*Monsters, agreement and anaphora: evidence from Tamil*

# Monsters, agreement, and anaphora: evidence from Tamil

Sandhya Sundaresan

Universitetet i Tromsø (CASTL)/Universität Stuttgart

## 1. Overview

Consider the sentence in (1) below from Tamil, an SOV language of the Dravidian family:

(1) raman$_i$ [$_{CP}$ taan$_{\{i,*j\}}$ jey-pp-een/*aan-nnu] so-nn-aan
raman-NOM SE-NOM$_i$ win-FUT-1SG/*3MSG-that say-PST-3MSG
"Raman said [$_{CP}$ that he$_{\{i,*j\}}$ would win]"
Obligatory *de se* reading: 'Raman said, "I will win"'

The structure in (1) contains a speech predicate *sonnaan* (told.3MSG) which embeds a (finite) clausal complement. This CP complement has a subject *taan*[1] – traditionally described as a nominative, 3SG. (optionally) long-distance (or SE- (Reinhart and Reuland 1993)) anaphor (Annamalai 1999).[2] *taan* is obligatorily coreferent with the matrix (3rd-person) subject and attitude-holder, *raman* and is, crucially, interpreted obligatorily *de se* with respect to its DP antecedent. Peculiarly, however, the embedded indicative verb *jey.pp.een* (will-win.1SG) agrees with the matrix 3MSG subject *Raman* and, it also doesn't *seem* to agree in person-feature with its own clausemate subject *taan*.[3]

*I thank Pavel Caha, Gennaro Chierchia, Alessandra Giorgi, Angelika Kratzer, and Peter Svenonius for taking the time to meet with me and discuss aspects of this paper – and Ken Safir for additionally sharing generously of unpublished work and curious binding data. I also thank the audiences at the Local Pronouns Workshop (Nijmegen), GLOW-in-Asia (Beijing), Peculiar Binding Configurations Workshop (Stuttgart), NELS 41 (Philadelphia), and the Grammar of Attitudes Workshop at DGFS (Göttingen), where versions of this were presented, for their feedback, and my Tamil native-speaker informants for confirming the existence of monsters in Tamil. I am especially grateful to Thomas McFadden and Gillian Ramchand for extensive meetings and discussions of the fledgling to final forms of this paper. Any errors are mine alone.

[1]Crucially, *taan* does indeed seem to be an embedded subject in such stuctures and not, for instance, a focus adjunct reflexive on the matrix subject DP. Emphasis of the form "Raman himself ..." in Tamil is created by means of a focus clictic *-ee* which attaches to the focussed DP.

[2]The glossing of *taan* as "SE" should thus be treated as purely descriptive for now.

[3]This, of course, is not a foregone conclusion since it is not clear what $\phi$-features, if any, anaphors have, to begin with (Kratzer 2009). I return to this point later.


Emily Elfner and Martin Walkow (eds.), NELS 41: 2–15

It has been proposed (Woolford 1999) that the (apparent) lack of person agreement between the embedded subject and verb in (1) is a grammatical conspiracy to avoid a violation of the Anaphor Agreement Effect/AAE – the idea, going back to Rizzi (1990), that anaphors are crosslinguistically disbarred from agreement(-triggering) positions. The problem for this hypothesis, however, is that when the same subordinate clause in (1) is embedded under the scope of a non-speech matrix verb like *kaɳḍupiḍi.tt.aan* ('find out') "real" agreement is not only possible, it is obligatorily required, as illustrated in (2):

(2) raman$_i$ [$_{CP}$ taan$_{\{i,*j\}}$ jey-pp-aan/*een-nnu] kaɳḍupiḍi-tt-aan
raman-NOM$_i$ SE-NOM$_i$ win-FUT-3MSG/*1SG-that find.out-PST-3MSG
"Raman$_i$ found out [$_{CP}$ that he$_{\{i,*j\}}$ would win.]"

Crucially, furthermore, this embedded 3MSG agreement may only surface with a 3MSG clausemate subject, thus cannot be dismissed as mere default agreement (3):

(3) seetha$_i$ [$_{CP}$ taan$_{\{i,*j\}}$ jey-pp-aaḷ/*aan-nnu] kaɳḍupiḍitt-aaḷ
seetha SE win-PST-3FSG/*3MSG-that find.out-PST-3FSG
"Seetha$_i$ found-out [$_{CP}$ that she$_{\{i,*j\}}$ would win]."

I thus propose, contra Woolford (1999) that structures in Tamil like (1) are not a function of the AAE. But then, what are the differences in agreement patterns in the embedded clauses in (1) and (2) due to? And how can this be related to the difference in syntactico-semantics of the embedding predicates in each?

## 2. A first inkling: similarities to indexical shift

The matrix predicate in (1), *sonnaan* (said.3MSG) is a reportive predicate; its clausal complement thus denotes a speech-report. The matrix predicate in (2) *kaɳḍupiḍittaan* (found-out.3MSG) is, however, a non-reportive predicate and its clausal complement does not constitute a speech report. Now consider the shifted indexical structure from Zazaki, an Indo-Iranian language, below (Anand and Nevins 2004, Anand 2006):

(4) hɛseni$_j$ (mɨ$_k$-ra) va kɛ ɛz$_{\{j,k\}}$ dɛwletia
Hesen-OBL I-OBL-TO said that I rich-be-PRES
"Hesen said that {I am, Hesen is} rich."

In (4), the indexical pronoun *ɛz* (I) can refer to the speaker or, anomalously, to the attitude-holder *hɛsen*. In other words, the embedded indexical pronoun is referentially ambiguous – it is a *shifting indexical* or, in Kaplanian parlance, a *monster* (Schlenker 2003, von Stechow 2002, Anand 2006). Significantly for the current discussion, shifted indexicality typically only occurs in contexts that are under the scope of reportive/speech-predicates. This provides us with an initial clue as to what might be going on in structures like (1) in Tamil. Just as in the Zazaki structure in (4), the 1st-person marking on the embedded verb in (1) refers to the (masculine, 3rd person) attitude-holder (*Raman*), and not to the speaker, as is standard. Crucially furthermore, this 1st-person marking appears to be induced under conditions that are identical to those that effect classic shifted indexicality, as in (4) – specifically, under conditions of embedding by a reportive/speech predicate. In Tamil, the

following speech-predicates appear to behave similarly to *soll* (say) in (1) above in requiring 1SG agreement on the embedded verb: *kattu* (shout/scream), *arivi* (inform), *pukaarsey* (advertise), *oppukkoḷ* (admit):[4]

(5) raman$_i$ [$_{CP}$ taan$_{\{i,*j\}}$ tiruḍi-n-een/*aan-nnu] ottukko-ɳḍ-aan
raman SE$_i$ steal-PST-1SG/*3MSG-that admit-PST-3MSG
"Raman admitted [$_{CP}$ that he$_{\{i,*j\}}$ stole/had stolen]"

Based on the above, I thus make the preliminary proposal that the structure in (1) involves a type of indexical shift.

## 3. Eliminating the usual suspects

Before embarking on an investigation of shifted indexicality in structures like (1), it is important to ascertain that what we're dealing with is actually an instance of indexical-shift. In this spirit, I discuss some of the usual suspects (see among others Schlenker To appear, for more discussion) – full-on quotatives and mixed/partial quotatives – candidates that may emulate the effects of monsters without actually being underlyingly monstrous.

### 3.1 Against a quotative analysis

Quotations may emulate shifted-indexicality effects because they "form a closed domain with respect to syntactic and semantic operators" (p. 81 Anand 2006).[5] This grammatical opacity property of quotatives itself yields some useful diagnostics for testing the existence of quotatives: an NPI inside a quotative may not be licensed by an operator outside the quote (6); *wh*-extraction outside of a quoted domain is expected to be ungrammatical (7):

(6) raman$_i$ [$_{CP}$ taan$_{\{i,*j\}}$ yenda tappu-m senjeen-nnu] ottukka-le
raman SE which mistake-even made-that admit-NEG
"Raman$_i$ didn't admit that he$_{\{i,*j\}}$ made any mistake."

(7) krishnan$_i$ yaar-ai$_1$ [$_{CP}$ taan$_{\{i,*j\}}$ t$_1$ paa-tt-een-nnu] so-nn-aan
krishnan who-ACC$_1$ SE t$_1$ see-1SG-that say-PST-3MSG
"Who(m)$_1$ did Krishnan$_i$ say [$_{CP}$ that he$_{\{i,*j\}}$ saw t$_1$]?"

(6) and (7) are both grammatical, showing that the embedded CP in these structures is not a full clausal quotative. For the sake of completeness, I present the minimal pairs to (6) and (7) below – the result of replacing embedded subject *taan* with the 1st-person indexical

[4] At the same time, non-speech attitude-predicates like *nene* (think/believe) don't seem to induce such shift. This is as expected if only (a proper subset of) speech-predicates may effect indexical shift (Schlenker 2003).

[5] To see this, consider:

i. Marie said, "I have a terrible hangover!"
$[\![I]\!] \rightarrow \{Marie, *Auth_{utterance-context}\}$

Although the indexical pronoun *I* obligatorily refers to the attitude-holder *Marie*, the structure in (i) is not an instance of indexical-shift, but one of full-on clausal quotation.

*naan* ("I"):[6] The structures in (8) and (9) are ungrammatical, indicating that the embedded CP *is* a full-on quotative in these cases.

(8) * raman$_i$ [$_{CP}$ naan$_i$ yenda tappu-m se-nj-een-nnu] ottukka-le
raman I which mistake-even make-PST-1SG-that admit-NEG
"*Raman didn't say "I made any mistake."

(9) * krishnan$_i$ yaar-ai$_1$ [$_{CP}$ naan$_i$ t$_1$ paa-tt-een-nnu] so-nn-aan
krishnan who-ACC$_1$ I t$_1$ see-PST-1SG-that say-PST-3MSG
"*Who(m)$_1$ did Krishnan$_i$ say "I saw t$_1$?"

It is, nevertheless, still possible that only a sub-part of the CP (specifically perhaps the part with the anomalous agreement) is actually quoted. Unlike full on quotatives of the kind discussed above, partial quotatives *can* influence the truth-conditional semantics of the larger context they occur in. However, in theory, any string, including a nonsense string, could be quoted. Thus, if the embedded CP in (1) did involve a partial quotative, we wouldn't expect 1SG embedded verb-agreement as in (1) to be restricted to clauses embedded under reportive predicates – showing that the embedded CP in (1) does not involve a partially quoted string either.

### 3.2 An important difference from (classical) indexical shift

The discussion above suggests that structures like (1) involve bonafide indexical in the embedded clause. At the same time, (1) shows one important difference from standard structures involving indexical shift. In the "classic" indexical structure in Zazaki (1), the morphosyntactic agreement mismatch, between the shifted indexical and the attitude-holder DP that it denotes, is cross-clausal. In (1), on the other hand, in addition to the cross-clausal person-feature mismatch between 1SG *-een* on the embedded verb and 3MSG matrix subject/attitude-holder (*Raman*), one could argue that there is a person-feature mismatch within the embedded clause itself. How? Because the embedded subject DP *taan* may only take 3rd-person antecedents and is, as such, traditionally treated as a 3SG anaphor. Its clausemate verb is, however, marked 1SG. To make matters even worse, and most problematically, this verbal suffix *-een* is an agreement marker! Under standard Minimalist assumptions of Agree for subject-verb agreement, verbal agreement is held to be inherited by the T head from the subject (in [Spec, TP]) and is not inherent to T. But the clausemate subject *taan* itself eems to show different person-features (3rd) from that on the verb (1st) – so where does this 1st-person agreement get inherited from?

Before proceeding further, however, a crucial caveat is in order. Notice that it is far from clear what $\phi$-features, if any, *taan* possesses precisely because it is anaphoric. Under the fairly common view (Kratzer 2009, Heinat 2008) that anaphoric DPs are born with

[6]When such operations are not performed, the sentence is fully acceptable:

i. * raman$_i$ [$_{CP}$ naan$_i$ yenda tappu-m seyya-le-nnu] so-nn-aan
raman I which mistake-even made-NEG-that admit-PST-3MSG
'Raman said "I didn't make any mistake."'

deficient or zero $\phi$-features and inherit these from their antecedents by some mechanism of feature transmission (e.g. Agree) in the course of the derivation, it is definitely premature to classify *taan* as a 3SG anaphor. By extension, it is also premature to talk about intra-clausal agreement mismatches between *taan* and a clausemate verb. Nevertheless, the agreement patterns in the embedded CPs in (2) and (1) do vary: the 3MSG embedded verb agreement in the former faithfully reflects the $\phi$-features of the 3MSG matrix subject *Raman* on the surface, whereas the 1SG in the latter doesn't. In both (1) and (2), furthermore, the same surface form of the DP *taan* shows up as the embedded subject and the clausemate verb in each presumably Agrees with *taan* in both structures, since Tamil is a uniformly subject-agreement language. This is thus a pattern that is definitely puzzling independent of the assumptions we might make about the $\phi$-features inherent to *taan*.

To summarize, the puzzle on our hands is as follows: what enforces 1st-person marking on the embedded predicate in (1) and 3MSG agreement marking on the embedded verb in (2)? Is it simply the case that there are two underlyingly different DPs both of which are spelled out syncretically as *taan*, or is there a more complex and potentially more interesting explanation? I investigate these issues in the following sections.

## 4. Problems with prior analyses

In the following sections, I consider three predominant analyses of indexical shift in the literature – those in Anand (2006), von Stechow (2002), and Schlenker (2003) – and demonstrate that the Tamil agreement patterns presented above cannot be (non-stipulatively) dealt with under these analyses.

### 4.1 Indexical shift due to context-overwriting: Anand (2006)

Consider again the case from Zazaki (4), repeated as (10) below:

(10) hɛseni$_j$ (mɨ$_k$-ra) va kɛ ɛz$_{j/k}$ dɛwletia
Hesen-OBL I-OBL-TO said that I rich-be-PRES
"Hesen said that {I am, Hesen is} rich."

One way to understand the different readings obtained above might be to claim the following. In the case of the conventional, unshifted reading, the AUTHOR function of the indexical ranges over the utterance context whereas the shifted reading obtains when the same AUTHOR function ranges over the context introduced by the speech predicate. The central intuition behind Anand (2006)'s approach is that the difference between the shifted and unshifted readings above is not the denotation of the indexical *ɛz* ('I'), which remains constant (= $\lambda c.Author(c)$) – but the value of the context that the AUTHOR function ranges over. In his account, shifted indexicality is thus taken to be the result of context shifting due to context-overwriting, argued to be implemented by parametrized operators introduced by the speech predicate and presumably located in C. Anand proposes that the operator responsible for context-overwriting in Zazaki is: $OP_{\forall}$; $[\![OP_{\forall}\alpha]\!]^{c,i} = [\![\alpha]\!]^{j,i}$ where j

= <$Auth(i), Addr(i), Time(i), World(i)$>. The shifted reading for (4) is then derived thus:[7]

[Hesen said [$_{CP}$ $OP_{\forall}$ I am rich]]$^{c,i,g}$
= [Hesen said [$_{CP}$ I am rich]]$^{i',i',g}$
= 1, iff [Hesen said [$_{CP}$ AUTHOR(i') am rich]]$^{i',i',g}$
= [Hesen said [$_{CP}$ I$_{hesen}$ am rich]]$^{i',i',g}$ (READING B: SHIFTED)

### 4.2 Problems with extending Anand (2006)'s analysis to Tamil

The essence of Anand (2006)'s analysis of Malayalam *taan* is that it is underlyingly a 1st-person indexical – thus, ultimately, a shifted indexical since it refers to the attitude-holder, not the utterance-context speaker. We can apply his context-overwriting approach to the Tamil data in (1) and (2). But there are three main problems with such an analysis.

The first major problem for a context-overwriting approach is that Tamil as well as Malayalam *taan* only take 3rd-person antecedents (11 vs. 12):

(11) raman$_i$ [$_{CP}$ taan$_{\{i,*j\}}$ jey-pp-een-nnu] sol-r-aan
raman SE win.FUT-1SG-that say-PRES-3MSG
Raman says, "I will win."

(12) * naan [$_{CP}$ taan jey.pp.een-nnu] sol.r.een
I SE win-FUT-1SG-that say-PRES-1SG
I say, "I will go." (Intended)

This antecedence restriction is unexpected if Malayalam and Tamil *taan* are indeed 1SG indexicals. Under Anand (2006)'s context-overwriting approach, the way to deal with this would be to claim that a speech-predicate like SAY in Tamil (and Malayalam) may introduce a context-shifting operator only if it is itself specified as 3rd-person.[8] But such a condition, while workable, is, as Anand himself concedes, very stipulative. The second major problem under a context-overwriting approach is how to account for "regular" embedded verbal agreement as in (2) (a problem that Anand (2006) doesn't deal with for Malayalam, again because it lacks verbal agreement.) In his system, we are essentially forced to claim that *taan* in structures like (2) is underlyingly different from that in (1). E.g. that it is not an indexical, but a logophor/SE-anaphor, which inherits the $\phi$-features of its antecedent and transmits these via Agree to the embedded verb (Kratzer 2009)). A potentially serious empirical challenge to this type of analysis is that structures like 2 also seem to be interpreted as *obligatorily de se* – there is no clear independent reason to assume that the syntactico-semantics of *taan* in (2) is different from that in (1). Finally, on top of the obligatorily shifting *taan*, Tamil has a purely "English-like" 1SG indexical *naan* (see also (8) and (9) above):

[7] **Unshifted Reading:** [Hesen said [$_{CP}$ I$_{sandhya}$ am rich.]]
There is no context overwriting; the indexical *I* is interpreted against the utterance context (perhaps introduced at root C) and is thus mapped onto the utterance speaker = me, Sandhya (in this utterance context).

[8] The other option would be to claim that *taan* is not an indexical but a 3rd-person logophor/anaphor after all, but the 1SG agreement on the embedded verb in Tamil (1) seems to independently argue against this.

(13) raman [$_{CP}$ naan jey-pp-een-nnu] so-nn-aan
raman I win-FUT-1SG-that say-PST-3MSG
"Raman said [$_{CP}$ that $I_{utt-speaker}$ will win]" (INDIRECT SPEECH)
"Raman said, "I will win" ". (QUOTATIVE/DIRECT SPEECH)

If indexical-shift obtains due to context-overwriting at LF, as Anand proposes, it is unclear how the morphophonological distinction between *naan* and *taan* is to be captured. Of course, one could postulate different spell-out rules for the unshifted indexical *naan* and the shifted indexical *taan*, as follows:[9]

UNSHIFTED: [AUTHOR (c)]$^{c,i,g}$ $\rightarrow$ SPELL-OUT: *naan*
SHIFTED: [$OP_{auth}$ AUTHOR (c)]$^{c,i,g}$ = [AUTHOR (i)]$^{i,i,g}$ $\rightarrow$ SPELL-OUT: *taan*

While this is certainly doable, it requires the context-shifting operator to already be present in Narrow Syntax (in order to affect both the semantics at LF and Spell-Out rules at PF). The syntacticization of context-shifting/overwriting, however, still needs to be precisely formalized in Anand's system which deals predominantly with the semantic aspects of indexical shift. This is thus perhaps more a problem of omission, in contrast to Problems I and II above which, I believe, pose a more serious challenge to Anand's analysis.

### 4.3 Variable binding and LF feature-deletion: von Stechow (2002)

Descriptively, shifted indexical pronouns show an inherent split along the semantic/morphology domains: Semantically, they denote 3rd person individuals (the attitude-holder). Morpho(syntactically), however, they are marked 1st person. von Stechow (2002) proposes that shifted indexicals are variable bound by attitude verbs (which, he proposes, quantify over contextual tuples: ⟨Author, (Addressee), Time, World, Location⟩). Furthermore, he proposes, the $\phi$-features of a DP that is variable bound are deleted at LF.

Such a model neatly accounts for the person-feature split noted in shifted indexical DPs, as follows. The indexical pronoun is born with a [+author] feature – this is the feature that it bears in the Narrow Syntax. At PF, this 1st-person feature is spelled out. At LF, the indexical DP is variable bound by the quantificational attitude verb that c-commands it – this ensures that the indexical denotes the attitude-holder. Crucially, by assumption, when the embedded indexical pronoun is thus variable bound, its [+author] feature is deleted (thus preventing a presupposition failure/clash). This is formalized below:

SENTENCE: $\text{Raman}_i$ says that $\text{he}_i$ will win. (*De se* reading: Raman says, "I will win".)
[Raman 3rd] $\lambda$ i [say 3rd] $\lambda \langle x_j{}^3\text{rd}, z_k, w_i, t_i \rangle$ [$x_j$ + ~~author~~] will win.

What happens when we apply this analysis to the Tamil data in (1)/(2) above? First of all, the 3rd-person antecedence restriction of the embedded DP *taan* in both (1) and

[9] I extend Anand (2006)'s proposal, that the context-shifting operator in Malayalam is $OP_{auth}$ (an operator which shifts only the Author coordinate of the context-tuple), to Tamil which behaves very similarly.

(2) would have to be reframed as a restriction on the quantificational attitude-verbs binding them. Specifically, we would have to stipulate that only quantificational attitude-verbs specified 3rd-person may delete and thus variable-bind [+author] elements. More problematically, under this model, there is no way to distinguish between non-shifting *naan* (13) and shifting *taan* ((1)/(2)), assuming that both are [+author] elements in the syntax, again because variable binding is held to occur only at LF. Finally, assuming that *taan* is the same underlying element, specifically an [+author] indexical pronoun, in both (1) and (2), it is unclear how to derive the distinction between agreement patterns (1st vs. 3rd) on their corresponding clausemate verbs. Thus, we end up with essentially the same problems as with Anand (2006)'s analysis.

## 5. Pronoun-centric model: Schlenker (2003)

Schlenker (2003) (and subsequent) attempts to derive indexical shift as a function of the syntactico-semantics of the indexicals themselves. Simply put, some indexical pronouns, like those in English, are specified as being rigid/unshifting. Thus, English $[\![I]\!]^{c,g}$ = $[\![x_i + author*(x_i)]\!]^{c,g}$ = $s(x_i) \wedge \delta(g(x_i)$ is Author(c*)) (for c* = $c_{utterance}$). Others, like those in Amharic and Zazaki are underspecified as to whether they are always shifting or rigid – in other words, they can either shift or not shift. Thus, Amharic $[\![I]\!]^{c,g}$ = $[\![x_i + author(x_i, c_j)]\!]^{c,g}$ = $s(x_i) \wedge \delta(g(x_i)$ is Author($c_j$)). Still others, like logophoric pronouns in e.g. Gokana and other languages, are semantically specified to always shift. Thus, logophoric $[\![I]\!]^{c,g}$ = $[\![x_i + author(x_i, c_j)]\!]^{c,g}$ = $s(x_i) \wedge \delta(\forall\ c_j \neq c^* \rightarrow g(x_i)$ is Author($c_j$)).

Since much of the action is localized to the denotations of the indexicals themselves, Schlenker's model would be able to capture the distinction between unshifting *naan* (13) and shifting *taan* in Tamil (1)/(2). Specifically, *naan* would have the denotation of an English-like "I" whereas *taan* would have the denotation of a logophoric "I" – as formalized above.[10] I take this to be an advantage of Schlenker's system over both Anand's and von Stechow's. Nevertheless, the other two problems mentioned above remain for Schlenker's analysis as well. That is, it is unclear how to derive the distinction between monstrous agreement in (1) and "regular" agreement in (2) given that both are presumably triggered by the same underlying indexical pro-form *taan*. It is also not apparent how to derive the 3rd person antecedence restriction of *taan* in both (1) and (2).

To sum up, it seems clear that the predominantly semantic analyses presented above are not (at least as yet) capable of deriving the complex agreement patterns evidenced in the Tamil structures under scrutiny. In the following sections, I outline an analysis of the Tamil patterns in terms of dual access to utterance and speech contexts for person and additionally postulate distinct patterns of underspecification for DPs and verbal agreement under a Late Insertion model to yield the correct agreement paradigms in (1) and (2).

## 6. An alternative analysis: access to both contexts

Consider again the relevant structures, repeated as (14) and (15) below:

[10] Of course, this still needs to be *syntactically* formalized to be able to effect Spell-Out rules at PF in a Y-modular architecture.

(14) raman$_i$ [$_{CP}$ taan$_{\{i,*j\}}$ jey-pp-een/*aan-nnu] so-nn-aan
raman-NOM SE-NOM$_i$ win-FUT-1SG/*3MSG-that say-PST-3MSG
"Raman said [$_{CP}$ that he$_{\{i,*j\}}$ will win]"

(15) raman$_i$ [$_{CP}$ taan$_{\{i,*j\}}$ jey-pp-aan/*een-nnu] kaɳɖu
raman-NOM$_i$ SE-NOM$_i$ win-PST-3MSG/*1SG-that find.out-PST-3MSG
piɖi.tt.aan

"Raman$_i$ found out [$_{CP}$ that he$_{\{i,*j\}}$ would win.]"

Described informally, the embedded clause in both structures represents an eventuality that is a *de se/self-report* of the author of the embedding context. Nevertheless, this eventuality is one that is ultimately also from the (utterance-context) speaker's/author's perspective (since it is represented as 3rd person). How do we capture this dual representation? One option is to propose that *both* utterance and embedding contexts are syntactically represented. Such a claim is not controversial – analogous proposals have been made to account for double access readings (DAR) for tense (Giorgi 2010) and mood (Schlenker 2004).[11] In a cartographic model (Rizzi 1997), both these contextual coordinates could be made locally and syntactically available in the exploded C-domain (see again Giorgi 2010, for a very similar proposal for DAR readings for tense).

But we still need an explanation for the systematic distinction in embedded verbal agreement between the two types of structures: 1SG in (14) and 3MSG in (15). I propose that the coordinates of the embedding context are syntactically present in the C-domain in both types of structures. Notice that this isn't much of a stipulation: the embedded eventuality, syntactically represented as a propositional complement in these structures, can never fully "escape" the embedding one in terms of its functional properties precisely by virtue of its being embedded (see Giorgi 2010, for more discussion of this for the temporal domain).[12] In contrast, the utterance context does not need to be syntactically represented (thus locally accessible) in the embedded clause. Let us specifically propose that the utterance context is not represented in the C-domain in structures like 14 but *is* represented in structures like 15.

How can this distinction be exploited to derive the difference in agreement patterns? Contextual coordinates provide the background with respect to which person features on a DP are evaluated. Semantically, an Author or Addressee could be seen as a function that takes a specific context as argument: $\lambda c.Author/Addressee(c)$, for c = context. Note then, that the inherent and interpretable person features on a DP themselves cannot be evaluated independent of a context. As discussed earlier, *taan* in both (14) and (15) appears to be interpreted as 1st person with respect to the embedding context and simultaneously as 3rd

[11] This similarity is potentially non-accidental, thus significant – though at this point, it is not clear how it should be formalized – because a contextual index expands to include Author, Addressee, Time, World (and perhaps also Location) coordinates. Thus, it is reasonable to expect that phenomena in the temporal and modal domains can be replicated and analogously derived in the nominal one (Schlenker 2003, and many others).

[12] However, the syntactico-semantic nature of clausal embedding does influence how "free" or "bound" the embedded clause can be, creating, in essence, a finiteness cline (Landau 2004, Sundaresan and McFadden 2009).

person with respect to the utterance context. This can actually be encoded directly in the syntax in (15) with two person features on the embedded DP because the 1st person feature can be interpreted with respect to the embedding context, and the 3rd person feature can be interpreted with respect to the utterance context. It might seem a bit controversial to propose two apparently contradictory person features on a single DP. Notice, however, that each person feature is obtained against a different context and the actually contradictory situation where a single DP has incompatible feature-specifications with respect to a single context is still excluded in this system. Returning to the discussion, if (as per my claim) the utterance context is syntactically absent in the embedded CP in (14), a 3rd person feature on the embedded DP would not actually have an appropriate context against which to be evaluated. Only a 1st person feature, interpreted with respect to the embedding context, can be present.

There are two ways to formalize this notion. One option is that the DP subject that gets spelled out as *taan* is born with no value for the person feature in both types of structures. This could be formally instantiated as an unvalued $\phi$-feature on the embedded DP, rendering this DP into a Probe for the Person feature (see also Reuland 2001, Heinat 2008, Kratzer 2009, for different implementations of this idea for anaphors).[13] Given the discussion of the role of contexts above, it seems reasonable to propose that the relevant contextual coordinates (Author, Addressee) serve to value these features, either by inheriting the person feature from the DP antecedent (in the sense of Kratzer 2009) or by already having these to begin with. There is some initial evidence, at least from structures like (15), that the inheritance approach is on the right track – although *taan* itself is underspecified for gender (i.e. can take MALE, FEMALE, and NEUTER antecedents), the agreement on the embedded verb in (15) is 3MSG – suggesting that the gender feature has been inherited. Nevertheless, there is a nontrivial problem of counter-cyclicity to an inheritance approach within a phase-based model, (though this is a problem that all feature-inheritance approaches face, as far as I know). I do not discuss this issue further for now, though it will of course have to be worked out in more detail.

If this approach is on the right track, we would end up with the following features for the embedded DP in the different structures:[14]

**The embedded DP in (15)**:

Starts out with these $\phi$-features: [Number: , Person: , Gender: ]
After valuation by context$_{embedding}$: [Number: sg, Person: 1, Gender: masc]
After valuation by context$_{utterance}$: [Number: sg, Person: 1, 3, Gender: masc.]
Values embedded verb as: [Number: sg, Person: 1, 3, Gender: masc.]

[13] And see again Heinat (2008) for arguments in favor of phrasal probing and Zeijlstra (2010) for arguments in favor of upward probing.

[14] I assume, following Giorgi (2010) that the coordinates of the embedding context are represented hierarchically below those of the utterance context in the C-layer – though nothing significant hinges on this assumption at the present stage. Under the approach below, a formal mechanism would also have to ensure that the embedded T Agrees with the clausemate embedded DP only *after* the latter has had any unvalued $\phi$-features of its own valued. It is also possible that a feature-sharing mechanism as in Pesetsky and Torrego (2007) might be called for to deal with this. I don't discuss this further here at this stage.

**The embedded DP in (14)**:

> Starts out with these $\phi$-features: [Number: , Person: , Gender: ]
> After valuation by context$_{embedding}$: [Number: sg, Person: 1, Gender: masc.]
> There is no context$_{utterance}$, thus there is no valuation as 3rd person.
> Values embedded verb as: [Number: sg, Person: 1, Gender: masc.]

Under the second option, the embedded DP could already be specified as: 1, 3, sg in line with the dual person-interpretation it seems to (descriptively) possess:
**The embedded DP in (15)**:

> Starts out with these $\phi$-features: [Number: sg, Person: 1, 3, Gender: masc.]
> Values embedded verb as: [Number: sg, Person: 1, 3, Gender: masc.]

**The embedded DP in (14)**:

> Starts out with these $\phi$-features: [Number: sg, Person: 1, Gender: masc.]
> Values embedded verb as: [Number: sg, Person: 1, Gender: masc.]

A potential disadvantage of this option is that it proposes distinct underlying $\phi$-features for the embedded DP that gets spelled out as *taan* in (14) and (15). What would happen if the embedded DP in (14) started out with identical $\phi$-features to that in (15) – i.e. as [Number: sg, Person: 1, 3, Gender: masc.]? Notice that the 1 and 3 features are with reference to two different contexts (in fact, they have to be to avoid internal inconsistency): the person feature 1 takes the embedding context and the person feature 3 takes the utterance context as its argument, respectively. Given the claim that the utterance context is entirely absent in the embedded C-layer in structures like (14), the 3rd person feature (despite being considered as inherent and interpretable under this option) would have no way of being semantically interpreted – and would thus crash. Again, the details have to be worked out more formally, but this is the general idea.

### 6.1 A [-R] feature on the embedded DP

One final important point is in order. I assume that, in addition to these $\phi$-features, the embedded DP also has a [-R] feature[15] (Reinhart and Reuland 1993, Landau 2004). I am using this as a formal featural encoding of nominal anaphoricity – one that essentially flags a DP as an anaphoric element for variable binding at LF.[16] This feature is present in

[15]It is also possible that this is actually encoded as an unvalued [R] feature which in itself might turn the anaphoric DP into a Probe at Narrow Syntax. This would potentially have different implications from those discussed here – I don't discuss this point further at this juncture.

[16]I realize that this is potentially at odds with some other recent claims (Kratzer 2009, Reuland 2001, and many others) that what makes a DP anaphoric is precisely its lack of $\phi$-features. One reason I don't adopt this idea is because of empirical evidence from Tamil, Malayalam and other languages (Jayaseelan 1997, Annamalai 1999) which contain fully $\phi$-specified DPs that may be anaphorically bound (even locally), the converse evidence of $\phi$-deficient DPs that do not behave like anaphors (see e.g. the discussion of quirky expletives in Richards 2008) and finally because I agree, in line with Hicks (2009) that while $\phi$-features may constrain the domain of antecedents for an anaphoric DP (presuppositionally, in the sense of Heim and Kratzer 1998, for instance), they do not uniquely define reference. I do not discuss this point further here.

both types of structures, by my claim. Thus, the feature-calculus described above for both structures must be updated with a [-R] feature that starts out on the embedded DP. Crucially, this [-R] feature can be exploited to explain the 3rd-person antecedence restriction of *taan*. Why? Because it flags *taan* as a DP that needs to be variable-bound: but DPs that are underlyingly specified 1st/2nd person cannot variable-bind since they are rigid designators across assignments.[17] This also explains why *taan* typically cannot appear in matrix subject position (16), except in free indirect discourse contexts (what is traditionally termed a logophoric usage (Reinhart and Reuland 1993)) whereas *naan* can (17):

(16) * taan$_i$ tuungi-n-aan/een [i $\rightarrow$ $[\![Attitude-holder]\!]$]
SE sleep-PST-3MSG/1SG
"He/I slept" (Intended)

(17) naan tuungi-n-een
I sleep-PST-1SG
"I slept"

### 6.2 Summary of patterns and Spell-Out Rules

The table below summarizes the relevant features and featural-interactions:

| | **Structure (14)** | **Structure (15)** |
|---|---|---|
| $\phi$ ON AGR$_{embedded}$ | 1, sg, masc. | 1, 3, sg, masc. |
| $\phi$ ON EMBEDDED SUBJECT | 1, -R, sg, masc. | 1, 3, -R, sg, masc |
| SEMANTICS OF EMBEDDED SUBJECT | *de se* | *de se* |
| REL: EMBEDDED SUBJECT TO CONTEXT$_{Att}$ | 1 | 1 |
| REL: EMBEDDED SUBJECT TO CONTEXT$_{Utt}$ | _ | 3 |

Table 1: Descriptive properties of Tamil patterns

Given the above table, we can now propose the spell-out rules in Table (2) under a Late Insertion model Halle and Marantz (1993), Starke (2009) – which proposes different underspecification patterns for DPs and Agr, to yield the right results.

## 7. Wrapping up: final issues

The analysis presented above deals with the $\phi$-featural patterns in the embedded CP in structures like (14) and (15) by means of tweaking these different parameters: The $\phi$-

[17] An immediate potential challenge this claim would be the existence of the so-called "fake-indexicals", in structures like the following (Kratzer 2009), where an indexical pronoun appears to participate in variable binding, as in (i) below:

i. I am the only one who did my homework.

Notice, however, that fake indexicals are typically not treated as "real"/underlying indexicals – crucially, they are typically held not to possess indexical features at the point of variable binding. Nevertheless, I concede that the issue of fake indexicals is an important one that crosscuts the current discussion – and needs to be explained further in this system.

| | | | |
|---|---|---|---|
| | [1,sg] | ⟷ | *-een* |
| AGR | [3, masc, sg] | ⟷ | *-aan* |
| | [1, sg] | ⟷ | *naan* |
| | [1, sg, -R] | ⟷ | *taan* |
| PRONOUN | [2, sg] | ⟷ | *nii* |
| | [3, masc, sg, +R] | ⟷ | *avan* |
| | [3, fem, sg, +R] | ⟷ | *avaɭ* |

Table 2: Spell-Out rules in a Late Insertion model

features that the embedded DP ends up with, the types and number of contexts that are syntactically represented, thus locally accessible, in the embedded C-layer, and distinct patterns of underspecification on the exponents for pronouns and verbal agreement in the Spell-Out component. This is a preliminary analysis – it is clear that a lot of details need to be further ironed out and non-trivial problems like potential counter-cyclicity dealt with.

Nevertheless, there is some promising independent evidence that the embedded CP in structures like (14) is structurally smaller than those in (15). Clauses like (14) seem to also be capable of being represented as gerundivals (18). However, structures (15) seem to resist this (19):[18]

(18) raman$_i$ [$_{CP}$ taan$_{\{i,*j\}}$ poo-v-adaaga] so-nn-aan
raman SE$_i$ go-FUT-GER say-PST-3MSG
"Raman$_i$ said [$_{CP}$ that he$_{\{i,*j\}}$ would go]"

(19) */?? raman$_i$ [$_{CP}$ taan$_{\{i,*j\}}$ poo-v-adaaga] kaɳḍupiḍi-tt-aan
raman SEgo-FUT-GER find.out-PST-3MSG
"Raman$_i$ found out [$_{CP}$ that he$_{\{i,*j\}}$ would go ] "

This reinforces the idea that different types of verbs select different-sized complements based on their own syntactico-semantics (attitudinal, non-attitudinal, reportive, etc) – an idea that also brings in the relevance of reportive vs. non-reportive embedding predicates from the shifted-indexicality discussion and literature.

## References

Anand, Pranav. 2006. De de se. Doctoral Dissertation, MIT.

Anand, Pranav, and Andrew Nevins. 2004. Shifty operators in changing contexts. In *Proceedings of SALT*.

Annamalai, E. 1999. Lexical anaphors and pronouns in Tamil. In *Lexical and anaphors and pronouns in selected South Asian languages: a principled typology*, ed. Barbara C. Lust et al., 169–216. Mouton.

[18]Interestingly, we have a similar pattern in English as well:

i. John$_i$ spoke of (his$_i$) going.
ii. * John$_i$ found out of (his$_i$) going.

Giorgi, Alessandra. 2010. *Towards a syntax of indexicality*. Oxford Studies in Theoretical Linguistics. Oxford: Oxford University Press.

Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In *The view from building 20*, ed. Hale and Keyser. MIT Press.

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

Heinat, Fredrik. 2008. *Probes, pronouns and binding in the Minimalist Program*. VDM.

Hicks, Glyn. 2009. *The derivation of anaphoric relations*. Benjamins.

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

Kratzer, Angelika. 2009. Making a pronoun: fake indexicals as windows into the properties of pronouns. *LI* 40:187–237.

Landau, Idan. 2004. The scale of finiteness and the calculus of control. *Natural Language and Linguistic Theory* 22:811–777.

Pesetsky, David, and Esther Torrego. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and clausal architecture: syntactic derivation an dinterpretation*, ed. Simin Karimi et al., 262–294. Amsterdam: Benjamins.

Reinhart, Tanya, and Eric Reuland. 1993. Reflexivity. *LI* 24:657–720.

Reuland, Eric. 2001. Primitives of binding. *LI* 32:439–492.

Richards, Marc. 2008. Quirky expletives. In *Agreement restrictions*, ed. d'Alessandro, Hrafnbjargarson, and Fischer, 181–213. Berlin: Mouton.

Rizzi, Luigi. 1990. On the anaphor-agreement effect. *Rivista di Linguistica* 2:27–42.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar*, ed. Liliane Haegeman, 281–337. Dordrecht: Kluwer Academic Publishers.

Schlenker, Philippe. 2003. A plea for monsters. *Linguistics and Philosophy* 26:29–120. Kluwer Academic Publishers.

Schlenker, Philippe. 2004. Sequence phenomena and double access readings generalized: two remarks on tense, person, and mood. In *The syntax of time*, ed. Jacqueline Guéron and Jacqueline Lecarme, 555–596. Cambridge, MA: MIT Press.

Schlenker, Philippe. To appear. Indexicality and *de se* reports. In *Handbook of Semantics*, ed. von Heusinger, Maienborn, and Portner. Mouton.

Starke, Michal. 2009. Nanosyntax: a short primer to a new approach to language. *Nordlyd: Special issue on Nanosyntax* 36:1–6. Edited by Peter Svenonius, Gillian Ramchand, Michal Starke, and Knut Tarald Taraldsen.

von Stechow, Arnim. 2002. Feature deletion under semantic binding: tense, person, and mood under verbal quantifiers. In *Proceedings of NELS 33*.

Sundaresan, Sandhya, and Thomas McFadden. 2009. DP distribution and finiteness in Tamil and other languages: selection vs. Case. *JSAL* 2.

Woolford, Ellen. 1999. More on the anaphor agreement effect. *LI* 30:257–287.

Zeijlstra, Hedde. 2010. There is only one way to agree. Talk given at GLOW 33, Wroclaw.

CASTL, Universitet i Tromsø
9037 Tromsø
Norway

sandhya.sundaresan@ uit.no