44). This interface requirement posits that

self cannot remain unconstrained (referring to all or any parts of the host individual). It must either be specified and thus constrained by means of an adjective, or else incorporate into the predicate giving rise to an equally constrained identity interpretation. In other words, the idea is that it must be made clear either by means of adjectival modification or by means of *self*-incorporation, which part of an individual *self* is meant to select.

(44) *Interface requirement on 'self' (at the LF interface):*
To guarantee successful communication, *self* cannot be unconstrained.
(i) Either a specific part of the host individual that *self* selects must be selected by means of an adjective;
(ii) Or, as an alternative strategy, the meaning of a *self*-containing clause is disambiguated by means of *self*-incorporation, which gives rise to the identity relation.

In the formal implementation, one may wonder why ɪz [z is a part of x] does not reduce to x (thus resolving the underspecification without *self*-incorporation), given that the iota operator has a maximalizing property. (This concern is similar to the concern addressed around examples (34)–(35).) However, observe that even expressions such as *the left part of my head* are contextually restricted in terms of their reference; e.g., if a speaker says that the left part of her head hurts, there is no entailment that the entire left half of her head hurts (which would be the maximal left part of her head). Moreover, it seems infelicitous (possibly due to scalar implicatures) to use concepts such as *part of*, *subset of*, etc., to refer to something in its entirety; to illustrate, even though every set is a subset of itself, it is generally deviant to refer to an entire set by means of the definite description the subset (for instance, the brief conversation A: 'Please draw a subset of set X.' — B: 'OK, I've drawn the subset.' would not make much sense if B simply redrew the original set X, which would be the maximal subset of X).

The question at this point is whether this analysis overgenerates. Specifically, do we find cases where *self*-incorporation occurs in the presence of an adjective, giving rise to Principle A effects (the meaning of *self* in such constructions amounting to identity? The next section is concerned with this question.[19]

3 The Syntactic Distribution of Principle A

3.1 Principle A as a Consequence of *self*-incorporation

In the section 1, we observed a correlation between the presence/absence of an adjective and the absence/presence of Principle A. The data in section 1 highlighted the fact that in Greek and Kutchi Gujarati unmodified *self* is always sensitive to Principle A, whereas modified *self* appears to be exempt from it. In the remainder of this paper, I argue for a syntactic analysis of Principle A effects involving *self*, which assumes that the anaphor covertly incorporates into the verb, as illustrated in example (45) and motivated in section 2.3. As we have seen, a transitive verb that incorporates *self* requires identity of subject and direct object (by means of the compositional semantics), thus giving rise to Principle A.

(45) Costas$_i$ admires himself$_i$
LF: Costas$_i$ self-admires himself$_i$

[19] As a final remark, it is worth addressing *Madame Tussaud* sentences, as discussed by Jackendoff (1992). The example in (i) in English (from Jackendoff 1992:4) has a reading in which it means that Ringo Starr (the actual person) started undressing the statue of Ringo Starr.

(i) The other day I was strolling through the wax museum with Ringo Starr, and we came upon the statues of the Beatles, and . . . . . All of a sudden Ringo started undressing himself.

In Kutchi Gujarati, *ena potha-ne* does not appear to allow for such readings, though it has suggested by an anonymous reviewer that they are possible in Modern Greek. More research is required to determine crosslinguistic empirical facts.

An approach along these lines was first argued for by Anagnostopoulou & Everaert (1999) (based on the Reinhart & Reuland 1993 system), and developed in Reuland & Winter (2009).[20] *self*-incorporation is generally viewed as the most economic way to encode binding dependencies (Reuland & Winter 2009:75).[21] I argued in section 2.3 that *self*-incorporation is actually due to an LF restriction that bans utterances with *self* from remaining unconstrained. Therefore, *self*-incorporation is generally obligatory unless it is blocked by independent syntactic principles (e.g., the coordinate structure constraint). In the latter case, we can assume that the configuration is still grammatical, as the interface requirement against unconstrained *self* can be flouted as a LAST RESORT. This is discussed in the following sections.

3.2 The Syntax of Anaphor Incorporation

Given that obligatory anaphor incorporation gives rise to Principle A effects, the question that remains to be answered is, what bans modified *self* from incorporating into the predicate? If modified *self* did incorporate, we would expect to see Principle A effects (and possibly a ban against modification with privative adjectives if incorporation is obligatory). This section aims to address these issues. I argue in section 3.2.1 that incorporation of modified *self* is blocked by the adjective, as incorporation across the adjective would violate independently motivated constraints that hold in the narrow syntax.

3.2.1 The Absence of Principle A with Modified *self*

The main claim I wish to make here, is that the absence of Principle A with modified *self* can be explained by locality; the presence of an adjective blocks *self*-incorporation due to Relativised Minimality (Rizzi 1990), a constraint which states that if movement to a certain position targets an element of a certain category (e.g., in the present case: a lexical head), then the closest appropriate element must move. This is illustrated in (46).

(46) *Relativised Minimality* (based on Rizzi 1990)

a. *possible movement*:

$Y_{\text{target position}}$ [$Z_{\text{element of a }\mathbf{\text{different}}\text{ category from X}}$ [$X_{\text{element to be moved}}$

b. *impossible movement*:

$Y_{\text{target position}}$ [$Z_{\text{element of the }\mathbf{\text{same}}\text{ category as X}}$ [$X_{\text{element to be moved}}$

Let us assume that *self*-incorporation involves a configuration where the verb attracts a lexical head (I assume that this is a generalized process that crosslinguistically underlies incorporation). In the examples in (47), the closest lexical head c-commanded by the verb is not *self* but the adjective *true* (this follows if the structure of a DP is [$_{\text{DP}}$ D [$_{\text{AP}}$ A [$_{\text{NP}}$ N]]]). Given the definition of Relativised

[20] Anagnostopoulou & Everaert (1999) proposed this approach for Greek, arguing that Greek *eafto* 'self' incorporates into the verb by covert head movement at LF. Anagnostopoulou & Everaert (1999) assume that head movement of *eafto* 'self' leaves a trace, but do not provide an explicit semantic analysis. Reuland & Winter (2009), in their analysis of English *self*, assume that incorporated *self* is only interpreted in its landing position.

[21] Reuland & Winter (2009) do not elaborate on the nature of this economy principle; the idea, as presented in Reuland (2001, 2005), can be roughly summarized as follows. If the subject and object of a transitive verb are coreferent, the object's interpretation is dependent on that of the subject. This dependency, which is indicated by the use of *self*, can be encoded computationally (i.e. syntax) by means of *self*-incorporation. Alternatively, it can be resolved at the interface by means of computational semantics. Reuland's general idea regarding economy is that it is more economical to encode such an interpretive dependency in the syntax, making it 'hard and fast' (Reuland 2005). This makes *self*-incorporation obligatory unless it is blocked by syntactic constraints, which suspend this economy principle. We can envisage this as follows. If *self* can incorporate, the competing derivation where it does not incorporate is eliminated by economy, given the more economic derivation with *self*-incorporation. However, if *self* cannot incorporate to begin with, no comparison between the two derivations will take place. See Reuland (2001, 2005, 2011) for further discussion.

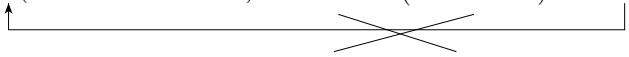
Minimality in (46), consequently in the Greek and Kutchi Gujarati examples in (47) the predicate can only attract *true* and not *self*, for the adjective is closer. This means that *self*-incorporation is blocked and in this sense, adjectives are interveners for *self*-incorporation.

(47) a. [I mitera tu Janni$_i$] agapai [ton **alithino** eafto$_i$ tu]. Greek
3.F.SG.NOM mother Janni$_i$.GEN loves 3.M.SG.ACC true self$_i$ 3.M.SG.GEN
'Jannis$_i$'s mother loves his true self$_i$.'

b. valji$_i$-ni ma [e-na **sacha** potha$_i$-ne] prem kar-e. KGu
Valji$_i$-GEN.F mother 3.SG-GEN true self$_i$-ACC love do-3.SG
'Valji$_i$'s mother loves his true self$_i$.'

The relevant configuration for relativised minimality is summarized in (48).

(48) V$_{\text{(attracts lexical head)}}$ ⋯ [AP *true*$_{\text{(lexical head)}}$ [NP *self*$_{\text{(lexical head)}}$]]



The above proposal explains the correlation between the two *selfs* (modified vs. unmodified) and Principle A, by highlighting that the core difference between them can be explained by locality. However, it is not immediately clear why the determiner is not an intervener for head movement in (3) and (4), repeated as (49) and (50) below. Greek has an overt determiner *ton* 'the'. In Kutchi Gujarati, I assume for now that the differential case marker *–ne* bears the properties of a determiner (as it correlates with definiteness/specificity).

(49) a. O Costas$_i$ vlepi [**ton** **eafto**$_i$ **tu**]. Greek
Costas$_i$ sees det.M.SG self$_i$.M.SG 3.M.SG.GEN
'Costas$_i$ sees himself$_i$.'

b. *O Costas$_i$ xeri oti Maria vlepi [**ton** **eafto**$_i$ **tu**]. Greek
Costas$_i$ knows that Maria sees det.M.SG self.M.SG 3.M.SG.GEN
'Costas$_i$ knows that Maria sees himself$_i$.'

(50) a. john$_i$ [**e-na** **potha**$_i$**-ne**] jo-yo. KGu
John$_i$ 3.SG-GEN self$_i$-ACC see-PFV.M.SG
'John$_i$ saw himself$_i$.'

b. *john$_i$ kidthu ke Maria [**e-na** **potha**$_i$**-ne**] jo-yo. KGu
John$_i$ said that Maria 3.SG-GEN self$_i$-ACC see-PFV.M.SG
'John$_i$ said that Maria saw himself$_i$.'

The difference between adjectives and determiners regarding their status as interveners can be explained by the distinction between the two categories; Baker & Hale (1990) demonstrate that lexical heads and functional heads must be treated as different categories by Relativised Minimality (i.e. there is no uniform Head Movement Constraint). They argue that lexical heads are interveners for lexical heads but not for functional heads and vice versa. An example of a lexical head (the noun *seuan* 'man') incorporating into the verb across a functional head (the demonstrative determiner *yede* 'that') is given in (51).

(51) a. [Yede **seuan**-ide] a-mu-ban. Southern Tiwa
that man-suf 2sS/A-see-past
'You saw that man.'

b. [DP Yede [NP [N t$_i$]]] a-**seuan**$_i$-mu-ban. Southern Tiwa
that 2sS-man-see-past
'You saw that man.' (Baker & Hale 1990:291, quoting Allen, Gardiner & Frantz 1984)

Thus, a functional head such as a determiner (*ton* in (49)), is not an intervener for a lexical head, such as the noun *eafto* 'self'.[22] Therefore, while in the examples in (47), *self*-movement is blocked due to relativised minimality (given that *true* is also a lexical head), in (49) and (50), it is not. This distribution is summarized in (52).

(52) a. * $V_{\text{(attracts lexical head)}}$ [AP $true_{\text{(lexical head)}}$ [NP $self_{\text{(lexical head)}}$]] *intervention*

b. $V_{\text{(attracts lexical head)}}$ [DP $the_{\text{(functional head)}}$ [NP $self_{\text{(lexical head)}}$]] *no intervention*

The above explanation, depicted in (52), assumes that incorporation of modified *self* cannot occur due to processes in the narrow syntax; however, incorporation of unmodified *self* is acceptable, as there is no relevant intervener. This derives the contrast between modified and unmodified *self*.

For completeness' sake, it is worth pointing out that possessor DPs do not intervene with *self*-incorporation either. It is plausible, as I have been assuming, that the possessor, *tu* in (53a) and *ena* in (53b), though taking the shape of a genitive-marked DP, is truly a complement of *self*, given that *self* is a relational noun. In Greek, it is likely that this complement is still in its base position, whereas in Kutchi Gujarati, the possessor has moved (plausibly as an XP) to the specifier of the DP to derive the surface order; therefore, neither can be an intervener. This is illustrated by the bracketed structures in (53).

(53) a. O Costas$_i$ vlepi [**ton** **eafto**$_i$ **tu**]. Greek
Costas$_i$ sees det.M.SG self$_i$.M.SG 3.M.SG.GEN
'Costas$_i$ sees himself$_i$.'

Costas [$_{\text{DP}1}$ ton [NP eafto [$_{\text{DP}2}$ tu]]]

b. Costas$_i$ [**e-na** **potha**$_i$**-ne**] jo-yo. KGu
Costas$_i$ 3.SG-GEN self$_i$-ACC see-PFV.M.SG
'Costas$_i$ saw himself$_i$.'

Costas [$_{\text{SpecDP}1}$[$_{\text{DP}2}$ ena] [$_{\text{D}'1}$ D [NP potha-ne t_{ena}]]]

For Kutchi Gujarati, we can now give a complete illustration (54a+b), assuming that the differential case marker *–ne* is located in D[23] (which is head-final). This contrasts with (55a+b), where *self*-incorporation is blocked by the presence of the adjective *mota* 'big'.[24]

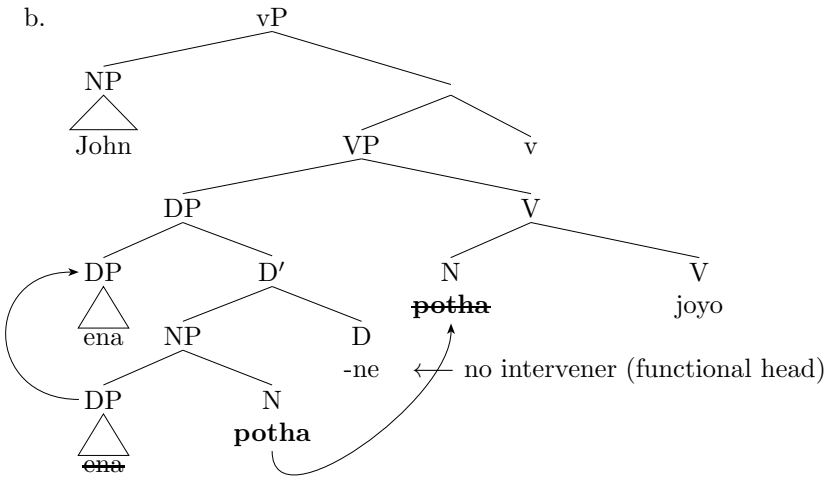
(54) a. john e-na potha-ne jo-yo. KGu
John 3.SG-GEN self-ACC see-PFV.M.SG
'John saw himself.'

[22]Note (although not relevant here), a lexical head is not an intervener for incorporation of a functional head either; cf. Baker & Hale (1990) for examples.

[23]The structures for Greek are analogous except for the difference in headedness (Greek being head-initial), and the fact that the genitive noun phrase does not move into SpecDP.

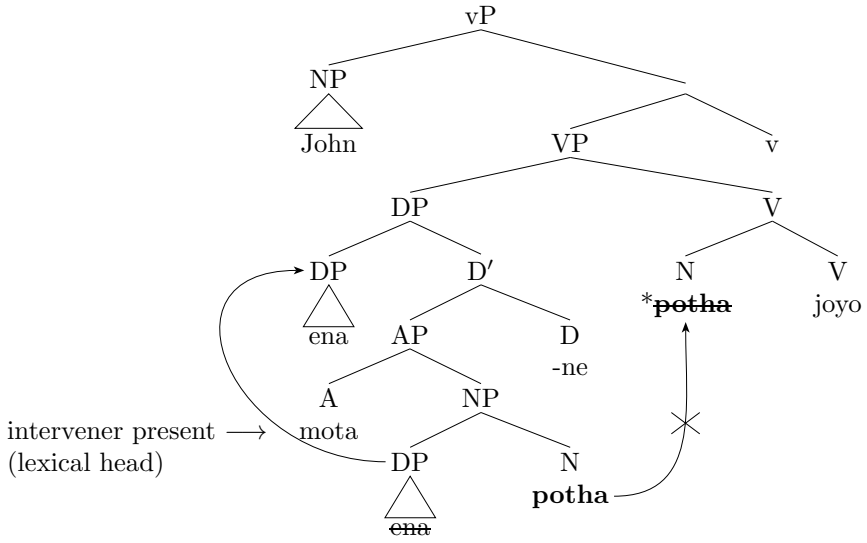
[24]Strike through marks unpronounced copies in these trees.

b.



(55) a. john e-na mota potha-ne jo-yo KGu
John 3.SG-GEN big self-ACC see-PFV.M.SG
'John saw his fat self.'

b.



4 Old Puzzles Revisited: Unbound Possessors in Greek

Greek allows for cases where the possessor of *eafto* 'self' is truly unbound (i.e. where it does not co-vary with any other expression in the clause). This is illustrated in (56). (Anagnostopoulou & Everaert call this the reified substantive reading of *eafto*, a term that I do not adopt). The possibility of such an unbound possessor (here: *tis Marias* 'of Mary') actually follows from my analysis whenever *self*-incorporation is blocked. In (56) the resulting meaning is that only *Jannis* knows the unique individual that is a part of *Maria's* and that is *good*.

(56) O Jannis xeri mono ton kalo eafto **tis Marias**
Jannis.NOM knows only the good self.ACC Maria.GEN
'Jannis only knows Maria's good self' (Anagnostopoulou & Everaert 1999:103)

5 Introducing A New Puzzle: Possessive Reflexives & Alienability

Kutchi Gujarati possessive pronouns seem to be subject to Binding Principles, illustrated by the examples in (57). If a possessor is c-commanded by a local antecedent, it must contain the reflexive *potha* 'self', giving rise to Principle A and B effects.

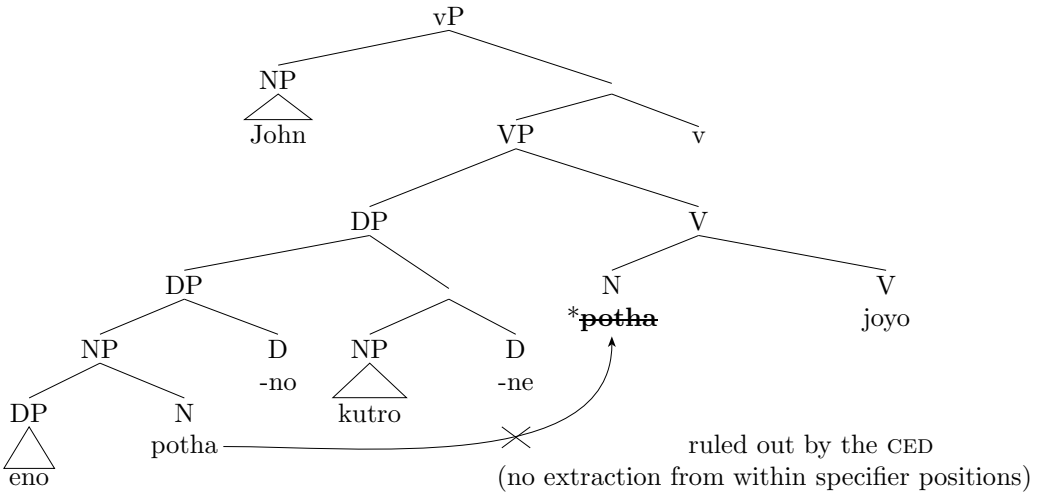
(57) a. valji$_i$ [e-na$_{k/\#i}$ mota kutra] jo-ya. KGu
Valji 3.SG-GEN.PL big.PL dog.PL see-PFV.PL
'Valji$_i$ saw his$_{k/\#i(\#own)}$ big dogs.'

b. valji$_i$ [[e-na potha-na]$_{i/\#k}$ mota kutra] jo-ya. KGu
Valji 3.SG-GEN.PL self.PL-GEN.PL big.PL dog.PL see-PFV.PL
'Valji$_i$ saw his$_{i/\#k}$ own big dogs.'

However, in this case, the presence of Principle A in possessive reflexive constructions cannot be due to *self*-incorporation, for if we were to force *self*-incorporation in (57), then we would derive a meaning that does not correspond to the meaning of the utterance, suggesting that *self*-incorporation cannot take place in the case of possessive reflexives. One could argue that the genitive possessor is located in the SpecDP of the DP that it modifies; the consequence of such an assumption is that incorporation is blocked by the same principle that gives rise to the Condition on Extraction Domains (CED), Huang (1982). The relevant generalization is that we can extract from complement positions, but not from specifier positions;[25] although we see Principle A effects, the possessive reflexive constructions show that incorporation of the anaphor is blocked by independent syntactic principles, in this case the CED, illustrated in (58).

(58) a. #john$_k$ [[e-na potha-no]$_k$ kutro] potha-jo-yo. KGu
John 3.SG-GEN self-GEN.M dog.M.SG self-see-PFV.M.SG
'John self-saw his dog.'

b.



Given that reflexive possessors cannot undergo *self*-incorporation, the question naturally emerges how Principle A applies to possessors. First of all, there is evidence that Principle A effects are rather limited in the case of possessive reflexives, as we can have genitive-marked (unmodified) *potha* in

25

(i) *Extraction from a complement position*
a. I met [a woman from England].
b. **Who** did you meet [**t** from England]?

(ii) *Extraction from a specifier position*
c. *I read [a woman from England]'s article on China.
d. ***Who** did you read [**t** from England]'s article on China?

the possessor position of subject DPs. So, possessive *potha* does not seem to require an antecedent outside the DP in any case. Then why do Principle A effects like in (57) occur?

(59) a. ama-ro potha-no kutro aav-yo. KGu
2.SG-GEN self-GEN.M dog.M.SG come-PFV.M.SG
'Our (own) dog came.' (literally 'Our self's dog came.')

b. john-nu potha-nu ghar bari g-yu. KGu
John-GEN.N self-GEN.N house burn went-PFV.N.SG
'John's (own) house burned down.' (literally 'John's self's house burned down.')

c. e-nu potha-nu ghar bari g-yu. KGu
3.SG-GEN.N self-GEN.N house burn went-PFV.N.SG
'His/Her (own) house burned down.' (literally 'His / Her self's house burned down.')

The following facts suggest that the necessity (and possibility) of *potha* inside possessors (cf. (59)) might be linked to alienable/inalienable possession, and it might be of a semantic nature. First, in cases of inalienable possession (e.g., kinship terms, body parts and properties), *potha* is optional. This is illustrated by the data in (60).

(60) a. john e-ni (potha-ni) ben-ne jo-yi. KGu
John 3.SG-GEN.F self-GEN.F sister-ACC see-PFV.F.SG
'John saw his own sister.'

b. john e-no (potha-no) hath-ne upar-yo. KGu
John 3.SG-GEN self-GEN.M arm-ACC raise-PFV.M.SG
'John raised his (own) arm.'

c. john e-na (potha-na) vaar-ne ketch-ya. KGu
John 3.SG-GEN self-GEN hair-ACC pull-PFV.PL
'John pulled his (own) hair.'

d. john e-ni (potha-ni) aakhi-ne bandth kar-i. KGu
John 3.SG-GEN.F self-GEN.F eyes-ACC close do-PFV.F.SG
'John closed his (own) eyes.'

e. john e-ni (potha-ni) uchai maapi. KGu
John 3.SG-GEN.F self-GEN.F height measured
'John measured his (own) height.'

The data in (60) contrast with (61), where *potha* is necessary.

(61) a. john e-no *(potha-no) kutro jo-yo. KGu
John 3.SG-GEN.M self-GEN.M dog.M.SG see-PFV.M.SG
'John saw his own dog.'

b. john e-nu *(potha-nu) ghar jo-yu. KGu
John 3.SG-GEN.N self-GEN.N house see-PFV.N.SG
'John saw his own house'.

c. john e-ni *(potha-ni) gaadi jo-yi. KGu
John 3.SG-GEN.F self-GEN.F car see-PFV.F.SG
'John saw his own car.'

d. john e-ni *(potha-ni) chopri jo-yi. KGu
John 3.SG.F-GEN.F self-GEN book see-PFV.F.SG
'John saw his own book.'

Finally, the examples of a reflexive object in (62a) does not even allow for *potha* inside the possessor of the argument *ena potha-ne*, cf. (62b), and thus contrasts with (59) above.

(62) a. valji$_i$ [e-na mota potha-ne]$_i$ jo-yo. KGu
Valji 3.SG-GEN big self-ACC see-PFV.M.SG
'Valji saw his fat self.'

b. *valji$_i$ [[e-na potha-na] mota potha-ne]$_i$ jo-yo. KGu
Valji 3.SG-GEN self-GEN big self-ACC see-PFV.M.SG
'Valji saw his fat self.'

The data in this section seem to suggest that *potha* in possessor position is required to mark co-reference between the possessor and a local antecedent if the possessive relation is alienable, whereas it is optional if this relation is inalienable. It seems as though *potha* in these constructions (or perhaps it is the constructions themselves), cannot be compared to the cases discussed in the main sections of this paper. It follows that in possessive reflexive constructions, any Principle A and B effects that we observe are plausibly semantic in nature, for alienability is a semantic property of a possessive relation. The observations discussed in this section extend beyond the scope of this project, and I leave them open for future research.

6 Extending the Analysis

As shown before, I derive Principle A effects from *self*-incorporation. In this section, I discuss another case where the reflexive and its antecedent are not co-arguments (i.e. arguments of the same verb); however, this time, this cases patterns like those in the main sections. The case I am referring to is exceptional case marking (ECM) in Greek and Kutchi Gujarati. ECM constructions are problematic for a view where Principle A is derived by *self*-incorporation or other mechanisms of subject-object identification, such as Reinhart & Reuland (1993). In the case of ECM constructions containing an anaphor, Principle A effects arise. Given that *self* and its antecedent do not belong to the same predicate, *self*-incorporation alone will not do, and an alternative must be pursued.

6.1 ECM Constructions in Greek and Kutchi Gujarati

The examples in (63) and (64) are ECM constructions in Kutchi Gujarati and Greek respectively.[26] The example in (65) illustrates acceptable ECM clauses containing an anaphor.

(63) a. valji-ne reena-ne dablu mar-va-nu khaptu tu. KGu
Valji-DAT Reena-ACC container.N.SG hit-INF-GEN.N want aux
'Valji wanted Reena to hit the container.'

b. valji-ne reena-ne sui ja-va-nu khaptu tu. KGu
Valji-DAT Reena-ACC sleep go-INF-GEN.N want aux
'Valji wanted Reena to go to sleep.'

(64) a. O Yiorghos perimene tin Maria na grapsi asxima sto djagonisma. Greek
the Jorghos expected the.ACC Maria subj write badly in.the exam
'Jorghos expected Maria to do badly in the exam.'

b. O Yiorghos ithele tin Maria na grapsi asxima sto djagonisma. Greek
the Jorghos wanted the.ACC Maria subj write badly in.the exam
'Jorghos wanted Maria to do badly in the exam.'

[26] For Greek ECM, see Schneider-Zioga (1992) and Kotzoglu (2002). The Kutchi Gujarati examples below indicate that we are dealing with ECM in (63). Example (i) shows that the accusative-marked argument can be an expletive with an associate that's an argument of the embedded clause (*it ... that John will come*).

(i) valji-ne aa-ne kevai ja-va-nu khaptu tu ke John av-se. KGu
Valji-DAT that-ACC said go-INF-GEN.N want aux that John come-FUT
'Valji wanted it to be said that John will come.'

Example (ii) shows that the accusative-marked argument can be part of an idiom in the embedded clause.

(ii) valji-ne tari jeeb-ne kap-vai ja-va-nu khaptu tu. KGu
Valji-DAT your tongue-ACC cut-pass go-INF-GEN.N want aux
'Valji wanted you to shut up.'

This indicates that the accusative-marked argument (while plausibly receiving accusative case from the matrix clause) is an embedded subject, i.e. we are not dealing with a control construction (or with a construction where the embedded clause has a *pro* subject).

(65) a. valji-ne **e-na** **potha-ne** sui ja-va-nu khaptu tu. KGu
Valji-DAT 3.SG-GEN self-ACC sleep go-INF-GEN.N want aux
'Valji wanted himself to go to sleep (... but he found it hard to stop watching TV).'

b. O Yiorghos perimene **ton** **eafto** **tu** na grapsi asxima
the Jorghos expected the.ACC self.M.SG 3.M.SG.GEN subj write badly
sto djagonisma. Greek
in.the exam
'Jorghos expected himself to do badly in the exam' (*ton eafto tu* needs to be focused)

c. ?O Yiorghos ithele **ton** **eafto** **tu** na grapsi asxima
the Jorghos wanted the.ACC self.M.SG 3.M.SG.GEN subj write badly
sto djagonisma. Greek
in.the exam
'Jorghos wanted himself to do badly in the exam.'

The problem is, that such constructions also exhibit Principle A effects of the same type as we find in matrix clauses. The examples in (66) show that Principle A effects surface with unmodified *self*, whereas the data in (67) show its absence with modified *self*. This is the same observation that has been illustrated throughout this paper.

(66) a. *valji kidthu ke reena-ne **e-na** **potha-ne** sui ja-va-nu khaptu tu. KGu
Valji said that Reena-ACC 3.SG-GEN self-ACC sleep go-INF-GEN.N want aux
'Valji said that Reena wanted himself (Valji) to go to sleep.'

b. *O Yiorghos xeri oti i Maria perimene **ton eafto** **tu** na grapsi
Jorghos knows that Maria expected the self.M.SG 3.M.SG.GEN subj write
asxima sto djagonisma. Greek
badly in.the exam
'Jorghos knows that Maria expected himself(=Jorghos) to do badly in the exam.'

(67) a. valji kidthu ke reena-ne **e-na** **bimar** **potha-ne** sui ja-va-nu khaptu
Valji said that Reena-ACC 3.SG-GEN sick self-ACC sleep go-INF-GEN. N want
tu. KGu
aux
'Valji said that Reena wanted his sick self (Valji's sick self) to go to sleep.'

b. ?O Yiorghos xeri oti i Maria perimene **ton aksiolipito** **eafto** **tu** na
Jorghos knows that Maria expected the pitiful self.M.SG 3.M.SG.GEN subj
grapsi asxima sto djagonisma. Greek
write badly in.the exam
'Jorghos knows that Maria expected his pitiful self to do badly in the exam.'

The pattern in (66)–(67) is not predicted from this analysis. Given that we see Principle A effects, we would expect *self*-incorporation; however in ECM clauses this may appear not to be possible, for the anaphor and its referent are arguments of different verbs (the anaphor is the subject of the embedded clause, its antecedent is the subject of the matrix clause). Thus, it is initially not clear how *self*-incorporation could apply. We may expect (66) to be good, as *self*-incorporation should be blocked anyway. Assuming that *self* cannot incorporate into the matrix predicate it is not immediately obvious why there are Principle A effects in ECM constructions, as the analysis predicts its absence.

At this point, there are two possible solutions to this puzzle. Either Principle A in ECM constructions is a completely different phenomenon from Principle A in non-ECM clauses, or *self*-incorporation into the matrix predicate can occur in ECM clauses. I have argued for the former in the case of possessive reflexives, which do indeed look rather different from regular reflexives. However, in the case of ECM predicates, this does not seem motivated, as they are parallel to matrix predicates. Therefore, I thus pursue the second option.

As Büring (2005:229) argues, ECM clauses appear to behave as though the embedded subject did also count as an object argument for the matrix predicate. Such a view has been substantiated by Bruening (2001), based on Pasamaquoddy facts.[27] The idea is that ECM really involves *Raising-to-Object*, i.e. the embedded clause expresses a property (type <e,<s,t>>). I adopt his entry for 'know' to Kutchi Gujarati and Greek.

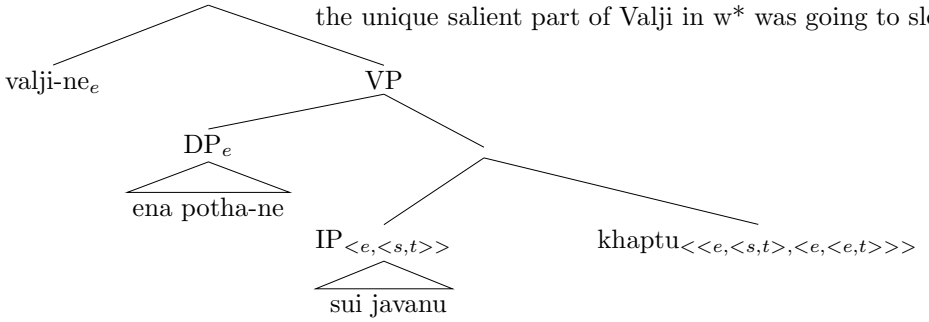
(68) a. $||\text{perimene/expected}||^{w} = \lambda P_{<e,<s,t>>}.\lambda x.\lambda y.$for all w' compatible with what y expected in w, P(x)(w')

b. $||\text{ithele/khaptu/wanted}||^{w} = \lambda P_{<e,<s,t>>}.\lambda x.\lambda y.$for all w' compatible with what y wanted in w, P(x)(w')

Without *self*-incorporation, we derive the truth conditions in (69b) from the above lexical entries (based on the structure in (69b)). We might now propose that *potha* once again incorporates into *khaptu* 'want' (leaving open the details of such an approach). The problem is, how *potha* 'self' (being of type <e,<e,t>) could compositionally combine with *khaptu* 'want' (being of type <<e,<s,t>>,<e,<e,t>>), as is illustrated by the diagram in (69b). Compositionally, there is no real issue for the semantic component, if we assume an alternative syntactic structure along the lines of Larsonian VP shells. The precise evidence for VP shell structures in Kutchi Gujarati and Greek exceeds the scope of this paper, and I leave it open for further research.

(69) a. valji-ne **e-na** **potha-ne** sui ja-va-nu khaptu tu. KGu
Valji-DAT 3.SG-GEN self-ACC sleep go-INF-GEN.N want aux
'Valji wanted himself to go to sleep (... but he found it hard to stop watching TV).'

b. ||CP|| = 1 iff for all w' compatible with what Valji wanted in w*, the unique salient part of Valji in w* was going to sleep.



7 Conclusion

The presence or absence of Principle A of the Binding Theory can be explained by a requirement at the syntax semantics interface. Principle A effects surface in the presence of unmodified *self*, where there is a requirement for the subject and object to be identical; this requirement triggers *self*-incorporation into the predicate. In contrast, modified *self*-incorporation is blocked in the syntax, giving rise to an asymmetric part-of relation. Given that *self*-incorporation is absent, we predict that Principle A effects do not surface, and this is exactly what we find.

Acknowledgements

This project has benefitted from numerous comments and discussions. I am particularly grateful to Patrick Grosz, Irene Heim and Sabine Iatridou for detailed comments and extensive discussions. I also wish to thank Adam Albright, Gennaro Chierchia, Robin Cooper, Alice Davison, Martin Everaert, Volker Gast, Martin Hackl, Daniel Hole, Barbara Partee, György Rákosi, Norvin Richards, Uli Sauerland, Rob Truswell, and Igor Yanovich. I would also like to thank the audiences at the semantics colloquium at the University of Frankfurt in 2009, and the audiences at the peculiar

[27] See also von Fintel & Heim's (2002:66–69) lecture notes. von Fintel & Heim do however not pursue or endorse such an analysis.

binding configurations workshop in Stuttgart in 2010. For Greek judgments and discussion regarding the data I thank Sabine Iatridou, Akis Kechagias, Marika Lekakou, Marios Mavrogiorgos, Dimitris Michelioudakis, Nina Topinzi, Vina Tsakali and Nikos Velegrakis. I would like to thank Shanta and Mavji Patel, Babita Seyani and Hema Nardani for the Kutchi Gujarati judgments.

References

Allen, Barbara, Donald Gardiner, and Donna Frantz. 1984. Noun incorporation in Southern Tiwa. *International Journal of American Linguistics* 52:292–311.

Anagnostopoulou, Elena and Martin Everaert. 1999. Toward a more complete typology of anaphoric expressions. *Linguistic Inquiry* 30(1):97–199.

Baker, Mark and Kenneth Hale. 1990. Relativized minimality and pronoun incorporation. *Linguistic Inquiry* 21(2):289–297.

Bruening, Benjamin. 2006. Discrepancies between projection and selection: Split coordination and raising to object in Passamaquoddy. University of Delaware.

Büring, Daniel. 2005. *Binding Theory*. Cambridge: Cambridge University Press.

Butt, Miriam and Tafseer Ahmed. 2011. The redevelopment of Indo-Aryan case systems from a lexical semantic perspective. *Morphology* 21(3):545–572.

Carlson, Greg. 1977. *Reference to Kinds in English*. Ph.D. thesis, University of Massachusetts, Amherst.

Chomsky, Noam. 1981. *Lectures on Government and Binding Theory*. Dordrecht: Foris.

Fox, Danny. 1999. Reconstruction, binding theory, and the interpretation of chains. *Linguistic Inquiry* 30(2):157–196.

Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33:63–96.

Gast, Volker. 2006. *The Grammar of Identity. Intensifiers and Reflexives in Germanic Languages*. Routledge.

Guerzoni, Elena. 2003. *Why Even Ask? On the Pragmatics of Questions and the Semantics of Answers*. Ph.D. thesis, MIT.

Heim, Irene and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Oxford: Blackwell.

Huang, C. T. James. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Ph.D. thesis, MIT.

Iatridou, Sabine. 1988. Clitics, anaphors and a problem for co-indexation. *Linguistic Inquiry* 19(4):698–703.

Jackendoff, Ray. 1992. Mme. Tussaud meets the Binding Theory. *Natural Language and Linguistic Theory* 10:1–31.

Kotzoglu, George. 2002. Greek 'ECM' and how to control it. In *Reading Working Papers in Linguistics*, vol. 6, pages 39–56.

Lewis, David. 1983. Survival and identity. In *Philosophical Papers Vol. 1*, pages 55–77. Oxford: Oxford University Press.

Lubowicz, Anna. 1999. Two views of Polish reflexives. In S. Bird, A. Carnie, J. D. Haugen, and P. Norquest, eds., *Proceedings of the Eighteenth West Coast Conference on Formal Linguistics (WCCFL XVIII)*, pages 337–350. Cascadilla Press.

Mistry, P.J. 1997. Objecthood and specificity in Gujarati. In J. Hill, P. Mistry, and L. Campbell, eds., *The life of language: Papers in linguistics in honor of William Bright*, pages 425–442. Berlin: Mouton de Gruyter.

Musan, Renate. 1999. Temporal interpretation and information-status of noun phrases. *Linguistics and Philosophy* 22:621–661.

Partee, Barbara. 2007. Compositionality and coercion in semantics: The dynamics of adjective meaning. In G. Bouma, I. Krämer, and J. Zwarts, eds., *Cognitive Foundations of Interpretation*. Royal Netherlands Academy of Arts and Sciences.

Patel, Pritty. 2010. Pronouns, anaphors and why Kutchi Gujarati binds like Modern Greek. Handout for the workshop Peculiar Binding Configurations, Stuttgart.

Quine, Willard Van Orman. 1960. *Word and Object*. The Technology Press of MIT and John Wiley & Sons.

Reinhart, Tanya and Eric Reuland. 1993. Reflexivity. *Linguistic Inquiry* 24(4):657–720.

Reuland, Eric. 2001. Primitives of binding. *Linguistic Inquiry* 32(3):439–492.

Reuland, Eric. 2005. Binding conditions: How are they derived? In S. Mueller, ed., *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar*. Stanford: CSLI Publications.

Reuland, Eric. 2011. *Anaphora and Language Design*. Cambridge: The MIT Press.

Reuland, Eric and Yoad Winter. 2009. Binding without identity: Towards a uniform semantics for bound and exempt expressions. In S. L. Devi, A. Branco, and R. Mitkov, eds., *Anaphora Processing and Applications: Lecture Notes in Computer Science*, pages 69–79. Berlin: Springer.

Rizzi, Luigi. 1990. *Relativised Minimality*. Cambridge: The MIT Press.

Saxon, Leslie. 1984. Disjoint anaphora and the Binding Theory. In Cobler, S. Mackaye, and M. Wescoat, eds., *WCCFL 3: Proceedings of the Third West Coast Conference on Formal Linguistics*, pages 242–251. Stanford: CSLI Publications.

Schneider-Zioga, Patricia. 1992. Exceptional case marking constructions in Modern Greek. University of Southern California.

Spathas, Giorgos. 2010. *Focus on Anaphora*. Ph.D. thesis, Utrecht University.

Spector, Benjaman. 2010. Topics in the semantics of interrogative clauses, Hand-Out 2.

Stechow, Arnim. 2008. Degrees are equivalence classes of individuals. Topics in Degree Semantics: 4 Lectures. Handout 1: Degrees.

von Fintel, Kai and Irene Heim. 2002. Lecture notes on intensional semantics (July/August, 2002). MIT.

von Fintel, Kai and Irene Heim. 2010. Intensional semantics (MIT spring 2010 edition). MIT.

Review of *Ibero-Asian Creoles: Comparative Perspectives*

Adam Roth Singerman, *University of Chicago*

Received November 2013

Bibliographic information:
Ibero-Asian Creoles: Comparative Perspectives. Edited by Hugo C. Cardoso, Alan N. Baxter, and Mário Pinharanda Nunes. Amsterdam: John Benjamins Publishing Company. 2012.

Introduction

Reading the eleven chapters in this volume — the proceedings of a conference held in October of 2010 at the University of Macau — has made two main impressions on me. The first is that the Spanish- and Portuguese-based creoles of Asia were born into and have developed within intricate linguistic ecologies, in which contact between genetically-unrelated and typologically-diverse languages has been the norm. The second point — perhaps a corollary to the first — is that research into these languages requires engaging head-on with multiple areas of scholarship. To tell a coherent story about these creoles, one must have extensive knowledge of the history of Spanish and Portuguese; enjoy deep familiarity with the grammatical structures of indigenous Asian languages; and pay detailed attention to the historical processes that have shaped contact between Iberian traders and colonizers, on the one hand, and various Asian populations, on the other. Fortunately, the chapters in *Ibero-Asian Creoles: Comparative Perspectives* largely rise to face these challenges. The result is a book which can hold appeal for many different audiences: creolists, typologists, specialists in Romance or Southeast Asian languages, and even historians interested in the interplay between linguistic and societal development.

Part of what makes research into the Ibero-Asian creoles so challenging is that the theoretical conceptions of Atlantic creolistics do not necessarily find natural counterparts in Asia. Even terms such as 'lexifier' and 'substrate' can be problematic, as Ian Smith points out in his essay on different sources of influence on creole word orders:

> The effect of the lexifier on creoles with which it remains in contact is well known from the Caribbean creoles. The substrates, however, have disappeared from the Caribbean, and only their early influence can be gauged. The Ibero-Asian creoles, on the other hand, remain in contact with their substrates, whose continuing influence in their role as adstrates must be considered. (p. 126)

As Smith points out, a single language can play different roles over the course of a given creole's development. But the editors, Hugo C. Cardoso, Alan N. Baxter and Mário Pinharanda Nunes, take this point further: in their words, "the substrate-adstrate opposition is often untenable" (p. 9) for the creoles discussed in this volume. They draw attention to a "circumstance which, though not exclusive of the Ibero-Asian creoles, is particularly typical of the Ibero-Asian creoles: the fact that they have coexisted for protracted periods of time with both their main lexifiers and various

JSAL Volume 6, December 2013.

adstrates" (pp. 8–9). Certain creoles in India, for instance, have enjoyed centuries of regular contact with both Portuguese and Gujarati/Marathi. Furthermore, as Cardoso explains in his study of comparative constructions, another "complicating factor . . . is the possibility that the various Luso-Asian Creoles establish relationships of progeny and/or diffusion" (p. 81). That is, as linguistic spread has been facilitated by the movement of goods and peoples across significant distances, creoles now spoken quite far from one another may share common historical roots and may have engaged in past contact. This point in turn adds another degree of complexity into an already packed ecology. Regardless of whether one wishes to salvage the substrate/adstrate/superstate distinction (or to maintain a sharp line between genetic and areal relationships), teasing apart the historical forces that have helped to shape the Ibero-Asian creoles is by no means straightforward.

This review discusses seven of the volume's eleven chapters, selected so as to represent the range of approaches visible in *Ibero-Asian Creoles*. I first address three chapters which tackle typological issues across creole languages, from both diachronic and synchronic perspectives. I then discuss two chapters which use variation within Spanish and Portuguese to better chart the provenance of words and particles that enjoy a wide areal distribution in Southeast Asian creoles. Finally, I discuss two chapters which examine how some Ibero-Asian creoles have interacted with other pidgins, creoles, and contact languages, and how they fit into a broader contact typology. This thematic division is largely expository, given that most of *Ibero-Asian Creoles*'s entries make use of the methodologies and findings of multiple subfields of linguistics. It is worth emphasizing that the four pieces not discussed here — J. Clancy Clements's 'Notes on the phonology and lexicon of some Indo-Portuguese creoles' (pp. 15–46), Baxter and Augusta Bastos' 'A closer look at the post-nominal genitive in Asian Creole Portuguese' (pp. 47–79), Eeva Sippola's 'Indefinite terms in Ibero-Asian Creoles' (pp. 149–179), and Carl Rubino's '*Bilug* in Zamboangueño Chavacano: The genericization of a substrate numeral classifier' (pp. 239–261) — are also interesting, insightful pieces of research. Their omission from this review is due to limitations of space only.

Historically-mindful typologies

Ian Smith's 'Measuring substrate influence: Word order features in Ibero-Asian Creoles' (pp. 125–148) examines and analyzes nine different morphosyntactic properties in one Spanish-based and seven Portuguese-based creoles of Asia. These are Zamboangueño Chabacano and Ternateño Chabacano (Spanish-lexified) and the creoles of Daman, Diu, Korlai, Sri Lanka, Malacca, Batavia, and Makista (all Portuguese-lexified). The properties under consideration include various ordering relationships on both the word- and morpheme-level: subject, verb, and object; possessor and possessum; adjective and noun; adposition and noun; demonstrative and noun; cardinal numeral and noun; relative clause and noun; degree word and adjective; and the position of interrogative phrases. For each property in each creole in the sample, Smith gives a score between −1 and 1, determined from the creole structures' proximity to the corresponding structures in the relevant substrates/adstrates. It is important to note that Smith himself calls these 'Substrate Influence Scores' "an unsophisticated overall measure of the penetration of substrate or superstrate word order patterns" (p. 143), and he acknowledges that his analysis makes sometimes crude distinctions when comparing structures across the different creoles under examination. The results, however, are still quite striking: the SIS ranking for an individual creole is an almost exact inverse of the historical presence of the Portuguese/Spanish languages in the locations where those creoles developed and are spoken. The creole spoken in Macau, Makista, has the highest SIS rank because it shows the greatest degree of typological affinity with Portuguese — and Macau was the very last of Lisbon's Asian holdings to undergo decolonization. On the other end of the spectrum, the Portuguese-based creole of Sri Lanka matches its substrates/adstrates, Sinhala and Tamil, throughout the surveyed morphosyntactic traits — and Sri Lanka was among the first of their Asian territories from which the Portuguese departed, in the mid-17th century.

Cardoso's essay, 'Luso-Asian comparatives in comparison' (pp. 81–123), examines comparative constructions in the creole languages of eight different sites of Portuguese colonization in Asia:

Diu, Daman, Korlai, Cannanore, Batticaloa, Malacca, Batavia/Tugu, and Macau. As in Smith's article, the number of creoles and substrates/adstrates to be examined is challengingly high; but Cardoso successfully compiles data from sources as varied as 19th-century documentation, recent description, and his own impressive array of fieldnotes. Most usefully, the chapter is organized by substrate/adstrate: Batavia/Tugu Creole and Malacca Creole are treated in the same section as Malay; the nearly-extinct Macau Creole immediately follows a discussion of comparative constructions in its adstrate, Cantonese; and so on. Following the typological classification of comparatives by Stassen (2008), the analysis classifies comparative constructions according to the kind of nominal case they impose upon the standard vis-a-vis the comparee and to the kind of morphosyntactic structures they utilize. In the discussion section, Cardoso explicitly compares the comparative constructions used by each of the eight creoles under discussion against the lexifier (Portuguese) and the relevant respective substrates/adstrates, and he provides an 'Index of Reliance on Lexifier' that weighs a given creole's Portuguese-like features against the features shared between Portuguese, on the one hand, and that same creole's substrate(s)/adstrate(s), on the other. The broader results point toward "a certain inverse correlation ... with time elapsed since break of significant contact with Portuguese" (p. 117): Portuguese-like comparative constructions are prominent in those creoles which were in contact with Portuguese for the longest amount of time. This finding closely mirrors Smith's conclusion concerning word order relationships; but, also like Smith, Cardoso acknowledges that his "generalizations must be taken as hints rather than as holistic, established facts" (p. 117).

I enjoyed both of these chapters and believe that their conclusions, however tentative, hold considerable promise for creolistics and for studies of language contact more broadly. My principal critique is one which the authors themselves directly acknowledge: comparing so many constructions across so many languages can lead to less-than-precise generalizations. In particular, Smith's typological classification suffers from the same challenge that impedes many broad typological surveys, namely, it is not clear how to weigh different properties against one another. More problematically, classificatory measures such as SVO and SOV are known to obscure other syntactic properties and to discount important considerations such as alternative word order configurations that relate to information structure. Regarding Cardoso's study, I found myself wondering about the different semantic properties of these comparative constructions. Although his attention is rightfully restricted to 'comparisons of superiority,' the constructions lumped together in Stassen's typology do not necessarily enjoy the same truth conditions; treating them as a natural class risks conflating distinctions that do not share the same meaning. (See Kennedy 2007 and Bochnak 2013 for discussion of this point.) As the two authors reach the same general conclusion – namely, that substrate influence on a given creole is inversely related to the duration and intensity of contact with the lexifier – I would like to see their chapters expanded and brought together into a more elaborated work. Should their hypothesis continue to accrue empirical support (as I suspect that it will), Smith and Cardoso will have provided us with a powerful new explanatory mechanism.

Pinharanda Nunes's 'Traces of superstrate inflection in Makista and other creoles' (pp. 289–326) sidesteps many of the methodological issues facing broader typological surveys. This chapter looks at Makista (Macau Creole Portuguese) and draws key comparative data from two other sources: the three closely-related Indo-Portuguese creoles of Diu, Daman and Korlai, and Kristang, or Malacca Creole Portuguese, from which Makista descends. While Makista had been described by earlier sources as possessing verbal morphology based on Portuguese infinitives and third person forms, Pinharanda Nunes shows that the language actually "presents a wider range of superstrate morphology than previously reported" (p. 319). He carefully and helpfully walks the reader through this argument: after outlining present, past perfective, and imperfective past verbal morphology in Portuguese, he provides explicit 'identification criteria' for such structures in Makista and then describes their distribution in his oral corpus of Makista. The data from the oral corpus are compared against 19th- and 20th-century written corpora, and the relevant structures in Kristang and the three Indian creoles are also examined. The ample sociohistorical discussion crucially contextualizes Makista within the linguistic ecology of Macau and the surrounding region: population movements

in Southeast Asia brought various ethnic groups to new locations and allowed for language contact and spread to take place across a wider and more diverse space. For substrate speakers, Pinharanda Nunes argues, there was "habitual inaccessibility ... to the ruling European minority's language", and multilingualism in pidgins, creoles, and Asian languages "allowed for local in-group models to emerge as the TL [Target Language] instead of Portuguese" (p. 316). Drawing on language-internal and broader historical evidence, Pinharanda Nunes concludes that the growth of superstrate verbal inflection in Makista represents a process of gradual decreolization; and, I would add, this process of decreolization cannot be divorced from the sociolinguistic setting of Macau within the region and the broader Lusophonic world. I very much appreciated the explicit precision and clear assumptions of this chapter, which provides a natural complement to the wider typological surveys featured elsewhere in this volume.

Variation within and across lexifiers and adstrates

Mauro A. Fernández's '*Nenang, nino, nem não, ni no:* Similarities and differences' (pp. 205–237) traces the development of the Kristang negation particle *nenang*, as well as other negators in Southeast Asian creoles and adstrates. The author argues that the Kristang particle developed directly out of the Portuguese *nem não*, which remains in use of some varieties of Portuguese and has been historically documented. The Spanish equivalent, *ni no*, in turn has entered Zimboanga Chabacano and other creoles of the Philippines. Part of what makes Fernández's argument so interesting is the fact that *nem não* and *ni no*, which translate (awkwardly) as 'neither not,' have fallen out of common use in the European varieties of Spanish and Portuguese but are well-attested in the varieties of particular ex-colonies. The bigger puzzle is how the creole forms acquired their meaning of 'not yet,' given that the original Romance meaning is akin to 'neither' or 'not even.' Fernández provides evidence that the present meaning of *nenang* in Kristang is due in part to relexification from Malay, whose negative particle *belum* could have shaped the semantics of the creole form. The extensive discussion section at times seems to raise more questions than it can answer, and it reads in parts as reliant upon conjecture. But if this chapter — which addresses deep questions regarding the genesis of the Spanish-based creoles of the Philippines — cannot claim to fully resolve the diachronic semantics of *nenang* and its relatives, it does succeed in shining "some light into this small, unstudied corner of the history of Spanish and Portuguese" (231).

Similarly, Nancy Vázquez Veiga and Fernández's chapter, '*Maskin, maski, masque...* in the Spanish and Portuguese creoles of Asia' (pp. 181–203), reiterates the sheer difficulty of tracing etymologies in the Ibero-Asian creoles. The authors challenge the common assertion that the *maski/maskin/masque* particle, now present in a host of Spanish-based creoles, must have descended from Portuguese; they instead argue that Spanish independently provided this form to the creoles of the Philippines. They show that Spanish also possessed a concessive or modal *mas que*, and that the descendant form *maskin* in Chabacano, a creole of the Philippines, has retained much of the original Spanish concessive and modal meanings. Furthermore, *maskin* has "acquired from the Philippine languages a new scalar or intensifying function, in addition to that of focal or indefinite quantifier" (p. 191). So the creole *maskin* has come to combine semantics from both Spanish and indigenous languages of the Philippines; and *maskin* has also entered the lexicons of many of those same indigenous languages. Did Spanish loan this particle directly to the native languages, or did a creole serve as an intermediary? Was transmission even more complicated, with some native languages acquiring the particle from the Spaniards and others from speakers of creole? The contact-abdundant linguistic ecology of the Philippines makes tracing the diachronic development of this particle (and perhaps any other lexical entry) extremely difficult. Yet Vázquez Veiga and Fernández show that progress can indeed be made on this front; and their chapter serves to remind the reader that to unravel the histories of creole languages, one must command a solid knowledge of the histories of their lexifiers.

That being said, a comment about this chapter's tone is in order. The opening pages challenge Keith Whinnom's mid-twentieth century work on the creole languages of Southeast Asia, in particular his theory that a Portuguese-Malay pidgin played a role in shaping Philippine creoles. This

part of the discussion reads as unfairly aggressive and disparaging; three times on pages 182–183, Whinnom is said to have 'failed' to draw some conclusion or connection that Vázquez Veiga and Fernández consider obvious. As the authors explicitly mention how the resources of the Internet have given them an "enormous advantage" by making their research into the diverse languages of the Philippines "far easier" (p. 197), more gracious criticism of those scholars who lacked these tools would seem warranted.

Creoles and language contact

Stephen Matthews and Michelle Li's 'Portuguese pidgin and Chinese Pidgin English in the Canton Trade' (pp. 263–287) examines the role of Portuguese-based pidgin in shaping the contact language of Chinese Pidgin English. Consulting phrasebooks that Chinese speakers used to learn pidgin, they show that "Portuguese and English lexical items coexisted for some time in the China trade" (p. 271). In terms of grammar, they argue that two oddities of Chinese Pidgin English — the uses of *have* as a copula and of *for* as a non-finite complementizer — "cannot be readily explained on the basis of English or Cantonese" (p. 280). Crucially, however, Portuguese can and does provide a ready template for these very constructions: the verb *ter* has been assuming the functions of an existential copula for centuries (and has largely supplanted the older *haver* in Brazilian Portuguese), and the preposition *para* regularly introduces non-finite clauses. Based on the lexical and grammatical evidence, they conclude — albeit tentatively — that Portuguese pidgin has played a greater role than previously recognized in the development of Chinese Pidgin English. Their evidence, even if limited, is persuasive, and it raises questions concerning the phylogeny of other creoles and pidgins. This is a topic worthy of continued investigation.

The volume's closing chapter, Anthony P. Grant's 'Mindanao Chabacano and other "mixed creoles": Sourcing the morphemic components' (pp. 327–364), aims to contextualize Mindanao Chabacano within a broader understanding of 'mixed creoles' and language mixing. Grant looks at how Chabacano has acquired over 10% of its Swadesh List lexicon from sources other than its chief lexifier, Spanish (with the 10% threshhold used to define mixed-lexifier creoles), and compares this and several other mixed-lexifier creoles' broader properties against typologies of contact and language mixing. Particularly interesting is the discussion that frames these mixed-lexifier creoles against the best-known cases of mixed languages: Ma'á (which brings together Cushitic and Bantu), Media Lengua (Spanish and Quechua), Mednyj Aleut (Russian and Aleut), and Michif (French and the Algonquian language Plains Cree). Grant argues that, whereas "mixed languages use (somewhat regularised, less allomorph-heavy and scaled-down) versions of sets of their contributory languages' inflectional (and often derivational) morphology," creolization is fundamentally different: "creators of creole languages construct new morphological systems over time . . . drawing on typological blueprints provided by their substrate languages" (p. 346). One wonders how this distinction will need to be qualified in light of new evidence concerning Australian mixed languages, such as Light Warlpiri (O'Shannessy 2013) and Gurindji Kreol (Meakins 2011), whose speakers appear to have happily innovated new structures. In some respects, Grant's survey comes across as more compilation and comparison than synthesis. Yet his findings — that "[t]here is only a rather weak correlation between the amount of borrowed basic lexicon in a mixed-lexifier creole and the proportion of borrowed structural features and function words" (p. 355) and that "[m]ixed lexifier creoles do not constitute anything more than a weakly defined class . . . as opposed to less mixed lexifier creoles" (p. 356) — are interesting and worthwhile *precisely because* they are hedged. They point toward the conclusion that the languages we call creoles, as but one instantiation of a much broader class of contact languages, form a highly heterogenous group (if they form any group at all!). That these languages' historical development and synchronic composition resist easy characterization and unified treatment reiterates the need for an approach that does not take 'creole' as any kind of primitive. Similar points have been made, for example, by Mufwene (2008).

Conclusion

It is to the authors' and editors' credit that *Ibero-Asian Creoles: Comparative Perspectives* — which is full of data from many different languages, addresses regional issues across an enormous territory, and examines evidence stretching back over the last half-milennium — coheres so successfully. Even though the eleven chapters tackle different questions and draw independent conclusions, the overall narrative is both internally consistent and thought-provoking. I hope and trust that this volume will generate more interest in the Ibero-Asian Creoles and will inspire other scholars to research these languages' historical development and present-day structures. In addition, I look forward to seeing extended versions of the cross-creole typologies presented here. If the insights and findings discussed in this book come to inform research on creoles and contact languages more broadly, the field of linguistics will surely benefit.

Acknowledgements

Thanks go to Hugo C. Cardoso and Salikoko Mufwene for helpful comments on this review.

References

Bochnak, Ryan. 2013. *Cross-linguistic variation in the semantics of comparatives*. Ph.D. thesis, University of Chicago.

Cardoso, Hugo C., Alan N. Baxter, and Mário Pinharanda Nunes, eds. 2012. *Ibero-Asian Creoles: Comparative Perspectives*. Amsterdam: John Benjamins.

Kennedy, Chris. 2007. Modes of comparison. In M. Elliott, J. Kirby, O. Sawada, E. Staraki, and S. Yoon, eds., *Proceedings of the 43rd Meeting of the Chicago Linguistics Socieyt (CLS 43)*. Chicago: Chicago Linguistic Society.

Meakins, Felicity. 2011. *Case-marking in contact: the development and function of case morphology in Gurindji Kriol*. Amsterdam: John Benjamins.

Mufwene, Salikoko S. 2008. *Language evolution: Contact, competition and change*. London: Continuum Press.

O'Shannessy, Carmel. 2013. The role of multiple sources in the formation of an innovative auxiliary category in light warlpiri, a new australian mixed language. *Language* 89(2):328–353.

Stassen, Leon. 2008. Comparative constructions, feature/chapter 121. In M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie, eds., *The World Atlas of Language Structures Online*. München: Max Planck Digital Library.