"Monstrous agreement", anaphora, and context: evidence from Tamil[1]

Sandhya Sundaresan, University of Tromsø (CASTL)/ Stuttgart

To appear in Proceedings of NELS41, University of Pennsylvania, 2010

1 The puzzle

The focus of this paper is structures like (1) in Tamil:

(1) raman taan poo.n.een/*aan-nnu so.nn.aan
raman.NOM SE.NOM$_i$ go.PST.1SG/*3SG-that say.PST.3MSG
"Raman said [$_{CP}$ that he$_{\{i,*j\}}$ went]"

The structure in (1) contains an (attitudinal) speech predicate *sonnaan* (told.3MSG) which embeds a (finite) clausal complement. This CP complement has a subject *taan*[2] – traditionally described as a nominative, 3SG. (optionally) long-distance (or SE- (Reinhart and Reuland, 1993)) anaphor (Annamalai, 1999). However, I suggest later in this paper that *taan* in (1) might be a 1st-person indexical. The glossing of *taan* as "SE" should thus be treated as purely terminological for now. This embedded subject DP *taan* is obligatorily coreferent with the matrix (3rd-person) subject and attitude-holder, *raman* and is, furthermore, interpreted as obligatorily *de se* with respect to its DP antecedent. Finally, and most puzzlingly, the embedded indicative verb *poo.n.een* (went.1SG) doesn't seem to agree with either its clausemate embedded subject *taan* or with the matrix subject *raman* with respect to the person-feature.

1.1 Against the Anaphor Agreement Effect

It has been proposed (Woolford, 1999) that the lack of embedded subject-verb agreement in (1) is a grammatical conspiracy to avoid a violation of the Anaphor Agreement

[1]Thanks to Gennaro Chierchia, Volker Gast, Alessandra Giorgi, Ken Safir and Peter Svenonius for fruitful discussions. I am especially grateful to Thomas McFadden and Gillian Ramchand for detailed commentaries on the syntactic and semantic (more or less respectively) aspects of the analysis presented in this paper. I also thank the audiences at the Local Pronouns workshop in Nijmegen, GLOW-in-Asia in Beijing, Peculiar Binding Configurations workshop in Stuttgart and NELS 41, UPenn, where earlier versions of this paper were presented, for useful comments and feedback.

[2]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.

Effect/AAE (originally due to Rizzi, 1990) – which claims that anaphors are crosslinguistically disbarred from agreement(-triggering) positions. The problem with such an analysis, however, is that when the same subordinate clause in (1) is embedded under a non-reportive matrix verb like *kee*ɭ*vipa*ʈʈ*aan* ('overhear') "real" agreement is obligatorily required, as shown below:

(2) raman taan poo.n.aan/*een-nnu keeɭvipaʈʈaan
raman.NOM$_i$ [SE.NOM$_i$ go.PST.3SG/*1SG-that] (over)hear.PST.3MSG
"Raman$_i$ (over)heard [$_{CP}$ that he$_{\{i,*j\}}$ went.]"

Crucially, furthermore, the agreement marking on the embedded verb in (2) cannot be dismissed as default agreement marking: first, because default verbal agreement marking in Tamil, as for instance, with a clausemate quirky subject, is instantiated as 3NSG and second, because the 3MSG agreement on the embedded verb in (2) may only surface with a masculine, 3rd-person clausemate subject. This second point is illustrated in (3) below:

(3) seetha taan poo.v.aaɭ/*aan-nnu keeɭvipaʈʈaaɭ
seetha SE go.PST.3FSG/*3MSG-that hear.PST.3FSG
"Seetha$_i$ heard [$_{CP}$ that she$_{\{i,*j\}}$ went]."

Based on these reasons, I propose, contra Woolford (1999) that structures in Tamil like (1) are not a function of the Anaphor Agreement Effect. Notice that this does not mean that there is no connection between anaphora and agreement – indeed, syntactic intervention effects such as those found in Chinese binding constructions (Huang and Liu, 2001, and others) strongly suggest otherwise – merely that the AAE cannot be maintained in its current form as a ubiquitous generalization about the world's languages.

But if the apparent agreement mismatch in the embedded CP in (1) is not due to the AAE, what, then, can it be due to? Conversely, why is subject-verb agreement necessitated in structures like (2)? This is the central puzzle and its investigation the central focus of this paper.

2 Intensionality and indexical shift

Let us consider, again, the structure in (1). The matrix predicate in this structure is *sonnaan* (said.3MSG), which is a reportive predicate. The clausal complement thus denotes a speech-report. In contrast, the matrix predicate in (2) *kee*ɭ*vipa*ʈʈ*aan* (overheard.3MSG)

is a non-reportive predicate: the embedded clausal complement, thus, does not constitute a speech-report.

Now let us consider the **shifted indexical** structure from Zazaki below (Anand and Nevins, 2004):

(4) hɛseni$_j$ (mi$_k$-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."

What is important for our current purposes is the 1st person indexical pronoun within the embedded clause in (4). This indexical pronoun ɛ*z* (I) can refer to the speaker as is standard. However, it can also, anomalously, refer to the attitude-holder *hɛsen* occupying the matrix subject position. In other words, the embedded indexical pronoun is referentially ambiguous – in common parlance, it is called a shifted indexical or Kaplanian monster (Schlenker, 1999; von Stechow, 2002; Anand, 2006). Significantly, shifted indexicality has been shown to crosslinguistically obtain only in clauses 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 (4), the 1st-person marking on the embedded verb in (1) is shifted to refer to the attitude-holder (*Raman*), not to the speaker. This 1st-person marking, furthermore, appears to be induced under conditions that are identical to those that induce 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), *oppukko*ɭ (admit).[3]

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

Based on the above, I thus propose that the structure in (1) involves a type of indexical shift – for now, I thus term the agreement on the embedded verb in (1) "monstrous agreement".

[3]At the same time, other attitude-predicates like *nene* (think/believe) don't seem to induce such shift. I am in the process of conducting a survey among native Tamil speakers to see exactly which intensional predicates facilitate indexical shift in this language.

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. This is because there are several grammatical phenomena that emulate the effects of indexical shift without being underlyingly monstrous. In this spirit, I discuss two of the usual suspects (see among others Schlenker, To appear, for more discussion) – full-on quotatives and mixed/partial quotatives.

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). To see this, consider the following example:

(6) 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 (6) is not an instance of indexical-shift, but one of full-on clausal quotation.

In this section, I consider two types of quotations – full-on quotations where an entire clause is quoted and partial quotations where only a subpart of a clause is quoted. I argue that the paradigm in (1) does not involve either type of quotation.

3.1.1 It's not a full-on quotation

In English, quoted elements may never be embedded under a complementizer – thus, the concomitant presence/absence of an overt complementizer can be used to tease quotatives and non-quotatives apart. But languages like Tamil and Japanese embed both quotatives and non-quotatives under the same complementizer – thus, different diagnostics are needed. Fortuitously, the grammatical opacity property of quotatives, described above, itself yields some useful diagnostics for testing the existence of full-on quotatives even in languages where this is not obvious from the surface paradigms. First, *wh*-extraction outside of a quoted domain is expected to be ungrammatical. Also, an NPI inside a quotative may not be licensed by an operator outside the quote.

I apply both tests to the relevant Tamil structure in (7) and (8) below. Notice that the embedded verb in both structures exhibits an apparently mismatched 1SG agreement, as desired:

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

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

Crucially, the structures in (7) and (8) are both fully grammatical – which would be entirely unexpected if the embedded CP were indeed a grammatical island. This shows conclusively that the embedded CP in these structures is not a full clausal quotative.

For the sake of completeness, I present the minimal pairs to (8) below – the result of replacing embedded subject *taan* with the 1st-person indexical *naan* ("I"). The structure in (10) is ungrammatical, indicating that the embedded CP is a full-on quotative.

(9) * krishnan yaarai naan paa.tt.een-nnu so.nn.aan
krishnan who(m) I see.1SG-that say.PST.3MSG
"*Who(m)$_1$ did Krishnan$_i$ say "I saw t$_1$?"

(10) * raman naan oru tappu.m senjeen-nnu ottukka.le
raman I one mistake.even made-that admit.NEG
"*Raman didn't say "I made any mistake."

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

(11) * raman naan oru tappu.m seyya.le-nnu so.nn.aan
raman I one mistake.even made.NEG-that admit.PST.3MSG
"Raman said "I didn't make any mistake."

Thus Tamil does have quotations – only, a quoted 1SG indexical pronoun surfaces as *naan*, not *taan*. In my analysis of these patterns toward the end of this paper, I account for this by motivating and specifying different underlying syntactic features for these elements, which in turns enables them to be spelled out differently.

3.1.2 It's not a partial/mixed quotation either

Unlike full on quotatives of the kind discussed above, partial quotatives (where only a subpart of the clause is quoted) can influence the truth-conditional semantics of the larger context they occur in. Thus, the sentence: ""James thinks "Casanova" is not having much

luck with girls" (where "Casanova" is James' nickname for his popular friend) could be true – even though history has taught us that the real Casanova had no problems in this respect. Could it be the case that the embedded CP in (1) involves a partially quoted element – perhaps only the embedded verb is quoted?

There is, however, a fairly straightforward way to show that this is actually not the case. If the embedded CP in (1) involved a partial quotative we wouldn't expect 1SG embedded verb-agreement as in (1) to be restricted to clauses embedded under reportive predicates. This is because, as Anand (2006) there are very few restrictions on what may be quoted – pretty much any string can be enclosed within quotes, including a nonsense string, as shown by the minimal pair below:

(12) Marx stuck out his tongue and said "pffft!"

(13) * Marx stuck out his tongue and said that pffft!

Based on this reasoning, I conclude that 1 doesn't involve a partial quotative either.

4 Differences from (classical) indexical shifting

The discussion above shows that structures like (1) do not involve either fully or partially quotes strings and, by extension, strongly suggests that (1) involves a bonafide shifted indexical in the embedded clause. Notice that I am using the term "shifted indexical" entirely pre-theoretically and descriptively, to indicate an indexical element that is capable of referring to an entity other than that encapsulated within the utterance context.

At the same time, (1) differs from standard structures involving indexical shift. In the section below, I consider some of the most significant of these differences.

4.1 Possibilities for denotation

Standard cases of indexical shift discussed in the literature (citations above) involve an indexical element that shows optionality in reference. In (4) from Zazaki above, the embedded indexical subject may refer either to the speaker or to the attitude-bearer. In (1), on the other hand, there is no referential optionality for the indexical. The 1st-person marking on the verb is obligatorily shifted to refer to the attitude-bearer; it may never refer to the speaker. Such obligatory shift is attested in other languages – for instance, the instances of logophora in many African languages – though it has not always been analyzed

as a case of indexical shift. For instance, Hyman and Comrie (1981) discuss a logophoric verbal suffix in Gokana which signifies obligatory coreference with a higher subject/agent. Schlenker (To appear) provides a shifted-indexical analysis of such logophors and it is tempting to simply extend his analysis to the Tamil cases. However, the Tamil structure in (1) differs from the relevant Gokana ones in two significant (and ultimately problematic) ways. First, the logophoric marker in Gokana does not have any obvious person-marking in the sense that that 1SG verbal marking *-een* in Tamil does – and can co-occur with 1st, 2nd, and 3rd-person subjects. Second, it doesn't obligatorily co-occur with a clausemate subject like *taan* (itself traditionally described as a logophor). Additionally, Anand (2006, p. 78) mentions one other instance of such obligatory shift (taken from the Athapaskan language Slave Rice, 1986, p. 53) the Athapaskan language, Slave (formatting mine):

(14) simon [rásereyineht'u] hadi
simon [2SG-hit-1SG] 3SG.say
"Simon said that you hit {him, *me}."

I do not discuss this distinction further in this paper – though it is clear that this is something that should be researched further.

4.2 (Non-)locality of the agreement mismatch

A more potentially problematic difference between Tamil (1) and Zazaki (4) is as follows. In the "classic" indexical structure in Zazaki (4), the morphosyntactic agreement mismatch between the shifted indexical and the DP that it denotes (the attitude-holder) 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 3SG matrix subject/attitude-holder (*Raman*), there is a person-feature mismatch within the embedded clause itself. That is, the embedded subject may only take 3rd-person antecedents; it is marked 3SG. Its clausemate verb is, however, marked 1SG.

Why is this potentially problematic? Because the verbal suffix *-een* on the embedded verb in (1) is an agreement marker![4] Under standard Minimalist assumptions of Agree (Chomsky, 2001) 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 clause-

[4]This seems to be in sharp contrast to corresponding structures in Gokana, for instance, where the logophoric marker, while being instantiated as a verbal suffix, does not behave like an agreement marker – in the sense that it is invariant regardless of the ϕ-features on the clausemate subject.

mate subject *taan* itself seems to exhibit different person-features (3rd) from that on its clausemate verb (1st). Where does this 1st-person agreement get inherited from?

The puzzle that arises for structures like (1) is thus as follows.m What enforces 1st-person marking on the embedded predicate? Relatedly, what does this 1st-person marking denote and how can be it be formally represented/derived?

5 Analysis of indexical shift in Anand (2006)

Here, I delineate the context-overwriting approach to shifted indexicality in Anand and Nevins (2004) and later in Anand (2006). Consider again the case from Zazaki (4), repeated as (15) below:

(15) hɛseni$_j$ (mi$_k$-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."

Reading A: Indexical $I \rightarrow$ Utterance-context author = *Sandhya*; UNSHIFTED.

Reading B: Indexical $I \rightarrow$ Intensional-context author = *Hesen*; SHIFTED.

Thus, the difference between the shifted and unshifted readings above is not the denotation of the indexical I, which remains constant ($= \lambda c.Author(c)$) – but the value of the context that the AUTHOR function ranges over. This is the central intuition behind Anand (2006)'s proposal. In his acccount, shifted indexicality is the result of context shifting due to context-overwriting.

This is formally derived as below using Stalnaker (1981)'s diagonalization operator:

(16) $[\![OP_{diag}]\!]^{c,i} = \lambda\chi_{<k,k,t>}(i)(i)$
$c_{<k>}$ = CONTEXT COORDINATES, $i_{<k>}$ = INDEX COORDINATES [5]

Anand proposes that the diagonalization responsible for context-overwriting in Zazaki is: $OP_{\forall}$, which is defined as follows:

(17) $[\![OP_{\forall}\alpha]\!]^{c,i} = [\![\alpha]\!]^{j,i}$ where j = $<Auth(i), Addr(i), Time(i), World(i)>$.

This causes the diagonalization operator to replace every contextual coordinate in its scope with the coordinates of the intensional index/context: this matches the indexical-shift facts for Zazaki. The idea is that different languages have different operators based

[5]Index coordinates represent a tuple of *de se* coordinates that are quantified over by an intensional predicate; this yields the obligatory *de se* semantics of shifted indexical structures.

on which contextual coordinates may shift – this accounts for the parametric variation in what indexical elements may shift. This $OP_{\forall}$ – which is presumably located in C - is optionally selected by shifting intensional predicates (e.g. *say*), resulting in shifted as well as unshifted readings.

The unshifted reading is the starightforward one which also obtains in English and can be formally derived as follows:

(18) **Unshifted Reading A:** [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). The shifted reading (Reading B) obtains when the $OP_{\forall}$ is selected.

(19) **Shifted Reading B:** [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]]$^{i',i',g}$

5.1 Extending Anand (2006)'s analysis to the Tamil data

In the discussion of Tamil above, I have treated the 1SG verbal agreement on the embedded verb as a shifted indexical. At the same time, by virtue of being an agreement marker, these ϕ-features must not inherent to the embedded T head but **inherited** via Agree with the embedded subject *taan*) in [Spec, TP$_{embedded}$]. The simplest way to achieve this would be to claim that *taan* in (1) must itself be specified for the valued/interpretable feature 1SG.

Is such a claim workable in Anand's system? Anand (2006) argues that Malayalam *taan* as a 1st-person indexical (based on its obligatory *de se* semantics), rather than an anaphor. Tamil *taan*, which is very similar to its Malayalam counterpart (Jayaseelan, 1997; Annamalai, 1999), also yields obligatory *de se* readings, suggesting that it too is a 1st-person indexical underlyingly.

(20) raman taan poo.n.een-nnu so.nn.aan
raman SE$_i$ go.PST.1SG-that say.PST.3MSG
"Raman said [$_{CP}$ that he$_{\{i,*j\}}$ went]"

Obligatory *de se* **reading:** *Raman said, "I went."*

If the discussion above is on the right track, the 1sg agreement on the embedded verb in (23) would be merely due to standard Agree between embedded T and the 1sg indexical pronoun (spelled out as *taan*) in [Spec, TP]. Nevertheless indexical shift does apply – to the contextual coordinates of the embedded 1st-person indexical subject *taan* – which refers to the attitude-holder rather than the speaker of the utterance-context.

This sounds promising. But a context-overwriting approach as in Anand (2006) for the Tamil data presented here is beset with two serious problems which, as far as I can tell, cannot be solved non-stipulatively.

Below, I discuss these problems in detail.

5.1.1 Problem I: Antecedent-restriction to 3rd-person DPs

The first major problem for a context-overwriting approach is that Tamil as well as Malayalam *taan* only take 3rd-person antecedents (21 vs. 22):

(21) raman taan poo.r.een-nnu sol.r.aan
raman IND go.PRES.1SG-that say.PRES.3MSG
Raman says, "I will go."

(22) * naan taan poo.r.een-nnu sol.r.een
I IND go.PRES.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, 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. The other option would be to claim that *taan* in these languages 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.

But such a condition, while workable, is, as Anand himself concedes, very stipulative.

5.1.2 Problem II: Regular agreement structures as in Tamil (2)

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 because it has no verbal agreement, as discussed earlier.) In

Anand (2006)'s system, we are essentially forced to claim that *taan* in structures like (2) is underlyingly different from that in (1). We would have to propose, for instance, that it is not an indexical, but a logophor/SE-anaphor, which inherits the ϕ-features of its antecedent and transmits these via Agree to the embedded verb (Kratzer, 2009)). A major empirical challenge to this type of analysis is that structures like 2 also seem to be interpreted as **obligatorily** *de se*.[6] Under Anand's analysis, this would necessarily be taken to mean that *taan* in such structures is also underlyingly a 1st person indexical pronoun. The problem is that, this would also mean that the distinction between 1sg vs. 3sg agreement on the embedded T/Agr in each can no longer be made purely in terms of the differential feature-specification of *taan* in each instance.

Nevertheless, the Tamil patterns in (1) and (2) can be accounted for under Anand's context-overwriting approach once we make these stipulations. In the Appendix at the end of this paper, I present a detailed analysis of the Tamil data under such an approach.

6 An alternative analysis: access to both contexts

Here I present an alternative analysis to these Tamil patterns. It should be noted at the outset that the formalized implementation of this idea are preliminary and thus tentative – and constitute an initial attempt to make the central intuition precise. While I agree with Anand (2006) in taking *taan* to be a 1st-person shifted indexical (again, in the purely descriptive, pre-theoretic sense), I differ from his approach in the following ways. First, I propose that indexical shift is effected, not by means of context-overwriting, but through allowing simultaneous access to embedding and utterance contexts in the syntax, specifically in the exploded embedded C-layer. Second, I claim that a 1st-person indexical pronoun, like the embedded DP that is spelled out as *taan*, can be simultaneously 1st and 3rd person in the syntax because each of these person features is evaluated against a different context in the C-layer. Finally, I make crucial use of the idea that an indexical pronoun can sometimes also be syntactically anaphoric and derive the difference between indexical pronoun *naan* ('I') and *taan* in Tamil in this manner.

Now consider again the relevant structures, repeated as (23) and (24) below:

[6]This is as far as I am able to ascertain based on my own grammaticality judgments. I am in the middle of conducting an online survey among other native speakers to test these intuitions.

(23) raman taan poo.n.een/*aan-nnu so.nn.aan
raman.NOM SE.NOM$_i$ go.PST.1SG/*3SG-that say.PST.3MSG
"Raman said [$_{CP}$ that he$_{\{i,*j\}}$ went]"

(24) raman taan poo.n.aan/*een-nnu keeḷvipaṭṭaan
raman.NOM$_i$ [SE.NOM$_i$ go.PST.3SG/*1SG-that] (over)hear.PST.3MSG
"Raman$_i$ (over)heard [$_{CP}$ that he$_{\{i,*j\}}$ went.]"

To see how these structures should be analyzed, I first informally lay out what each of these structures means descriptively. 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 interpretation? One option might be 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 for tense (Schlenker, 1999; Giorgi, 2010) and mood (Schlenker, 2003). 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 – an idea discussed in Partee (1973) and much analyzed since (Heim, 1994; Kratzer, 1998; Schlenker, 2003, and many others). In a cartographic model (Rizzi, 1997; Cinque, 1999), 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 (23) and 3msg. in (24). Let us suppose that the coordinates of the embedding context are syntactically present in the C-domain in both types of structures. This in itself is not stipulative. 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). However, the syntactico-semantic nature of clausal embedding does influence how "free" or "bound" the embedded clause can be, creating, in essence, a finite-

ness cline (Landau, 2004; Sundaresan and McFadden, 2009). 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 (23) but is represented in structures like (24).

How can this distinction be exploited to derive the difference in agreement patterns? Consider again the role and significance of a context (utterance or embedding) to the person feature on a DP. Informally speaking, contextual coordinates provide the background with respect to which person features on a DP are evaluated. More formally, 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. Allowing the context to be present and interact actively in the syntax thus means that a syntactic instantiation of person, such as the person ϕ-feature on a DP can be syntactically relativized with respect to the contextual coordinates of a given context in the syntax, specifically in the exploded C-layer. As discussed earlier, *taan* in both (23) and (24) is interpreted as 1st person with respect to the embedding context and as 3rd person with respect to the utterance context. This can thus actually be encoded directly in the syntax in (24) 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. However, if (as per my claim) the utterance context is syntactically absent in the embedded CP in (23), 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. Notice again that this is only possible because I am treating the syntactic person feature on the DP as being relative to a context. Thus, it is not controversial to propose two apparently contradictory person features (1st and 3rd) on a single DP. 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. There are two ways to formalize this notion each of which I discuss briefly below.

6.1 Option 1: Unvalued person-feature on the embedded DP

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 ϕ-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).[7] 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 evidence, at least from structures like (24) that the inheritance approach is on the right track. (i) Notice that *taan* itself is underspecified for gender – i.e. it can take male, female, and neuter antecedents. (ii) Nevertheless, the agreement on the embedded verb in (24) is 3msg. (iii) Given (phase-locality), this would entail that it must have inherited the gender feature from its clausemate subject *taan* which itself must have inherited it from the DP antecedent, via the contextual coordinates in the C-layer. (iv) Nevertheless, under the PIC, there is a nontrivial problem of counter-cyclicity to such an approach, (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. Returning to the main discussion, if this approach is on the right track, we would end up with the a different feature calculus for the embedded DP in the different structures. 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 ϕ-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 different features are represented below:

The embedded DP in (24):

Starts out with these ϕ-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.]

[7]See Heinat (2008) for arguments in favor of phrasal probing and Zeijlstra (2010) for arguments in favor of upward probing.

The embedded DP in (23):

Starts out with these ϕ-features: [Number: , Person: , Gender:]

After valuation by $\text{context}_{embedding}$: [Number: sg, Person: 1, Gender: masc.]

There is no $\text{context}_{utterance}$, thus there is no valuation as 3rd person.

Values embedded verb as: [Number: sg, Person: 1, Gender: masc.]

6.2 Option 2: "Person-overspecification"

Under this option, the embedded DP could be specified as: 1,3, sg in line with the dual person-interpretation it seems to (descriptively) possess. In the discussion below, I assume that the embedded DP starts out fully specified for all relevant ϕ-features. But of course, intermediate options are also possible (potentially even more plausible) – e.g. we might propose that some features get inherited, in the manner described above, to account for the masculine gender feature on the embedded verb in (24).

Under the fully pre-specified version, the structure in (24) could be dealt as follows:

The embedded DP in (24):

Starts out with these ϕ-features: [Number: sg, Person: 1, 3, Gender: masc.]

Values embedded verb as: [Number: sg, Person: 1, 3, Gender: masc.]

The embedded DP in (23):

Starts out with these ϕ-features: [Number: sg, Person: 1, Gender: masc.]

Values embedded verb as: [Number: sg, Person: 1, Gender: masc.]

What would happen if the embedded DP in (23) started out with identical ϕ-features to that in (24) – 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 (23), 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.

6.2.1 An unvalued R feature on the embedded DP

One final point is in order. I assume that, in addition to these ϕ-features, the embedded DP also has an unvalued [R] feature (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. 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 ϕ-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 ϕ-specified DPs that may be anaphorically bound (even locally), the converse evidence of ϕ-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 ϕ-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. This feature is present in both types of structures, by my claim. Thus, the feature-calculus described above for both structures must be updated with an unvalued R feature that starts out on the embedded DP.

6.3 Summary of patterns

The table below summarizes the relevant features and featural-interactions for (23) and (24):

	Structure (23)	**Structure (24)**
ϕ ON $\text{AGR}_{embedded}$	1, sg, masc.	1, 3, sg, masc.
ϕ 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

6.4 Spell-Out Rules

Given the above table, we can now propose the following spell-out rules in a Late Insertion model Halle and Marantz (1993); Starke (2009) – which yields the right results.

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

	[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*

7 Wrapping up: final issues

The analysis presented above deals with the ϕ-featural patterns in the embedded CP in structures like (23) and (24) by means of tweaking three different parameters: the ϕ-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 under a feature-inheritance approach) dealt with.

Nevertheless, there is some promising independent evidence that the embedded CP in structures like (23) is structurally smaller than those in (24). Clauses like (23) seem to also be capable of being represented as gerundivals:

(25) raman$_i$ [$_{CP}$ taan$_i$ poo.v.adaaga] so.nn.aan
raman SE$_i$ go.FUT.GER say.PST.3MSG
"Raman said [$_{CP}$ that he$_i$ would go]"

However, structures (24) seem to resist this:

(26) */?? raman$_i$ taan$_i$ poo.v.adaaga kaɳḍupiḍi.tt.aan/keeɭvipaṯṯaan
raman$_i$ SE$_i$ go.FUT.GER discover.PST.3MSG/hear.PST.3MSG
"Raman$_i$ discovered/heard [$_{CP}$ that he$_i$ would go] "

If the embedded C-layer in structures like (23) is structurally smaller than that in (24) because the former (but not the latter) lacks the head that encodes featural information pertaining to the coordinates of the utterance context, as I propose here, the corresponding paradigm in (25) and (26) would be expected, perhaps even predicted. This correlation

also 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.

8 Appendix: Analysis under Anand (2006)

Here's how the rest of the analysis for Tamil (23) would play out in Anand (2006)'s system: $[\![taan]\!]^{c,g} = \lambda c.Author(c)$. The matrix intensional verb *sol* (say) introduces a context-shifting diagonalization operator which minimally c-commands the embedded CP: [Raman said $[_{CP}$ $\text{OP}_{diag-Tamil}$ I will-go_{1sg}]]^{c,i,g}. Indexical shift in Tamil (and as Anand discusses, also in Malayalam) seems to apply only to 1st-person indexical pronouns – and not other contextual coordinates, like Addressee and Time. Thus (27) illustrates that the Time indexical in Tamil doesn't shift – a point that is easy to show for other types of indexicals like Addressee and World:

(27) seetha krishnan naaḷeikki school.ukku poo.v.aan-nnu so.nn.aaḷ
seetha krishnan tomorrow school.DAT go.FUT.3MSG-that say.PST.3FSG
"Seetha said $[_{CP}$ that Krishnan will go to school tomorrow.]"
tomorrow → $\text{TOMORROW}_{\{Utt-time, *Att-time\}}$

Thus, I propose, in line with Anand (2006) for Malayalam, that the Tamil diagonalization operator:

(28) $[\text{OP}_{diag-Tamil}]^{c,i,g} = [\text{OP}_{auth}]^{c,i,g}$

(29) $[\text{OP}_{auth}\ (\alpha)]^{c,i,g} = [\![\alpha]\!]^{j,i,g}$, for j = $< Auth(i), Addr(c), Time(c), World(c) >$

But there is still one other complication – namely that not all 1st-person indexicals shift in Tamil. Thus, in (30) below which forms a minimal pair with (23), the 1st-person indexical pronoun *naan* (I) is unshifted and may only refer to the author of the utterance-context (which happens to be me, Sandhya).

(30) raman naan poo.n.een-nnu so.nn.aan
raman I go.FUT.1SG-that say.PST.3MSG
"Raman said $[_{CP}$ that $\text{I}_{sandhya}$ would go]" (INDIRECT SPEECH)
"Raman said, "I went" (DIRECT SPEECH/QUOTATIVE)

To account for this option, I propose the Spell-Out rules below. Note that such an option also entails that the context-overwriting operator be *syntactically* represented in order to

be able to influence Spell-Out rules.

[AUTHOR (c)]^{c,i,g} → SPELL-OUT: *naan* (**Unshifted**)

[OP$_{auth}$ AUTHOR (c)]^{c,i,g} = [AUTHOR (i)]^{i,i,g} → SPELL-OUT: *taan* (**Shifted**)

The various Tamil patterns can now be derived as follows:
Shifted reading: (23):

(31) [Raman said$_{3sg}$ [$_{CP}$ OP$_{Auth}$ I went$_{1sg}$]]^{c,i,g}
= [Raman said$_{3sg}$ [$_{CP}$ AUTHOR(i') went$_{1sg}$]]$^{i',i',g}$
= Raman said$_{3sg}$ [$_{CP}$ *taan*$_{\{raman\}}$ went$_{1sg}$]$^{i',i',g}$

Unshifted reading: (30):

(32) [Raman said$_{3sg}$ [$_{CP}$ I went$_{1sg}$]]^{c,i,g}
= [Raman said$_{3sg}$ [$_{CP}$ AUTHOR(c) went$_{1sg}$]]^{c,i,g}
= Raman said$_{3sg}$ [$_{CP}$ *naan*$_{\{sandhya\}}$ went$_{1sg}$]^{c,i,g}

References

Anand, Pranav. 2006. De de se. Doctoral Dissertation, Massachusetts Institute of Technology.

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, Kashi Wali, James W. Gair, and K.V. Subbarao, 169–216. Mouton de Gruyter.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz. Cambridge, Mass.: MIT Press.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective.* Oxford Studies in Comparative Syntax. Oxford University Press.

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: Essays in linguistics in honor of Sylvain Bromberger*, ed. Ken Hale and Samuel Jay Keyser. Cambridge, Mass.: MIT Press.

Heim, Irene. 1994. Comments on Abusch's theory of tense. In *Ellipsis, tense, and questions*, ed. Hans Kamp, 143–170. University of Amsterdam.

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

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

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

Huang, James, and Luther Liu. 2001. Logophoricity, attitudes, and *ziji* at the interface. In *Long-distance reflexives*, ed. Peter Cole, Gabriella Hermon, and James Huang, 141–195. New York: Academic Press.

Hyman, Larry M., and Bernard Comrie. 1981. Logophoric reference in Gokana. *Journal of African Languages and Linguistics* 3:19–37.

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

Kratzer, Angelika. 1998. More structural analogies between pronouns and tenses. In *Proceedings of SALT VIII*. MIT.

Kratzer, Angelika. 2009. Making a pronoun: fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40.

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

Partee, Barbara. 1973. Some structural analogies between tenses and pronouns in English. *Journal of Philosophy* 70:601–609.

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. Wendy K. Wilkins Simin Karimi, Vida Samiian, 262–294. Amsterdam: John Benjamins.

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

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

Rice, Keren. 1986. Some remarks on direct and indirect speech in Slave (Northern Athapaskan). In *Direct and Indirect Speech*. Mouton de Gruyter.

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

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. 1999. Propositional attitudes and indexicality: a cross-categorial approach. Doctoral Dissertation, MIT.

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

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

Stalnaker, Robert. 1981. Indexical belief. *Synthese* 49:129–151.

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. *Journal of South Asian Linguistics* 2.

Svenonius, Peter, and David Adger. To Appear. Features in Minimalist Syntax. In *Oxford Handbook of Linguistic Minimalism*. (See also: *http://ling.auf.net/lingBuzz/000825*).

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

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