JSAL

Volume 2, Issue 1 December 2009

Journal of South Asian Linguistics

CSLI PUBLICATIONS

Contents

Editorial Statement

At the close of the year, we are happy to be able to bring out the second issue of the *Journal of South Asian Linguistics*. This issue includes a total of four articles, three of which are on Dravidian languages (Malto, Tamil and Telugu) and one of which is on Persian. Strictly speaking, Persian is of course not a South Asian language. However, given that Persian is an Indo-Iranian language and given that Persian has been in close language contact with South Asian languages for millenia and that therefore South Asian languages and Persian share many linguistic characteristics, we have decided that Persian clearly also lies within the remit of our journal. Indeed, the paper on Persian by Marina Pantcheva is on Persian complex predicates, which show many structural similarities to South Asian complex predicates. The other three papers deal with verbal structure in Malto (by Chaithra Puttaswamy), and control constructions in Tamil (by Sandhya Sundaresan and Thomas McFadden) and Telugu (by Youssef Haddad), respectively.

Surprisingly, at least to us, this issue thus contains no papers on Indo-Aryan languages. This fact is surprising because the literature on South Asian languages tends to be dominated by papers written about Indo-Aryan languages. We therefore see the composition of this issue as a welcome development and as a sign that this Journal, JSAL, really is providing a much needed forum.

In the initial announcement of the journal, we had indicated that we hoped to include theoretically oriented descriptions of underdocumented languages. In this volume, we are excited to present the first of these theoretically oriented descriptions: a detailed description of verbal morphology in Malto, a severely underdocumented North Dravidian language by Chaithra Puttaswamy.

Work on the next issue is already well underway. So far, the balance is 50-50 between Indo-Aryan and Dravidian in the papers that we have already accepted. However, we note that we have no papers so far on Tibeto-Burman, Dardic or Munda languages. Of course, how the journal will develop further and what mix of South Asian (and adjacent) languages we are able to publish papers on is up to you: our readers.

Please do send us your papers and if you see a paper that will be relevant, please tell us about it and the author about the journal. JSAL is intended to provide a venue for dissemination of theoretical proposals inspired by data from the South Asian languages as well as theoretically informed descriptive work on the South Asian languages. Many a paper on South Asian languages lies lost in the author's filing cabinet or does not make it beyond the talk/handout stage for want of a proper venue. So please help us get the word out! However, do note that we are only able to accept papers that meet the standards of high international quality. So before sending us papers, do also make sure that you are aware of the international state-of-the-art and that your paper indeed makes a significant contribution to the state-of-the-art on an international level.

In closing, we would like to thank a number of people. As always, Dikran Karaguezian of CSLI Publications for taking on board on-line journal publication as part of the activities of CSLI and for working with us on making this journal a reality. We would also like to thank the University of Konstanz for supporting the close editorial connection between CSLI Publications and Konstanz by financing a research assistant to help with the editing and publishing of books and journals which take shape as part of the CSLI-Konstanz cooperation. In terms of this second issue of JSAL, we would also like to thank our reviewers for taking on the additional work involved in reviewing papers for us, Fatemeh Nemati for help with editing, Jaouad Mousser for help with type-setting and Sebastian Sulger for help with editing, type-setting and the continued maintenance of the Journal website and server. Indeed, we migrated JSAL to a new server this year in order to allow faster and better access to the papers published in the journal and hope you will have already appreciated this improvement!

Rajesh Bhatt, University of Massachusetts, Amherst

Miriam Butt, University of Konstanz

Subject Distribution in Tamil and Other Languages: Selection vs. Case

Sandhya Sundaresan and Thomas McFadden, *CASTL, University of Tromsø/ University of Stuttgart*

Received February 2009; Revised November 2009

Abstract

This paper presents an alternative account of DP distribution that is based on DPs being selected rather than being Case-theoretically licensed. We argue that the fundamental prediction made by Case theory, namely that obligatorily controlled PRO and overt DPs are in complementary distribution, is not empirically justified. To this end, we provide data from non-finite clausal adjuncts, complements and gerundivals in Tamil where subject controlled PRO and overt subject DPs seem to alternate in free variation. We further illustrate, with supporting evidence from Malayalam, Sinhala, Latin, Irish, and Middle English as well as the Present-Day English gerundival construction, that this type of problematic alternation is not a language-specific quirk but a widely attested crosslinguistic phenomenon. While standard Case theories are equipped to handle *either* the occurrence of PRO or that of an overt subject, they are unable to consistently handle the alternation between both types of elements. Our selection analysis is designed to handle the alternations as well as instances where only one DP type is allowed.

1 Introduction

In Tamil,[1] simple infinitive clauses can function as adjuncts with purpose or temporal interpretation, as in (1). Such infinitives can appear with an implicit subject which has to be coreferent with a matrix argument — i.e. we get obligatory control PRO, as in (1a). However, it is also possible to have an overt non-coreferent subject in the nominative case, as in (1b).

(1) a. raman$_i$ [PRO$_{i/*j}$ puuri porikk-a] maavu vaangi-n-aan
Raman.NOM PRO puuri.ACC fry-INF flour.ACC buy-PST-M.3SG
'Raman bought flour to fry puuris.' (PURPOSIVE)
'Raman bought flour while frying puuris.' (TEMPORAL)

b. raman [**vasu** puuri porikk-a] maavu vaangi-n-aan
Raman.NOM Vasu.NOM puuri.ACC fry-INF flour.ACC buy-PST-M.3SG
'Raman bought flour for Vasu to fry puuris.' (PURPOSIVE)
'Raman bought flour while Vasu fried puuris.' (TEMPORAL)

[1] One of the four major languages of the Dravidian family, spoken, among other areas, primarily in parts of South India, Sri Lanka, and Singapore, Tamil is a subject *pro*-drop language with rich case and agreement morphology, wide scrambling, and SOV clause structure.

JSAL Volume 2, Issue 1, December 2009.

Such data are problematic for standard theories of DP distribution based on abstract Case, as it is not clear what could be assigning or checking Case to license the overt subject in (1b).[2] Indeed, we have an alternation between PRO and overt subjects in such clauses which does not seem to correlate with any independently identifiable morphological, syntactic or semantic property. As Case theory is standardly based on the assumption that PRO and overt DPs are in complementary distribution, this presents a puzzle.

In this paper we examine these and related data in more detail and explore the implications they have for syntactic theory. To this end, we present new data from Tamil non-finite complements, adjuncts and gerundivals. However, the conclusions we draw have more general theoretical relevance, and we will present considerable additional data from other languages to show that the patterns we report are not quirks of Tamil which may call for a language-specific explanation but a more widespread issue that is attested in several languages. We begin in Section 2 with a brief review of standard Case-theoretic assumptions about the licensing of DPs and the basic data that motivated them. Then in Section 3 we present the main Tamil data and explain why they are problematic for Case theory. In Sections 4 and 5 we bring in comparative data to support the conclusions we draw on the basis of Tamil, first from the South Asian languages Malayalam and Sinhala, and then from several European languages such as Latin, Irish, Middle English, and Present-Day English. We briefly consider and reject three possible ways to accommodate the problematic data within Case theory in Section 6, including discussion of the previous theoretical treatment of some of the Tamil data by Sarma (1999). In Section 7, we propose and motivate an alternative analysis in which the distribution of DPs is based on selectional requirements of lexical and funcional heads rather than licensing needs of the DPs themselves. Finally, Section 8 presents a summary of our data and analysis and their relevance for syntactic theory.

We will ultimately be arguing that a simple (binary) notion of finiteness is insufficient for understanding DP distribution. However, in order to avoid confusion we will continue to use standard terminology in the more descriptive portion of this article in Sections 3—5. That is, for now we will still talk of infinitives, and finite and non-finite clauses, as identified on essentially morphological grounds.[3] These are crucially intended only as convenient descriptive labels. When we present our analysis in Section 7, we will propose a more nuanced scale of dependency to replace finiteness in the discussion of DP distribution.

2 Background: standard Case theory

To set the stage, it will be helpful to consider the main ideas and motivation behind accounts of overt DP distribution in terms of abstract Case.[4] The fundamental assumption is that overt DPs are somehow defective and need help in order to be licensed. Specifically, they can only appear in places where some element can assign or check their Case. In all other contexts, overt DPs are disallowed. Object DPs are assumed to get Case from the verb and the functional heads related to voice and aspect found immediately above, so their licensing depends on the identity and properties of the lexical verb and the voice and aspect of the clause. Subjects, on the other hand, are assumed to get Case from the functional heads in the Infl complex or further up in the left periphery. Subject licensing and distribution are thus directly related to the finiteness of the clause. Specifically, finite

[2] We will follow the common convention of using capitalization to help distinguish abstract syntactic **Case** — the theoretical construct used to account for DP distribution in terms of licensing — from morphological **case** — the typically overt marking on DPs in many but not all languages. Hence we refer to nominative case to describe the overt form of *Vasu* in (1b), but abstract Case below when discussing how the same DP might be licensed.

[3] E.g. Tamil "infinitives" are those forms built by suffixing *-a* to the primary stem. "Finite" verbs are any forms bearing both tense and agreement suffixes. We do not necessarily expect that either of these morphological categories will display unified syntactic behavior as regards subject type or anything else for that matter.

[4] Case theory has of course undergone significant revisions since it was first proposed early in the Government and Binding era. In this section and elsewhere, we will try as much as possible to focus on the foundational ideas that are common to all vintages of Case theory and abstract away from points of detail where they differ from each other. Where it is necessary to focus on one particular version or non-universal detail, we will be as explicit as possible.

inflection (in T, Agr for instance) is assumed to assign nominative Case to the subject position, so overt subjects are licensed in finite clauses, whether matrix or embedded, as in (2):

(2) Mitch believes [that Lazlo lives in the steam tunnels].

Infinitives, on the other hand, lack this inflection and thus do not assign nominative Case. This means that, by default, overt subjects are impossible in infinitives, as the following examples show:

(3) a. * [Lazlo to live in the steam tunnels] would be strange.
b. * Mitch tried [Lazlo to live in the steam tunnels].

However, there are infinitives with overt subjects. In English these include infinitives that are introduced by the prepositional complementizer *for*, as in (4a), and those that follow so-called ECM (exceptional Case-marking) verbs like *believe*, as in (4b).

(4) a. [**For** Lazlo to live in the steam tunnels] would be strange.
b. Mitch **believes** [Lazlo to live in the steam tunnels].

This is where Case theory gets interesting. The claim is that elements like *for* and *believe* assign Case to the following subjects, thereby allowing them to be overt even though the clause is non-finite.[5] This idea has some empirical plausibility, since prepositions and transitive verbs do determine the morphological case on following DPs in languages with rich case systems. In German, for example, the object of the preposition *mit* 'with' is marked dative, while that of *ohne* 'without' is accusative. Similarly, among verbs *helfen* 'help' takes a dative object while *unterstützten* 'support' takes an accusative. So what we see on the surface in German is supposed to happen abstractly in English infinitives.

Crucially, analogous case-assigners are conspicuously lacking in the sentences in (3) above where an overt subject was impossible. If the subject is instead left non-overt in such sentences, the result is grammatical:

(5) a. [PRO$_{arb}$ to live in the steam tunnels] would be strange.
b. Mitch$_i$ tried [PRO$_{i/*j}$ to live in the steam tunnels].

So it appears that an overt subject is possible in non-finite clauses just where a Case-assigner is available, and where one is not available, the subject must be non-overt PRO. If the Case requirement has something to do with (overt) morphology, it is perhaps plausible that it should treat a silent element like PRO differently. In later versions of Case theory, it has been commonly assumed that PRO does in fact get a special kind of Case called null Case, which licenses PRO and nothing else (see e.g. Chomsky and Lasnik 1993, Martin 2001). For our purposes this still means that PRO and overt subjects are distinguished in terms of Case.

There is now an extensive body of work which has identified serious problems with Case theory, either proposing significant revisions or arguing that it should be abandoned entirely. Some representative contributions in this area are Zaenen et al. (1985), Yip et al. (1987), Marantz (1991), McFadden (2004), Landau (2006) and Sigurðsson (2008, 2009). We intend the current paper to be understood as a continuation of this tradition, adding to the case against abstract Case. However, we will be presenting new evidence which allows a novel kind of argument against Case.[6] We also

[5]The term ECM refers to the fact that the structural configuration for Case-assignment here is a bit different from that normally found with DP objects of verbs and prepositions. While the latter are generally assumed to be the complements of their Case-assigners, the embedded subjects in sentences like (4a) and (4b) are in the specifier position of the phrase that is the complement of *for* and *believes* respectively. This situation is less awkward in more recent versions of Case theory within Minimalism: these assume that Case assignment or checking operates via Agree, which in turn depends not on a specific structural configuration like specifier or complement but on minimal c-command, which would be equally satisfied in all relevant constellations.

[6]Much of the previous work was devoted to showing that case morphology is dissociated from DP licensing in ways that go against the predictions of Case theory, often concentrating on so-called quirky Case phenomena. We will only tangentially touch on that topic, concentrating rather on problems in the distribution of PRO and overt DPs that are independent of case morphology.

present a distinct proposal about what should replace abstract Case in places where it did work within the theory. For these reasons we will not provide significant discussion of the earlier work in this vein, directing the reader instead to the citations above.

3 Tamil non-finite clauses and the licitness of subjects

In this section, we present a detailed description of non-finite clausal structures attested in Tamil, specifically: obligatory control complements (such as those of 'try'-class verbs), complements allowing both controlled PRO and overt embedded subjects (such as those of 'want'-class verbs), purposive and temporal adjunct infinitives as well as gerundivals showing the same alternation. At the end of this section, we also present evidence to show that the null coreferent element in the subject position of the embedded clauses here is controlled PRO and not little *pro* and that the relative hierarchy of arguments is as presented and not solely the result of scrambling.

3.1 Obligatory control complement infinitives

The first type of Tamil infinitive we'll look at appears as the complement of verbs like *paar-* 'try'.[7] Such infinitives require a non-overt subject, which is obligatorily coreferent with the matrix subject, as in (6a). Adding an overt subject, as in (6b) yields ungrammaticality:

(6) a. raman$_i$ [PRO$_{i/*j}$ saadatt-ai saappiḍ-a] paa-tt-aan
Raman.NOM PRO rice-ACC eat-INF try-PST-3m.sg
'Raman tried to eat rice.'

b. * raman [anand saadatt-ai saappiḍ-a] paa-tt-aan
Raman.NOM Anand.NOM rice-ACC eat-INF try-PST-3m.sg
'Raman tried Anand to eat the rice.'

This pattern taken on its own fits in very nicely with standard Case theory. The complement clause is non-finite, so by default an overt subject should be impossible. Furthermore, there is no special Case licensor like a potential ECM verb or prepositional complementizer like English *for* to override this default and exceptionally license an overt subject. As Case theory predicts, we instead get obligatorily controlled non-overt subject PRO.

Indeed, what we see here is entirely parallel to the behavior of infinitives embedded below obligatory subject control verbs in English, often called the *try*-class. Sentences (3b) and (5b) discussed in Section 2 above are examples of this type, as are those in (7):

(7) a. John$_i$ tried [PRO$_{i/*j}$ to eat turkey]
b. * John tried [Bill to eat turkey]

(7a) shows that the infinitival complement of *try* can have a covert subject which is coreferent with the matrix subject, while (7b) shows that an overt subject in such a clause is ruled out. Tamil obligatory control complement infinitives thus behave just like one of the classes of English infinitives that is central to the motivation for abstract Case. So far, then, Tamil presents no problem for Case theory.

3.2 Alternating complement infinitives

A second type of infinitive clause in Tamil appears as the complement of verbs like *veṇḍ-* 'want'. Verbs like *veṇḍ-* take a dative subject which co-occurs either with a nominative object, as in (8a), or with an infinitival complement, as in (8b):

[7] We use the primary stem as the citation form for Tamil verbs, which has all stem-forming and inflectional material stripped off. The data are based on the native-speaker intuitions of the first author and are essentially from Spoken Tamil, although Written Tamil forms have been used where necessary to make the morphological structure more clear.

(8) a. **Nominative object:**
champa-vukku oru samosa veɳɖ-um
Champa-DAT a samosa.NOM want-N.3SG
'Champa wants a samosa.'

b. **Infinitival complement with PRO:**
champa-vukku$_i$ [PRO$_i$ oru samosa-vai saappiɖ-a] veɳɖ-um
Champa-DAT PRO a samosa-ACC eat-INF want-N.3SG
'Champa wants to eat a samosa.'

In (8b), the embedded subject is non-overt and is again obligatorily coreferent with the matrix subject. However, an overt (noncoreferent) nominative DP subject is also licit, as in (9):

(9) **Infinitival complement with overt nominative DP:**
champa-vukku [sudha oru samosa-vai saappiɖ-a] veɳɖ-um
Champa-DAT Sudha.NOM a samosa-ACC eat-INF want-N.3SG
'Champa wants Sudha to eat a samosa.'

This pattern is again very similar to something we find in English, namely infinitives appearing as the complements of so-called *want*-class verbs, as in (10):

(10) a. Sue$_i$ wanted [PRO$_{i/*j}$ to drink beer]
b. Sue$_i$ wanted [Jill/her$_{*i/j}$ to drink beer]

The presence of PRO as a non-finite subject in (10a) is quite unproblematic within standard Case theory given the assumption that non-finite clauses normally cannot take overt subjects because they lack adequate Case-assigners. But under this assumption, the presence of the overt DP in (10b) is entirely unexpected. The English pattern has traditionally been analyzed in two different ways within Case theory. Either verbs like *want* can license an embedded overt subject via ECM, or there is a null variant of the prepositional complementizer *for* which can do the same thing (see e.g. Bošković 1997, Martin 2001, for discussion of these possibilities). In either case, some sort of optionality is required, since the conditions for overt DP licensing must obtain in (10b) but not in (10a), where the distinct requirements for PRO licensing must hold instead.[8]

The Tamil data are similar to English, with the following important differences. First, the embedded overt DP in Tamil is marked nominative and not accusative. Second, Tamil also allows dative subjects in the infinitive if the embedded verb is "quirky"-dative assigning (like *puriy-* 'understand'):

(11) **Infinitival complement with overt dative DP:**
champa-vukku [sudha-vukku viʃiyatt-ai puriy-a] veɳɖ-um
champa-DAT sudha-DAT the.matter-ACC understand-INF want-N.3SG
'Champa wants Sudha to understand the matter.'

[8]Instead of optionality, one could assume that the examples in (10a) and (10b) have different syntactic structures due to a different semantics for *want* in each case. Accordingly, we might propose two alternative structures for (10b): an object control structure or an ECM structure (Bošković 1997, Martin 2001), as follows:
(i) Sue$_i$ wanted Jill$_j$ [$_{CP}$ PRO$_{*i/j}$ to drink beer] (Object Control)
(ii) Sue$_i$ wanted Jill$_j$ [$_{TP}$ t$_j$ to drink beer] (ECM)
But an immediate problem arises with the above alternative structures: if *Jill* is in the matrix clause as a direct object, we would expect it to be passivizable. But this prediction is not borne out:
(iii) *Jill was wanted to drink beer
In contrast, the direct object DP in both object control- and ECM structures can be passivized, just as expected:
(iv) Matthias$_i$ was persuaded t$_i$ [$_{CP}$ PRO$_i$ to attend the AA meeting]
(v) Matthias$_i$ was expected t$_i$ [$_{TP}$ t$_i$ to become an alcoholic]
Given such immediate empirical problems with potential alternative structures for problematic constructions like (10b), we for now do not pursue this line of argumentation. Instead, we assume that (10a) and (10b) have the same syntactic structure, as discussed.

The example below shows that *puriy-* 'understand' takes a dative subject in a finite clause:

(12) **Finite clause with dative DP subject:**
champa-vukku viʃiyatt-ai puri-nd-adu
champa-DAT the.matter.ACC understand-PST-N.3SG
'Champa understood the matter.'

The possibility of structures like in (11) above suggests that the morphological case on the embedded subject has something to do with the embedded predicate and is, as such, being supplied internal to the embedded clause. Our analysis of overt DP subjects in non-finite complement clauses should therefore extend this analysis to the case-marking in structures like (9), where a nominative subject shows up in the same position. This in fact, as we discuss later, is the kind of analysis that Gair (2005) makes for corresponding clausal structures in Sinhala.

Sarma (1999) proposes a Case-theoretic account for the kind of Tamil data shown in (9) which is analogous to the ECM analysis for English. Thus at first glance these data might not seem to present a challenge for Case theory. However, as we will argue below, neither treatment of English *want*-class complements is convincing, nor is Sarma's story for Tamil. The main issue will turn out to be that Case theory is not designed to accommodate an alternation between overt subjects and PRO in a single context as we find here. Rather, complementary distribution of the two is expected; thus, alternating complement infinitives will present a problem.

3.3 Adjunct infinitives

A third type of Tamil infinitive clause, which we briefly introduced at the beginning of the article, appears as an adjunct in the matrix clause, with either a temporal or a purposive interpretation. These too allow both obligatorily controlled PRO and overt DPs in the embedded subject position. Consider the examples in (13):

(13) a. PRO:
[PRO$_{i/*j}$ saadatt-ai saappiḍ-a], naan$_i$ veḷiya poo-n-een
PRO rice-ACC eat-INF I.NOM outside go-PST-1SG
'I went out (in order) to eat rice.' PURPOSE INTERPRETATION
'As I ate rice, I went out.' TEMPORAL INTERPRETATION

b. **Overt subject:**
[avan saadatt-ai saappiḍ-a], naan veḷiya poo-n-een
he.NOM rice-ACC eat-INF I.NOM outside go-PST-1SG
'I went out (in order) for him to eat rice.': PURPOSE INTERPRETATION
'As he ate rice, I went out.': TEMPORAL INTERPRETATION

From the perspective of Case theory, (13a) is as expected. The embedded clause is non-finite, so by default a PRO subject is predicted. The fact that we get an overt subject *avan* 'he' in (13b), on the other hand is, however, quite surprising, as it is not at all clear what could be licensing it. Just as with the PRO example in (13a), special Case-licensors seem to be absent. First, there is no evidence for a prepositional complementizer like *for* to license the overt subject.[9] Second, the matrix verb *pooneen* 'went.1sg' is clearly not an ECM verb: standard ECM verbs are transitive, whereas this verb is intransitive. Third, the matrix verb does not c-command the embedded subjects in (13a) and (13b), since the infinitival clauses are adjuncts. So the structural conditions for ECM are not met. It appears then, that Tamil adjunct non-finite clauses with overt subjects pose a serious problem for standard theories of Case. Furthermore, we have the same problem here as with the alternating complements: we do not just get unexpected overt subjects, but an alternation between overt subjects and PRO. This alternation again will prove to be a challenge for Case theory.

[9] Postpositions and prepositions in Tamil typically take oblique DP complements whereas in the types of structures under consideration here, the DP is marked nominative.

3.4 Gerundivals

Gerundival constructions in Tamil are also similar in important respects to the adjunct infinitival clauses. Instead of the infinitive form of the verb, these are built around a gerundival clause:

(14) raman veelai-ai **sey-v-ad-ukku** college-ukku poo-n-aan
Raman.NOM work-ACC do-FUT-GER-DAT college-DAT go-PST-M.3SG
'Raman went to college for doing work.'

The gerundival element, boldfaced in the example above, is overtly marked for future tense as well as for dative case.[10] The dative marking contributes the meaning of purpose whereas the future tense marking expresses an unmarked temporal interpretation. Although the future tense marker shown in (14) is the most common form used in this type of construction (presumably because the purposive construction most often refers to an event that still needs to be completed relative to the time-frame of the main event in the clause), it can be readily replaced by either present or past forms. Additionally, no sequence-of-tense (SOT) effects (Enç 1987, Abusch 1997, and many others) are found: the tense of the gerudival clause is not anaphoric on that of the "matrix" clause in any way and can be varied independently from it. The structure in (15) below shows that the tense of the gerundival can be other than future while that in (16) illustrates that this tense can, furthermore, be different from that of the "main" clause:

(15) **Past-gerundival under past**
raman$_i$ [PRO$_{i/*j}$ veelai-ai **sey-d-ad-ukku**] paris-ai **vaangi-n-aan**
Raman.NOM PRO work-ACC do-PST-GER-DAT prize-ACC get-PST-M.3SG
'Raman got a prize for having done the work.'

(16) **Past-gerundival under future**
raman$_i$ [PRO$_{i/*j}$ veelai-ai **sey-d-ad-ukku**] paris-ai **vaangu-v-aan**
Raman.NOM PRO work-ACC do-PST-GER-DAT prize-ACC get-FUT-M.3SG
'Raman will get a prize for having done the work.'

As concerns interpretation, the subject of the gerundival clauses in (15) and (16) is a null element that is obligatorily coreferent with the matrix subject *Raman* — i.e. obligatorily controlled PRO. The purposive interpretation is conveyed by the dative marker on each gerund, but the tense meanings vary according to the tense marker used. That the tense on the gerundival can be varied independently from that of the "matrix" clause demonstrates that the former is real semantic tense and neither lexicalized/"dead" nor syntactico-semantically anaphoric in any way, as has been argued for structures with SOT effects. Nevertheless, there is still a difference between the T head in such constructions and those in standard finite clauses: the T head within the gerundival structure is deficient because it does not bear overt subject agreement. The reason such tensed gerundivals are interesting for our purposes is that they allow not only a PRO subject coreferent with the matrix subject, but also an overt non-coreferent subject, as in (17):

(17) **Past-gerundival with overt subject:**
⌈raman veelai-ai sey-d-ad-ukku⌋ sudha paris-ai vaangi-n-aa⌊
Raman.NOM work-ACC do-PST-GER-DAT Sudha.NOM prize-ACC get-PST-F.3SG
'Sudha got a prize for Raman's having done the work.'

Note here that in contrast to the null subject structure in (15), the matrix verb in (17) above does not agree with *Raman* but with the feminine subject *Sudha*. *Raman* is the agent in the event described by the gerundival, namely the event of doing the work and, thus, a part of this gerundival structure. What makes it potentially problematic for Case theory is that the DP *Raman* shows up with nominative case, not genitive, the case we would expect if it were introduced outside the

[10] The dative marker seems to be functioning like a prepositional complementizer (e.g. English *for*) here, marking the entire gerundival clause.

gerundival and which could be accounted for in standard Case theories (see e.g. Chomsky 1981). What could be assigning nominative case to *Raman*? It could not be the finite T of the main clause, because this should already be checking off its nominative case feature against the main nominative subject *Sudha*, which the main verb agrees with. Instead, it looks like what is being gerundivalized is an entire clause including the subject *Raman*, which means that *Raman* must be getting its nominative internal to this gerundival clause. But how is this licensed? We could argue that the tense within the gerundival is responsible for structural nominative Case, licensing *Raman*. But there are a couple of problems with this. One is that the gerundival structure does not appear to be a fully finite clause. Although tense is overtly marked, there is, as we've noted, no overt subject agreement. It is not clear that we should expect a thus deficient T to be able to check Case on *Raman* in the standard way.[11] The other problem is that this type of gerundival structure again does not have to show up with an overt DP but can alternate with obligatorily controlled PRO in the same position as the overt DP. This is an issue because, while either of the structures in (15) or (17) can be individually explained within the premises of Standard Case theory, a unified analysis of both structures would be unfeasible. For instance, the structure with the overt DP in (17) could be explained in the way we've already described, namely by saying that the subject of the gerundival clause *Raman* gets Case from the (semi-)finite T within the gerundival structure. For the structure in (15), we could say that the PRO gets special null Case from the deficient T within the gerundival clause. But having both possibilities requires the same T to have two different statuses.

3.5 Brief excursus: evidence for obligatory control PRO vs. little *pro*

In all the examples discussed so far, we have tacitly assumed that the non-overt embedded subject is controlled PRO. For a null subject language with rampant scrambling, such an assumption is, however, not trivial. Thus we will now present two pieces of evidence for the presence of obligatory control PRO: obviation of Weak Crossover (WCO) effects (also termed the "PRO-gate" effect), and obligatory coreference effects. Then we will give data showing that in the non-finite adjunct and complement structures under investigation, the relevant arguments are indeed in the embedded clause, not the matrix clause.

3.5.1 Obviation of WCO effects

Jaeggli and Safir (1989) note that controlled PRO, but not little *pro*, can obviate WCO effects — in other words, obligatory control PRO seems to function as a "gate" for WCO, rendering a normally ungrammatical phenomenon, grammatical (the observation being originally due to Higginbotham 1980). This is shown below for English:

(18) * $[_{CP}$ Who(m)$_i$ did $[_{DP}$ John$_i$/him$_i$ washing his$_i$ car] upset e$_i$]?
(19) $[_{CP}$ Who(m)$_i$ did $[_{DP}$ PRO$_i$ washing his$_i$ car] upset e$_i$]?

The ungrammaticality in (18) is held to arise because the *wh*-operator "crosses over" the coreferent possessive pronominal *his* in the DP *his car* while raising to matrix [Spec, CP]: the term "Weak Crossover (WCO)" refers to this phenomenon and the term "WCO-effect" standardly refers to the ungrammaticality resulting from it. In (19) there is also WCO, but the structure is nevertheless grammatical. Since the structures in (18) and (19) essentially form a minimal pair, with the only difference between them being the identity/nature of their gerundival subjects, the conclusion is that it is the nature of the gerundival subject in (19) — specifically the presence of PRO — that allows WCO in this structure. Jaeggli and Safir (1989) show, furthermore, that in *pro*-drop languages like Spanish, *pro* patterns with overt pronouns and not with PRO.[12] The Spanish examples from Jaeggli

[11] A reviewer points out that what we characterize here as tense markers might actually be aspect, as suggested for related Dravidian languages, Kannada and Malayalam, in Amritavalli and Jayaseelan (2005). We will not take a strong position on this distinction, but note that our arguments here would not be weakened either way, since aspect is not standardly thought to license overt subjects within Case theory, and in any case the problematic alternation would remain.

[12] This shows, in turn, that the obviation of WCO effects is not merely a phonological matter: specifically, it shows that it is not the mere absence of phonological material in gerundival-subject-position that obviates WCO-effects.

and Safir (1989) are reproduced below (formatting ours):

(20) ?/* A-quién$_i$ acusó la mujer$_j$ que$_j$ bailó con él$_i$?
whom$_i$ accused the woman who danced with him
'[$_{CP}$ Who(m)$_i$ did [$_{DP}$ the woman$_j$ [$_{CP}$ who$_j$ [$_{TP}$ e$_j$ danced with him$_i$]]] accuse e$_i$]'?

(21) ?/* A-quién$_i$ acusó la mujer$_j$ con quien bailó?
whom$_i$ accused the woman with whom danced
'[$_{CP}$ Who(m)$_i$ did [$_{DP}$ the woman$_j$ [$_{CP}$ with who(m)$_j$ [$_{TP}$ *pro*$_i$ danced e$_j$]]] accuse e$_i$]'?

The example in (20) shows that a *wh*-operator may not cross over an overt pronoun, even when this is contained within another DP. The structure in (21) shows that this same ungrammaticality arises when this pronoun is covert. Jaeggli and Safir (1989) thus take this to be a diagnostic for PRO vs. *pro*: specifically, a null element that obviates WCO-effects is PRO; one that preserves the ungrammaticality due to WCO is little *pro*. Implementing this useful diagnostic for Tamil, we obtain the following in the control structures in (22) and (23). WCO-effects obtain in (22) (which has an overt pronoun *avan* as the gerundival subject), but do not obtain in (23) where the gerundival subject is covert and corefers with the *wh*-operator *yaarai* 'whom'. From this, we conclude that the silent gerundival subject in (23) is PRO and not *pro*.

(22) * yaar-ai$_i$ [avan$_i$ tann-ooɖa$_i$ car-ai alambinadu] t$_i$ sandooʃappaɖutt-idu?
who-ACC$_i$ him$_i$ self-GEN$_i$ car-ACC having-washed made-happy-3N-SG
'[$_{CP}$ Who(m)$_i$ did [$_{DP}$ him$_i$ having washed his$_i$ car] make happy?]'

(23) yaar-ai$_i$ [PRO$_i$ tann-ooɖa$_i$ car-ai alambinadu] t$_i$ sandooʃappaɖutt-idu?
who-ACC$_i$ PRO$_i$ self-GEN$_i$ car-ACC having-washed made-happy-3N-SG
'[$_{CP}$ Who(m)$_i$ did [$_{DP}$ PRO$_i$ having washed his$_i$ car] make happy?]'

3.5.2 Obligatory coreference effects

The second type of evidence we present in favor of obligatory controlled PRO vs. little *pro* is that of obligatory coreference readings between the silent embedded subject and the matrix antecedent. We show this with all four types of constructions in Tamil discussed in this paper which involve a silent embedded subject:

(24) **Non-finite controlled clausal complement (*paar-* 'try'-class):**
raman$_i$ [PRO$_{i/*j}$ kaapi-ai kuɖikk-a] paattaan
Raman.NOM PRO coffee-ACC drink-INF tried.3m.sg
'Raman tried to drink coffee.'

(25) **Non-finite purposive clausal adjunct:**
raman$_i$ [PRO$_{i/*j}$ pariccai-ai erud-a] school-ukku poo-n-aan
Raman.NOM PRO exam-ACC write-INF school-DAT go-PST-M.3SG
'Raman went to school to write the exam.'

(26) **Non-finite clausal complement (*veɳɖ-* 'want'-class):**
champa-vukku$_i$ [PRO$_{i/*j}$ kaapi-ai kuɖikk-a] veɳɖ-um
Champa-DAT PRO coffee-ACC drink-INF want-N.3SG
'Champa wants to drink coffee.'

(27) **Past gerundival:**
sudha$_i$ [PRO$_{i/*j}$ avvaɭavu kaapi-ai kuɖi-tt-ad-ukku] paris-ai vaangi-n-aaɭ
Sudha.NOM PRO so.much coffee-ACC drink-PST-GER-DAT prize-ACC get-PST-F.3SG
'Sudha got a prize for having drunk so much coffee.'

As the indexation shows, in all the structures attested in (24)–(27), the covert subject has to be coreferent with the matrix subject and not with any other entity. Before we reach a decisive conclusion from these data, a little more needs to be clarified. While the impossibility of a non-coreferent reading is probably expected in non-alternating structures such as *try*-class complements

(24), it is unexpected for alternating non-finite clauses such as purposive/temporal adjunct clauses (25), *want*-class complements (26), and gerundivals (27): if a (trivially) non-coreferent overt DP can show up in non-finite subject position, then why not non-coreferent *pro*? This is especially puzzling given that Tamil is a null-subject language with exhibits rampant *pro*-drop in matrix and crucially also embedded subject position:

(28) *pro* sadatt-ai saappi-ṭṭ-aan
pro rice-ACC eat-PST-M.3SG
'*pro*$_i$ ate rice.' (i → singular male discourse antecedent)

(29) raman$_i$ [$_{CP}$ *pro*$_j$ sadatt-ai saappi-ṭṭ-aan]-nnu so-nn-aan
raman$_i$ *pro* rice-ACC eat-PST-M.3SG-COMP say-PST-M.3SG
'Raman$_i$ said that [$_{CP}$ *pro*$_{i/j}$ ate rice].' (j → singular male discourse antecedent)

While *pro*-drop is possible in the embedded finite subject position in (29), however, yielding both (accidental) coreference and disjoint readings, it seems to be impossible for the subject of a non-finite clause. The related Dravidian language, Malayalam shows the same distinction.[13] While we are not entirely sure of the explanation for this, we believe that it might have something to do with the lack of tense and agreement in non-finite clauses in Tamil[14] which hinders recovery of information if the subject is dropped (along the lines of standard analyses of Rizzi 1986, and others).

All that matters for the point made here, however, is that a reading with disjoint reference between embedded and matrix subjects is impossible in the non-finite structures (25)–(27) above. Under an assumption that the silent embedded subject is *pro* (which is standardly treated as a null pronoun), these facts could not be easily explained; this strongly suggests that the silent element is rather a controlled PRO.

3.5.3 Disambiguating argument identity and position

Tamil exhibits pervasive scrambling, thus surface order cannot always be taken to be representative of underlying structure. This fact combined with that of finite subject *pro*-drop can lead to ambiguity both with regards to the position and the identity of arguments. The adjunct non-finite sentence, repeated from (25) above, is thus ambiguous between two readings shown below:

(30) **Non-finite purposive clausal adjunct:**
raman pariccai-ai erud-a school-ukku poo-n-aan
raman.NOM exam-ACC write-INF school-DAT go-PST-M.3SG
Purposive/temporal reading 1: 'Raman went to school in order to write the exam.'/'Raman went to school as he wrote the exam.'
Purposive/temporal reading 2: '(Some guy) went to school (in order) for Raman to write the exam (purposive).'/'(Some guy) went to school as Raman wrote the exam.'

This interpretive ambiguity becomes especially obvious when we introduce an adverbial focus-operator *maṭṭum* ('only'/'alone') modifying the overt subject to (30) as shown below:

(31) raman **maṭṭum** pariccai-ai erud-a school-ukku poo-n-aan
Reading 1 ("High reading"): 'Raman alone went to school to write the exam'/'Raman$_i$ alone went to school as he$_i$ wrote the exam.'
Reading 2 ("Low reading"): '(Some guy) went to school (in order) for Raman alone to write the exam'/'(Some guy) went to school as Raman alone wrote the exam.'

Szabolcsi (2009) and Barbosa (2009) use very similar diagnostics to argue for a structural ambiguity in such constructions in flexible word-order and subject-drop languages like Italian and European Portuguese. Following their analyses, we propose that the above readings correspond to the following different structures:

[13] Rosmin Mathew (p.c.)

[14] Malayalam has no verbal agreement marking but it does have tense marking which also serves to distinguish finite and non-finite clauses.

(32) **Structure 1 ("High reading"): matrix overt subject, embedded controlled PRO:**
[$_{CP}$ raman$_i$ maṭṭum [$_{CP}$ PRO$_i$ pariccai-ai erud-a] school-ukku poonaan]

(33) **Structure 2 ("Low reading"): matrix *pro*, embedded overt subject:**
[$_{CP}$ *pro*$_i$ [$_{CP}$ raman$_{j/*i}$ maṭṭum pariccai-ai erud-a] school-ukku poonaan]

It is not so important for our current purposes that the matrix subject *pro*-drop structure corresponding to the "low reading" always be attested. What *is* important is that an obligatory control structure with embedded controlled PRO should always be possible in these structures — and it is. We thus conclude that the embedded clause types discussed in this section do indeed contain controlled PRO as indicated.

4 Evidence from other South Asian languages

Other South Asian languages show similar properties with respect to DP distribution as Tamil. Here, we present data taken from Malayalam (Mohanan 1982) and Sinhala (Gair 2005) and show that these are also similarly problematic for standard theories of Case.

4.1 Malayalam alternating complement infinitives

Malayalam also shows free variation between overt subjects and subject-controlled PRO in the non-finite complements of certain verbs like *aagɽahiccu* 'want', as discussed by Mohanan (1982). Consider the examples in (34):

(34) a. amma [PRO$_{i/*j}$ wisaːkk-aan] aagɽahiccu
mother.NOM PRO$_{i/*j}$ be.hungry-INF wanted
'Mother wanted to be hungry.'
b. amma [kuṭṭi-kkə wisaːkk-aan] aagɽahiccu
mother.NOM child-DAT be.hungry-INF wanted
'Mother wanted the child to be hungry.'

These Malayalam data, while similar enough to the English examples above, show one significant difference, namely that the embedded overt subject in (34b) is marked with a 'quirky' dative while the matrix subject is marked nominative. This is the dative normally assigned by the verb root *wisaː-* 'be hungry' to its subject, as can be seen from the following example:

(35) kuṭṭi-kkə wisaɳɳu
child-DAT was.hungry
'The child was hungry'.

This strongly suggests that the case on the embedded subjects in such clauses does not come from the matrix clause, but is established internal to the infinitive.

However, this in itself does not necessarily pose a problem for Case theory, as the presence of a dative subject does not entirely rule out an ECM-type analysis. One could still propose, along the lines of what has often been claimed for Icelandic quirky subjects (see e.g., Sigurðsson 1992, for discussion of the 'double-case' approach) that *kuṭṭi.kkə* first gets quirky dative case in the embedded clause, which determines its morphological form. On top of this, it then gets structural accusative Case by ECM, which is what licenses it to be overt, but this case does not surface in the morphology because the dative takes precedence. But this would predict that, if we have a *want*-class verb with an embedded infinitive and an overt subject but no quirky dative-assigning embedded verb, then the structural accusative should surface on the embedded subject. However, this prediction is not borne out, as the data in (36) show:

(36) amma [kuṭṭi ṭaḷaɽaan] aagɽahiccu
mother.NOM child.NOM be.tired.INF wanted
'Mother wanted the child to be tired.'

What we get on the embedded subject in (36) is nominative.[15] This is strong evidence against the idea that it should be licensed by ECM from the matrix verb.

Now we have a problem for Case theory, because it is not clear where else the licensing for these overt subjects could be coming from. On top of this, note that here again we have the alternation between PRO and overt DPs within a single environment. On the basis of such data, Mohanan (1982) came to the conclusion which we have also come to, namely that the Case-theoretic treatment of PRO could not handle the facts of Malayalam.[16]

4.2 Non-finite clauses in Sinhala

Gair (2005) presents data on overt and non-overt subjects in a wide array of clause types, both finite and non-finite, in Sinhala. These too are problematic for Case theory, because overt DPs show up in positions where they would not be expected to receive abstract Case. Also, as Gair himself points out, the subject-taking abilities of different clause types seem to correlate more with the properties of the embedding environment than with the finiteness of the clauses themselves.

We'll concentrate here on clauses with the verb form that Gair calls the infinitive, which bears no marking for tense.[17] When the infinitive occurs in the complement of certain verbs like *kəməti* 'like', we get obligatory control, i.e. overt subjects are ruled out, and there is obligatory coreference between the understood subject of the infinitive and the matrix subject:

(37) gunəpaalə$_i$ [*taman/*eyaa(mə)/$\emptyset_{i/*j}$ həmədaamə yannə] kəməti.
Gunapala.NOM$_i$ *self/*he/$\emptyset_{i/*j}$ every.day go.INF like.ASSM
'Gunapala likes to go there every day.'

This on its own looks like the sort of behavior we find in the complements of verbs like *try* in English. It seems that non-finite T in the embedded clause does not assign nominative case, and the matrix verb does not license accusative by ECM, so an overt subject is not allowed, so all is as predicted by Case theory.

Other classes of infinitives in Sinhala cast some doubt on this, however. For example, infinitives also appear in clauses introduced by elements like *issella* 'before'. As the examples in (38) show, overt subjects are possible in such clauses:

(38) a. [mamə ennə issella] miniha kaarekə wikka
I.NOM come.INF before man.NOM car sell.PST
'The fellow sold the car before I came.'
b. [maʈə teerennə issella] ləkcərekə iwərə unaa
I.DAT understand.INF before lecture finish become.PST
'The lecture ended before I understood (it).'

In a Case-theoretic approach, one might be tempted to see these as parallel to English *for-to* infinitives, with *issella* as a prepositional complementizer which assigns Case to the embedded subject. However, the morphology suggests that this is not the case. In (38a), the embedded subject is not marked with an oblique case — which is what one would expect on the complement of a preposition — but with the nominative associated with subjects of finite clauses. In (38b), the embedded subject

[15] The verb *ʈaɭar* in Malayalam shows up with a default nominative subject, as the following example shows:

(1) kuʈʈi ʈaɭarŋŋu
child.NOM tired
'The child was tired.'

[16] See also Jayaseelan (1984, 1985) for additional discussion of several types of non-finite clause in Malayalam, including some adjuncts which show alternating PRO and overt DP subjects, similar to those in Tamil.

[17] It also bears no agreement marking, but this is true of all verb forms in spoken Sinhala, thus does not really show much. As Gair (2005) discusses in detail, it is difficult to decide which Sinhala forms should count as finite and which should not, and it is questionable whether the distinction is a useful one for the analysis of the language. In any case, the 'infinitive' which we discuss here comes pretty close to a prototypical non-finite form, lacking tense specification and mostly being restricted to embedded environments, including those with obligatory control.

is marked dative, as is usual for subjects of *teerenəwa* 'understand'. In other words, the case on the embedded subject is fully determined internal to the infinitive clause and is independent of both the matrix clause and the embedded P or C element *issella*.

Sinhala also has a construction similar to what we saw with Tamil *veɳḍ-* 'want'. Infinitival clauses can be embedded under the verb *oonə* 'want', and can either have a covert subject coreferent with the matrix subject as in (39a), or an overt non-coreferent subject as in (39b):

(39) a. maʈə$_i$ [∅$_i$ heʈə koləmbə yannə] oonə
I.DAT ∅$_i$ tomorrow Colombo go.INF want
'I want to go to Colombo tomorrow.'

b. ammaʈə [laməya wibaage paas-wennə] oonə
mother.DAT child.NOM examination pass.INF want
'Mother wants the child to pass the examination.'

c. guruwərəyaʈə [laməyaʈə paaḍəmə teerennə] oonə
teacher.DAT child.DAT lesson understand-INF want
'The teacher wants the child to understand the lesson.'

Again, the subject of the embedded clause is in the nominative here. We might suppose that this comes from the finite matrix T, as the matrix subject is dative and thus does not use up the matrix nominative. In (39c), however, we again find the subject of an infinitive clearly getting dative case from the embedded verb. It thus again seems that the case needs of the non-finite subject are handled internal to the infinitival clause. This raises the possibility that the nominative in (39b), like that in (38a) above, is actually supplied within the infinitive as well, as Gair (2005) in fact concludes.

Finally, like Tamil, Sinhala allows infinitivals as purpose adjuncts, again at least optionally with overt subjects, as in (40):

(40) kaaʈəhari kaapu gamaŋ, [**kaurawat** nohoyanḍə] sarpəyə koheehari
someone bite.PST.REL when anyone.NOM NEG.find.INF serpent.NOM.PL somewhere
həngenəwa, nee?
hide.PRES no?
'When serpents have bitten someone, they hide so that no one can find them, don't they?'

The subject of the embedded infinitive, *kaurawat* 'anyone' is again in the nominative case. This cannot be coming from the matrix finite clause for two reasons. First, since the infinitive is an adjunct, it is not in the right structural configuration for ECM-like assignment of a matrix case to its subject position. Second, the nominative of matrix T is already assigned to the matrix subject *sarpəyə* 'serpents'. There is also no hint of a prepositional complementizer heading the embedded clause like English *for* or even the *issella* we saw in (38) above. It is thus a mystery under standard Case theory how an overt subject should be licensed in examples like this, just as it was with the parallel Tamil adjunct infinitives. Note that, however licensing works, the source of the nominative morphological case on *kaurawat* 'anyone' must be internal to the infinitival clause, again since the structural configuration is not right for something to come from the matrix clause. This supports the suspicion noted above that the embedded nominative case in (39b) is supplied internal to the embedded clause, further undercutting support for an ECM-like analysis of complements of *oonə* 'want'.

The conclusion that Gair (2005) draws from these Sinhala data is very similar to what we have said for Tamil. He points out that the occurrence of non-overt subjects in infinitives embedded under *kəməti* cannot be attributed to the non-finite verbal inflection, precisely because other infinitives do allow overt subjects. Instead, he argues, whether we get overt or non-overt subjects in an infinitive depends on the matrix clause and the relationship between it and the embedded clause. This is the line we ourselves will pursue in Section 7 where we formalize a syntax and semantics for DP distribution in terms of selectional properties of the matrix predicate.

5 Problematic data from non-South-Asian languages

Problematic data for standard Case theory are of course not only found in South Asian languages. In this section, we describe the AcI construction attested in a number of languages, Middle English infinitive clauses which cast doubt on Case-theoretic analyses of the *for-to* construction in Present-Day English and, finally, a problematic structure in Present-Day English itself, namely the gerundival. These data show that DP distribution presents challenges for Case-theoretic analyses quite generally across languages.[18]

5.1 AcI constructions in Irish and Latin

A number of languages including Latin, Ancient Greek and Modern Irish have a type of non-finite construction usually called AcI or accusative-with-infinitive. They involve an overt subject DP in the accusative case, which makes them superficially look like English ECM infinitives. However, they have a radically different distribution from the relevant English clauses, which are restricted to the complement of verbs like *believe*. AcI clauses instead appear in a wide range of syntactic environments: as clausal subjects, as arguments of matrix nouns and adjectives, as adjuncts and sometimes even in special root contexts. We give as examples here from Irish a subject clause in (41a) and a clausal argument of a noun in (41b) (from McCloskey 1985), and from Latin a clausal argument of an adjective in (42a) and a root infinitive in (42b) (from Gildersleeve 1895).[19]

(41) **Modern Irish**

a. Ghoillfeadh se orm [**tu** me a ionsai].
would.bother it on.me you.ACC me INF attack
'It would bother me for you to attack me.'

b. Níl iontas [**é** mac mí-nádúrtha a thógáil].
is-not wonder him.ACC son un-natural INF raise
'It is no wonder that he should raise an unnatural son.'

(42) **Latin**

a. Est inūsitāt-um [**rēg-em** re-um capit-is esse].
is extraordinary-N.sg king-ACC answerable-ACC head-GEN be.inf
'It is an extraordinary thing for a king to be tried for his life. (C., *Dei.*, 1. 1)

b. **Homin-em**-ne Rōmānum tam Graecē loqu-ī?
man-ACC-Q Roman-ACC.M.sg such Greek-ABL speak-INF
'A Roman speak such good Greek? (To think that a Roman should speak such good Greek!)' (PLIN., *Ep.*, IV. 3, 5)

Clearly, in the examples above and most of the contexts where AcI infinitives show up, there is no c-commanding transitive verb. So ECM is simply ruled out as a possible source of licensing for the overt subject, morphological accusative aside.[20] There is also no evidence for anything like English *for*. Thus, as in the Tamil adjunct infinitive examples, it is not clear how the overt subjects could be Case-licensed. Note furthermore that in both languages, PRO is found in infinitival clauses in these positions as well, so here again we find the PRO/overt subject alternation.

5.2 Middle English infinitives

Similarly problematic data for Case theory have been reported by McFadden (2008) for Middle English (ME). E.g., throughout the Middle English period, the vast majority of *for-to* infinitives do not have an overt subject:

[18] The reader is also referred to Szabolcsi (2009) for an extensive discussion and case-studies of nominative-marked subjects in non-finite control and raising constructions crosslinguistically.

[19] See also vanden Wyngaerd (1994) for examples from Ancient Greek and further discussion.

[20] See Pillinger (1980), Cann (1983) for additional arguments against a general ECM analysis of Latin AcI.

(43) a. I ne come not in-til erþe [**for to** do mi wille]
'I didn't come to the earth to do my will.' (BENRUL,10.333)
b. . . . and wente wythout the wal [**for to** walke]
'. . . and went outside the wall in order to walk.' (REYNAR,11.179)

The few that do place the subject **before** *for*, not between it and *to*:

(44) For it es a velany, [**a man for to** be curyously arrayed apon his heuede with perré and precyous stanes]
'For it is a disgrace for a man to be strangely adorned on his head with jewels and precious stones.' (ROLLTR,29.609, from Pak 2006)

So we do not see the familiar correlation between overt subjects and *for* in infinitive clauses that is familiar from Present-Day English. This casts some suspicion on a Case-based analysis of Present-Day English *for-to* infinitives, since *for* does not seem to be playing a role in licensing overt subjects.[21] Furthermore, this constitutes another instance of the PRO/overt subject alternation.

The really interesting data from Middle English, however, are the infinitives which are not complements of transitive verbs, have no *for*, and yet still have an overt subject (see Fischer 1988, Garrett in press, Pak 2006, McFadden 2008, for discussion). Much like the AcI infinitives just discussed, they occur as surface clausal subjects, extraposed subjects of predicate adjectives and nouns, as well as adjuncts:[22]

(45) 'That were shame unto the,' seyde sir Launcelot, '[**thou** an armed knyght to sle a nakyd man by treson].'
"'That would be a disgrace on you," said sir Lancelot, "for you.NOM, an armed knight, to slay a naked man by treason".' (MALORY,206.3373)

As (45) shows, the subject of the infinitive is often in the nominative form when it is a pronoun. Other examples have non-nominative forms:[23]

(46) ȝiff itt like to thy most gracious lordshipp [**me** to do þis message]
'If it would please your most gracious lordship for me.OBL to do this message. . . '
(ROYAL,258.322)

Again, it is unclear how the overt subjects are licensed in these clauses. The parallel Present-Day English examples have an overt *for* and are assumed to be grammatical precisely because *for* is there to handle Case.

5.3 Gerundival clauses in Present-Day English

Tamil gerundivals are not the only ones that are problematic for Case theory. Present-Day English gerundival clauses appear in various positions with no c-commanding transitive verb, have nothing like overt *for*, and no tense or agreement marking. Yet they also allow overt subjects, with no apparent source for Case, and they show the alternation with PRO that we have seen in several clause types:

[21] Henry (1992) proposes a Case-theoretic account of similar facts in Belfast English. However, in that dialect, overt subjects can come before *for* only in the complements of *want*-class verbs, when the embedded subject is adjacent to the matrix verb. Otherwise *for* comes before the subject. This is what makes the ECM analysis that Henry proposes for SUBJ-*for-to* examples plausible. Such an analysis will not work for a language like ME, where this order is also found in sentences like (44) in which the infinitival clause is not the complement of the matrix verb, and where the order *for*-SUBJ-*to* is simply not found.

[22] Clauses of this kind are mostly restricted to the second half of the ME period, roughly 1350–1500, after which they are replaced by the modern *for-to* infinitive. During this period they are, however, found often enough in the texts that we can be sure they are not errors. See Fischer (1988) for a large collection of examples, and McFadden (2008) for corpus-based statistical evidence that this kind of infinitive occurs at a very similar frequency in two of Chaucer's *Canterbury Tales* as the *for-to* infinitive occurs in Early Modern English.

[23] Fischer (1988) suggests that the non-nominative forms are older here, with the nominative an innovation. In any case the non-nominatives here cannot be construed as evidence for an ECM analysis of such clauses, since here as well c-commanding transitive verbs are lacking.

(47) [Barry/PRO$_{i/*j}$ having no hot sauce], we$_i$ went to the store. [adjunct]

These are problematic for the same reason that the Tamil ones are. It is not obvious how the overt subjects could be getting Case, and if we come up with a reason why Case should be available here, we lose our explanation for the versions with PRO. Although a Case theoretic analysis of *either* the structure with PRO or that with the overt subject might be possible, a unified analysis of both types of structures is not viable.[24]

5.4 Quick excursus: finite control and case-marked PRO

Even though this is not the main focus of this paper, we would like to point out for the sake of exhaustiveness that the converse scenario of obligatorily controlled PRO in finite clauses is also crosslinguistically attested. This is just as problematic for standard Case theories as overt subjects in non-finite clauses are — since Case theory postulates a strictly one-one mapping relationship between controlled PRO and the null Case or lack of Case assigned by non-finite T. This in turn entails that obligatory control PRO only occurs in [Spec, TP$_{-fin}$].

Landau (2004) and Darzi (2008) present examples of finite control in Hebrew and Persian, respectively. Here, we reproduce the Persian examples from Darzi (2008) below (formatting ours):

(48) **Finite obligatory control PRO with overt complementizer**
Žian mi.tun.e (ke) be.r.e
Jian DUR.be.able.3.SG (that) SUBJ.go.3SG
'Jian can/is able to go.'

Darzi (2008) presents convincing evidence that the null subject in (48) is indeed controlled PRO and not little *pro* and also shows that the embedded clause is not a restructuring infinitive (Wurmbrand 2001) but a full-fledged finite CP. The grammatical appearance of obligatory-controlled PRO in a finite clause as in (48) above is thus entirely unexpected given the null Case Minimalist hypothesis and also argues strongly that subject distribution should be divorced from Case-licensing and, in turn, from clausal finiteness.

Additionally, the Icelandic floating-quantifier structure in (49) below (due to Sigurðsson 2008), shows convincingly that obligatory control PRO can bear case, gender, and number features even in its canonical non-finite subject position — which, again, is a serious blow to licensing accounts of obligatory control PRO in terms of Case:

(49) **Obligatory control PRO with case, number, and gender: Icelandic**
Bræðrunum likaði illa [að PRO vera ekki báðir kosnir]
brothersD.M.PL liked ill [to PRO be not both.N.M.PL elected]
'The brothers$_i$ disliked [$_{CP}$ PRO$_i$ not being both elected].'

5.5 Interim summary

To summarize, then, the apparent free variation between overt subject and obligatory subject-controlled PRO in various embedded clause types in several languages is seriously problematic for standard Case theory. Any proposal that might account for one type of alternant (PRO or the overt DP) appears incompatible with the other alternant, so that a unified analysis of both variants taken together seems unfeasible. Furthermore, in many of the relevant clause types, the fact that overt subjects are possible at all challenges basic assumptions about what can license such DPs. It is also difficult to explain, within a Case-theoretic analysis, the differing behavior of adjunct non-finite clauses and complement non-finite clauses in Tamil, namely why it is that complement infinitives come in both obligatory control and alternating varieties, while all adjunct clauses are alternating.

[24] See Pires (2007) for recent discussion of the properties of gerundival clauses and the problems they pose for standard Case theory.

6 Can Case theory be salvaged?

Before turning to an alternative account for the data above, it is perhaps both prudent and logical to exhaust the possibilities for explaining them within the assumptions of standard Case theory. In this spirit, we consider three options that might make sense within Case theory but show that each of these options is untenable. Along the way, we give critical discussion of a previous attempt by Sarma (1999) to account for some of the Tamil data in Case theory.

6.1 Option 1: Case licensing from the matrix clause

The first idea is that licensing for (some) overt non-finite subjects comes not form the matrix verb, but from matrix T. Sarma (1999) proposes such an analysis for *want*-class complements with quirky-dative matrix subjects, like the following:

(50) **Alternating non-finite complement with** PRO**/overt subject:**
champa-vukku$_i$ [PRO$_i$/sudha oru samosa-vai saappiḍ-a] veṇḍ-um
Champa-DAT PRO/Sudha a samosa-ACC eat-INF want-N.3SG
'Champa wants to eat a samosa'/'Champa wants Sudha to eat a samosa.'

The idea is that, since the matrix subject gets dative, matrix T still has a nominative Case left over, which it can assign into the embedded clause, licensing the overt subject there.

But this proposal has the following problems. First, it's only viable for infinitives in the complement of verbs with a dative subject; if the matrix subject were nominative, it (and not the embedded subject) would be assigned nominative Case by matrix T. Second, while this idea works nicely for each alternant individually, it fails to consistently explain the data when taken together: i.e. why *both* PRO and overt DPs are possible in (apparent) free variation. Thus, short of motivating distinct structures for the complements with overt subjects and those with PRO, a consistent Case-based analysis along these lines is unviable.

6.2 Option 2: special licensing due to a null P

A second proposal would be that it is the presence of a null P head in the non-finite clause — a null version of English *for* — that licenses Case on the overt subject. Such a proposal has been put forward by Bošković (1997) and Martin (2001) for English *want*-class complements, for instance. However, such a proposal also has serious flaws. The main problem is that there is no independent evidence for such a P head (overt or covert) in the constructions we have discussed; in fact, as we have noted, the nominative case-marking on the non-finite subjects strongly argues against such an idea, since Ps typically and in these languages do not assign nominative Case. And finally, the free variation between PRO and the overt DP finds no insightful explanation within this approach either, and a potential way out — positing the presence of an optional null P — is both inelegant and stipulative.

6.3 Option 3: Case-licensing by non-finite T

The third possibility would be that the non-finite T in non-finite complement and adjunct clauses with overt subjects itself possesses exceptional Case licensing properties. Sarma (1999) tentatively suggests that something along these lines might be the explanation for the alternating purposive/temporal non-finite adjuncts in Tamil discussed in this paper, repeated again below:

(51) [PRO/sriram saadatt-ai saappiḍ-a], naan veḷiya poo-n-een
PRO/Sriram rice-ACC eat-INF I.NOM outside go-PST-1SG
'I went out to eat rice.'/ 'I went out for Sriram to eat rice.' PURPOSIVE INTERPRETATION
As I ate rice, I went out/As Sriram ate rice, I went out.' TEMPORAL INTERPRETATION

This approach would certainly account for the presence of overt DP within non-finite clauses, but is problematic for the following reasons. First, it is essentially a stipulation with no independent evidence. Second, an exceptional Case-licensing non-finite T would, like in the other approaches, not

be able to account for the alternation with subject-controlled PRO in the same positions. Finally, this approach raises more theoretical problems than it seems to solve: if both non-finite and finite Ts can license overt DPs, how do we predict when we get PRO vs. overt DPs in a Case-theoretic approach? The possibility of distinguishing PRO vs. overt DPs in terms of differing licensing conditions due to the finiteness of T — one of the fundamental claims of Case theory — is lost.

For the discussion above, we conclude that an account for subject DP distribution in terms of licensing due to Case is unfeasible and propose below an alternative analysis based on selection rather than licensing of subjects.

7 Analysis

In this section we will present an alternative analysis of DP distribution which can handle the data from Tamil and other languages which are problematic for Case theory. We will first describe and motivate the basic idea, which is based on the selectional needs of lexical and functional heads rather than on the licensing needs of the DPs themselves. Then we will lay out one possible formal implementation of the idea within the Minimalist framework in terms of Agree and the $[\pm R]$ feature introduced by Reinhart and Reuland (1993) and later in Landau (2004)[25] and finally we will demonstrate it with sample structures for the most important clause types.

7.1 Introducing the selection alternative

Recall that the idea behind Case theory is that DPs are inherently defective and require external licensing. Under this view, (overt) DPs are ruled out by default, and what must be explained are the places where they are licit. The approach we would like to pursue is essentially the opposite of this. We propose that DPs have no special needs, but simply have to be integrated into the structure and interpretation like any other syntactic elements. Thus they can show up anywhere this is possible — as long as independently motivated principles of grammar are satisfied. Under this view, DPs are fine by default in all A-positions, and what we must explain are the places where they are illicit.

The facts about overt DP distribution that have been handled in terms of Case thus need a different kind of explanation. Following Marantz (1991) and others, it is now standardly assumed that one set of these facts — having to do with required movement operations to derived subject positions in passives, unaccusatives, raising and other contexts — should be handled in terms of the EPP or something similar (see e.g. Chomsky 2001). The other main set of facts has to do with what we have been concerned with here — the choice between overt DPs and PRO. What we would like to propose is that instances where overt DPs are impossible — as well as those where they are required — should be explained in terms of selection by c-commanding functional and lexical heads.

The basic idea is that, by default, both overt DPs and PRO are licit in all A-positions (as long as the EPP is satisfied), and are in fact in alternation with each other. This is of course precisely the situation of alternation that we find in a number of the clause types we have discussed here, including the Tamil alternating complement infinitives, adjunct infinitives and tensed gerundivals, Latin and Irish AcI infinitives, English gerundivals, and ME adjunct infinitivals. In such clauses, the choice of DP type is determined by the intended interpretation. That is, if coreference with a matrix argument in the syntax is intended, then controlled PRO appears.[26] If not, then an overt DP/little *pro* is used. In contexts where overt DPs are impossible, this is because something is explicitly selecting PRO. This is what we get e.g. in the complements of *paar-* in Tamil, of *try* in English and of *kəməti* in Sinhala. The ungrammaticality of overt subjects here has nothing to do with a lack of licensing for them — they need no licensing. The problem is rather that the selectional requirements imposed in the clause are not satisfied, causing the derivation to crash. In contexts where PRO is typically impossible, this is because something is explicitly selecting an overt DP/*pro* (we'll discuss

[25] However, it will become clear from the discussion in the rest of the paper that we deviate in significant points from both these proposals in our conception and formal implementation of this feature.

[26] We leave aside arbitrary PRO for the time being, which has not appeared in the data we've examined in this paper. We believe that it can be unified satisfactorily with obligatory control PRO, but the discussion would take us too far afield.

below how to unify these). We see this in all of the languages discussed in prototypical main finite clauses. Again, the problem with PRO is not that it itself is not licensed, but that it does not satisfy the selectional requirements of the relevant local functional head.

A good way to understand the difference between the licensing-based approach of Case theory and our selection-based approach is in terms of the probes and goals of Minimalist Agree relations. Under both approaches, the distribution of DPs in particular syntactic positions is regulated by relationships between those DPs — the goals — and the verbs and functional heads that c-command them — the probes. In Case theory, it is the needs of the goal — relating to the Case features on the DP — which must be met for a particular configuration to be allowed. In our selection-based theory, on the other hand, it is the needs of potential probes — relating to the selectional features on verbs and functional heads — which must be satisfied.

7.2 On finiteness and dependency

Implicit within our analysis is a rather different take on finiteness and its relation to subject positions than that embedded within Case theory. Crosslinguistically there is a clear tendency for prototypical finite clauses to have overt subjects and for prototypical infinitives to have PRO subjects. We of course do not intend to deny this tendency, but we do take issue with the rather simplistic way that it has been understood in theoretical terms. Standard Case theory relies on the idea that the finiteness of a clause, or more accurately of T/Infl, determines and explains the choice between overt DPs and PRO. It is not just that overt subjects and finiteness go together, but that finiteness is what licenses overt subjects. This presupposes and relies on a basic binary distinction between finite and non-finite. More fine-grained distinctions between clause types may be made for other purposes,[27] but in terms relevant for Case assignment by T/Infl, there are only these two possibilities.

We reject this understanding of the connection between finiteness and DP distribution. First, we question whether a merely binary notion of finiteness can be defended and maintained crosslinguistically.[28] There are rather several distinct ingredients which go into what is traditionally thought of as finiteness, including at least agreement, tense marking, temporal interpretation, referential dependence and modality. These can combine in different ways in different languages and in different constructions within a single language. In principle, several different binary distinctions could be drawn, but it is not clear which distinction should be privileged for purposes of DP distribution, especially if we strive for crosslinguistic consistency (see Adger 2007, and the other papers in Nikolaeva 2007 for discussion). In fact, the evidence we've presented in this article shows that, for the purposes of subject distribution, at least three types of embedded clause must be distinguished: those which require PRO, those which require overt subjects, and those which allow either. This undermines the claim that finiteness is responsible for licensing certain types of DP. The existence of the alternating type of clause suggests that the non-alternating types actually involve restrictions on the generally available possibilities for the subject position, not the creation (i.e. licensing) of possibilities.

Thus in what follows, we will speak in terms of a notion of clausal *dependency* rather than finiteness.[29] This is intended to make clear the distinction between our position and that of standard Case theory, and we think it more accurately reflects the facts. What is meant here is not a single

[27] E.g. the complements of verbs like *believe* are usually assumed to be smaller than the complements of verbs like *try*. As a consequence, the former are transparent for Case coming from outside (i.e. ECM) and for A-movement (i.e. subject-to-subject raising) while the latter are not. Nonetheless, both types of clause count as non-finite in the sense that T itself does not assign nominative Case and license an overt subject.

[28] See Landau (2004) for some related discussion and arguments that finiteness, at least as is relevant for the distribution of PRO, needs to be broken down into more basic components. Also see Gair (2005) for arguments that a simple notion of finiteness is not particularly useful for an understanding of the properties of embedded clauses in Sinhala, including the data we discussed in Section 4.2 above.

[29] Again, see Landau (2004) for much related discussion and conclusions that are quite similar to those we draw here. Landau proposes a rather explicit formal calculus for determining whether an embedded clause will have PRO or an overt subject, which is independent of Case. Important differences from our analysis are that Landau's system posits an important role for agreement and provides no obvious way to deal with clauses showing the PRO/overt DP alternation.

distinction of dependent versus independent, but a range of possibilities determined by the interaction of multiple factors. Dependency in our sense is meant to encode the extent to which (at least) the temporal and nominal referents of an embedded clause are defined in terms of or in relation to those of the clause in which it is embedded. It is of course clear that clausal embedding can set up relationships between elements of the matrix and embedded clauses on both nominal and temporal levels. E.g. the reference of pronouns (including PRO) in an embedded clause is determined in part on the basis of the reference of nominals in the matrix. And the temporal reference of an embedded clause is systematically determined relative to that of the matrix. Furthermore, it is reasonable to think that these two different types of dependencies are related and treated in parallel fashion by the grammar. Consider in this connection the tradition of semantic work documenting the parallels between pronouns and tenses, which are based in large part on the similar dependency effects that they display in embedded clauses (see Partee 1973, Kratzer 1998).

From this perspective, obligatory control PRO (as well as other anaphoric DPs such as reflexives and bound variable pronouns) just represent the maximum degree of dependence on the matrix clause in the area of nominal reference. In temporal reference this would correspond to embedded clauses with anaphoric tense which cannot refer to a time distinct from that of the matrix, as in (52), where the matrix and embedded clauses have incompatible temporal adverbs:

(52) * **Yesterday** Andrew tried *Aloisius/PRO to work **tomorrow**.

Similarly, protoypical finite root clauses in out-of-the-blue contexts represent the maximum degree of independence. In the area of nominal reference, an overt subject (or little *pro*) is required whose reference can be established solely on the basis of the discourse, independent of anything in the syntactic context. In the area of temporal reference, the tense information necessary to compute the reference time from the speech time must be specified in such clauses, as there is no matrix time specification to work from.[30]

How many degrees of dependency need to be distinguished is an open question.[31] As noted above, for the purposes of DP distribution, it is clear that at least three are required, and we will thus speak here in terms of a three-way distinction. Clauses which require overt DP subjects we will call **independent**, those which require PRO we will call **anaphoric**, and those which allow the alternation we will call **dependent**.[32] In syntactic terms, independent clauses will contain a functional head which selects for overt DPs, anaphoric clauses will contain a head which selects for PRO, and dependent clauses will contain a head that selects for neither. A more nuanced typology will probably be needed for other purposes (and perhaps even for DP distribution when more data are taken into account), but we will stick to the bare minimum as far as it will take us.[33] Crucially, the two degrees distinguished in traditional Case theory are not sufficient.

Note then that our notion of dependency is defined in essentially semantic terms relating to temporal and nominal reference. We expect that it will be related to the morphological properties of a particular embedded clause, but only indirectly. This is another clear distinction from standard finiteness, which is commonly defined at least in part in terms of morphological tense and agreement marking. For us, what really matters is not the presence or absence of overt tense marking, but the relative dependency of the embedded temporal reference. The embedded clause in (53) below has

[30] Whether the nominal-temporal parallelism holds in the general case, in particular at the intermediate points on the scales, is an important question that can shed light on how exactly the dependencies are related. However, it is far less clear, and we will not take a position on it here.

[31] It should, however, be a finite number. The dependencies under discussion here are syntactic and must be implemented in terms of discrete features and structures, so we should not be dealing with a true continuum or cline.

[32] Compare Landau (2004)'s similar use of anaphoric and dependent tense.

[33] Note that our three-way distinction in clause types maps onto a two-way distinction in DP types: PRO versus overt DPs. One clause type requires PRO, a second type requires overt DPs, and the third allows either. We will argue below that other DP types pattern either with PRO or with overt DPs in terms of their syntactic behavior relative to dependency. However, it will almost certainly turn out to be necessary to make a finer distinction in the degree of dependency of DP types.

some degree of temporal independence, unlike that in (52) above, yet morphologically they are both "infinitives", lacking any tense morphology.

(53) **Yesterday** Andrew$_i$ wanted Aloisius/PRO$_i$ to work **tomorrow**.

Similarly, we care about what referential types of embedded subjects are allowed, not so much about the availability of agreement morphology. Here again, the contrast between (52) and (53) shows that lack of agreement can be found in both anaphoric and dependent clauses.

We also find the dissociation in the opposite direction — finite morphological marking in dependent situations. The "finite" control phenomena discussed in Section 5.4 are examples of nominal reference being highly dependent even though there is morphological agreement. And the well known sequence of tense effect found in many languages represents an instance where temporal interpretation and morphological tense marking go their separate ways.

What regulates the distribution of the various types of clauses in the syntax, then? That is, how is the dependency between matrix and embedded clause actually computed? We propose that this is a function of the type of embedding. This includes the distinction between adjunction and complementation (and probably conjunction), and also crucially the relationship between matrix predicate and embedded clause under complementation. Specific verbs and other predicates differ in what kinds of clausal complements they can select. In English, e.g. *want* takes independent *that*-clauses and dependent *to*-infinitives, *remember* takes independent *that*-clauses, independent gerundivals and anaphoric *to*-infinitives, while *try* takes anaphoric gerundivals and anaphoric *to*-infinitives, and these selectional restrictions must be stated somehow in the grammar. The distribution of subject DP types is then largely a question of what kind of clauses a particular predicate selects. So we always get PRO and not overt DPs in *try*-class complements because verbs like *try* select anaphoric clausal complements, which in turn contain a functional head that selects PRO.

7.3 Selectional semantics

Of course, we would ultimately like to clarify **why** it is that English *try* and Tamil *paar* 'try' take anaphoric complements, while English *want* and Tamil *veɳɖ* 'want' take dependent ones. That is, what independent factors determine the selectional properties of specific heads? Can we derive the selectional restrictions (at least in part) from something deeper, or do they just have to be lexically stipulated? It is reasonable to think that selectional behavior ultimately has a semantic basis — an idea that is reinforced by the observation that predicates meaning roughly 'try' across languages tend to take anaphoric complements, while verbs meaning 'want' tend to be more flexible. To be more precise, the meaning 'try' seems to imply the involvement of its subject in the embedded proposition, whereas 'want' does not. I.e. while 'want' expresses a relation between an individual and a proposition, 'try' expresses a relation between an individual and a property which that individual will saturate.[34] We can think of this in terms of the following rough semantic denotations:[35]

(54) a. $[\![try]\!]_{<<e,t>,<e,t>>} = \lambda P_{<e,t>} \lambda x_e .\textsc{try}(x, P(x))$
b. $[\![want]\!]_{<t,<e,t>>} = \lambda Q_t \lambda x_e .\textsc{want}(x, Q)$

The denotations above show that both predicates take two arguments, namely the matrix subject and the embedded clause. Crucially, the denotation for 'try' states furthermore that the highest argument

[34] What we are saying here about 'try' is clearly related to the proposal by Chierchia (1989) and others that control infinitives are properties rather than propositions. However Chierchia (1989), argues that control should be treated as a purely semantic phenomenon and proposes that the control complement, despite being semantically a property of type <e,t>, nevertheless has an entailed subject; this entailed subject is one of the arguments of the matrix clause giving obligatory coreference effects between the "antecedent" asserted DP and the embedded entailed one. For us it is crucial, however, that selection be syntactically represented - though the selectional semantics of the predicate as well as binding operations later at LF are also key players.

[35] Note that we do not claim that these are the actual precise denotations for English *try* and *want*. They are simply intended to show the abstract form that those denontations take with respect to complementation and coreference requirements between matrix and embedded arguments.

of the embedded clause will be coreferent with the matrix subject. In contrast, the denotation of 'want' places no such restrictions on the arguments of its complement clause.

Extensive arguments in favor of a semantic theory of control are put forward by Culicover and Jackendoff (2006). The approach we outline here shares important intuitions with their proposal, specifically the idea that the selectional behavior of matrix verbs with respect to control has a semantic basis. However, while they argue for a primarily semantic treatment of control phenomena, we implement the crucial selectional relationships syntactically. That is, the semantics of a predicate may help to determine what sorts of structures it can appear in, but it is these syntactic structures that determine the distribution of PRO and overt DPs.[36]

7.4 Comparing approaches: DP selection vs. licensing

The key advantage of our approach over more traditional Case-theoretic ones is that it can elegantly handle the alternation between PRO and overt subjects within a single syntactic context. As discussed above, such alternations are seriously problematic for Case theory, which attempts to account for the distribution of overt DPs and PRO in terms of distinct licensing conditions. The logically possible Case-theoretic responses to such alternations are ad hoc at best, and have essentially no predictive power. Because the possibility for an alternation has to be essentially stipulated, such a theory makes no predictions about where alternations should and should not be expected.

On the other hand, a selection-based approach to the choice between PRO and overt DPs predicts the existence of alternations between the two and allows an explanation of where they are and are not found. Because the two types of DPs do not have distinct special licensing needs, we expect that they will be in alternation with one another whenever neither is explicitly selected; this, in fact, is the default scenario. Our empirical investigations reported above have shown that clause types with alternating subject types are not only attested, but rather common crosslinguistically. Thus the fact that our approach can handle them so much better than Case-theoretic approaches can is a decided advantage. Furthermore, because selection is a syntactic relationship, we can expect it to be subject to standard structural and locality restrictions. So, for instance, matrix verbs will only be able to have an affect on the DP distribution in clauses which are their direct complements. Adjunct clauses and clauses which contain enough structure to place a locality boundary between their subject and the matrix verb thus cannot have a selectional effect on the type of the embedded subject. This is why Tamil has both obligatory control and alternating complement infinitives, depending on the class of the matrix verb, but only alternating adjunct infinitives.[37] In addition, the minimality of selection will also ensure that matrix verbs and clausal functional heads can affect the type of the subject, but not the object. Selection will always see the highest DP, with subjects being interveners for potential selection of object DP types. This is why we have nothing like control of embedded objects by matrix arguments.

Note in comparison that both theories are equally well equipped to accommodate clauses which require only one or the other type of DP. Standard Case theory has to assume a formal distinction on overt DPs and PRO — either the overt DPs bear Case features and PRO does not, or PRO only bears a special null Case feature — and a series of formal distinctions on c-commanding verbal

[36] On standard conceptions of the interaction between syntax and semantics which we accept here, the semantics of a predicate cannot directly determine its selectional (or other syntax) within a derivation. The narrow syntax feeds into the LF branch, not the other way around. The way that we intend for semantic considerations to affect the syntax is more indirect. Roughly speaking, if you have a syntactic structure with a predicate that selects an anaphoric clausal complement, the result will be the kinds of meanings where a matrix argument is necessarily involved in the embedded predicate. On a conceptual level, 'try' has such a meaning whereas 'want' does not. Perhaps a more accurate way to state the tendency is that 'try' is one of the meanings consistent with structures where an anaphoric dependent class is selected. And a predicate that appeared in a different structure would have to mean something slightly different.

[37] This of course does not mean that all adjunct clauses should allow alternations. Adjunct clauses can also be anaphoric, e.g. English gerundivals headed by *while* are anaphoric:

(i) Paul watches TV while *Dan/PRO working.

Crucially, though, the matrix verb has nothing to say about the fact that such adjuncts are anaphoric, while gerundivals headed by *with* are dependent.

and functional heads, e.g. transitive and ECM verbs assign structural accusative, finite T assigns structural nominative, non-finite T assigns no Case or null Case, and ECM and raising verbs like *believe* and *seem* select TP complements, while obligatory control and alternating verbs like *try* and *want* select different flavors of non-finite CP. We also must assume a comparable amount of formal information on both DPs and the heads in their environment. Overt DPs and *pro* must be distinguished formally from PRO, which, we will argue, can be done in terms of a single binary feature. We must also posit a series of distinctions on c-commanding heads, e.g. we will need similar selectional features on verbal heads like *believe*, *try* and *want* for different categories of embedded clauses. We will not need to assume Case features on verbs or functional heads, but we will need to posit selectional features on some of them for either PRO or overt DPs/*pro*. Thus in terms of the number and complexity of formal devices needed to account for the standard DP distribution facts, our account fares no worse than Case theory.[38] The fact that it can also handle the less well-known alternation facts discussed here, is then decisive.

7.5 Formal implementation

We have already said that there are degrees of dependency in both the nominal and temporal domains. In the nominal domain, this dependency is for reference: on the one end, we have complete referential independence and on the other, we have complete referential dependence or anaphoricity. This corresponds to "R-expressions", such as proper names, on the one end and anaphors such as controlled PRO and SELF-anaphors (Reinhart and Reuland 1993) on the other, respectively. In this section we will explore one way in which a theory of DP distribution based on selection for dependency types can be formally implemented. Given the focus of this article, we restrict ourselves to dealing with the semantics of nominal reference.[39] What we describe here is the subject of ongoing research, but it should be sufficient to give an idea of how such a theory can work.

7.5.1 The semantics of nominal reference

To be able to talk intelligently about how referentiality might be formalized, we need to ask ourselves the following question: what makes a DP referentially anaphoric, as in the case of controlled PRO and SELF-anaphors, and what makes it referentially independent, as in the case of proper names? The answer, we believe, lies in the semantics of the DP itself: specifically, following by-now standard analyses of reference and anaphora (in the tradition of Heim and Kratzer 1998, and many others), we propose that an anaphoric DP has a variable index on the D head. This essentially means that it does not denote/point to any particular individual in the real world. In contrast, a (maximally) referentially independent DP such as a proper name (e.g. *Martha*) does not have a variable but a fixed reference index which denotes a particular individual in the world. Additionally, we believe that both types of DPs inherently contain (interpretable) ϕ-features. For an anaphoric DP these introduce a presuppositional semantics which constrains the domain of possible antecedents for this DP.[40] Following the analysis of Heim (2008) and others for deictic pronouns, we propose that these

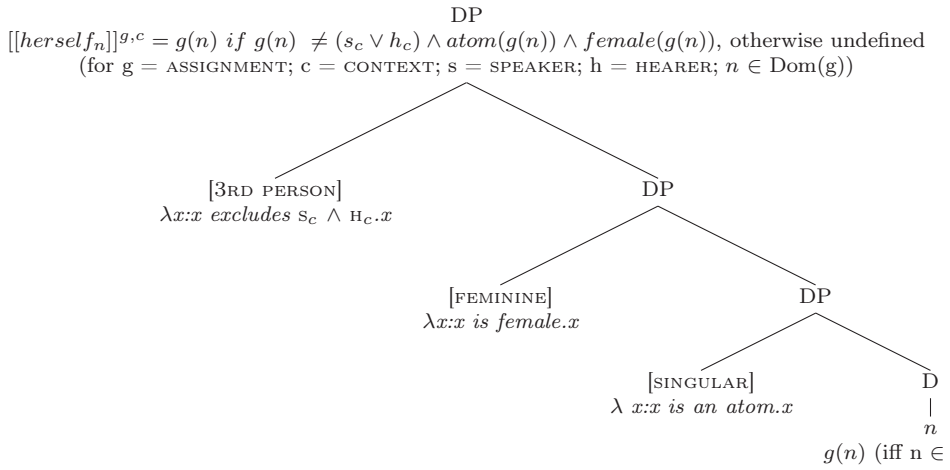
[38] In fact, to the extent that we succeed in motivating the features we use to distinguish PRO from overt DPs/*pro* in our implementation below, our account may be seen as less stipulative. Given recent demonstrations that morphological case is dissociated from DP distribution (see e.g. McFadden 2004, Sigurðsson 2009), abstract Case features would at present serve only to model the DP distribution facts, and thus would be more stipulative. Additionally, the null Case treatment of PRO due to Martin (2001, and others) has also come under wide criticism for being theoretically stipulative (see Hornstein 1999, Sigurðsson 2008, and many others.).

[39] We suspect that temporal semantics is related, as discussed above, but will not attempt a formal treatment of this for now. It is very likely that analyses are on the right track where clausal temporal and aspectual heads mediate between anaphoric DPs and their antecedents in the syntaxt/semantics, as argued in Heim (2008), Kratzer (2009), Borer (1989) and others. We ourselves propose such a treatment via selection due to Agree in the narrow syntax but leave the details with respect to temporal semantics aside for the time being.

[40] This stands contra analyses by Kratzer (1998), Heim (2008), Kratzer (2009) and others, which propose that bound variable DPs are additionally also born without any ϕ-features. This proposal definitely has its attractions: for instance, in the syntax, their (putatively) minimal feature-structure immediately distinguishes these pronominals from non-anaphoric (deictic) pronouns and flags them for variable-binding at LF. Additionally, LF/PF featural mismatches — where an indexical pronoun gets pronounced with 1st/2nd person features but is interpreted as a non-indexical

ϕ-features are syntactically adjoined to the D head in an arbitrary (or probably parametrized) hierarchical order. Thus, under our treatment, an anaphoric DP like *herself* has the following structure and corresponding LF semantics (cf. Kratzer 2009, 188 for the denotation of pronouns).

(55) **Syntax and semantics for an anaphoric DP (here: *herself*)**



Since the anaphoric DP does not have a reference of its own, it must somehow inherit this reference from another DP. Again following standard semantic analyses (citations above), we propose that this happens through variable binding of the anaphoric DP at LF. Variable binding ensures that the anaphoric DP ends up getting assigned the same index as that of its syntactic antecedent — yielding obligatory coreference readings such as those exhibited in the control data described earlier in this paper. Crucially also, semantic binding (as defined within the tradition of Heim and Kratzer 1998) does not require locality between the antecedent and anaphoric DPs — only that a c-command relationship (built on the simpler notion of sisterhood) exist between them. This in turn means that controlled PRO can be bound by its obligatorily non-local antecedent. For a referentially independent DP, on the other hand, like the proper name *Martha*, the story is much simpler: it has its own reference and does not need to acquire this from any other DP, in the syntax or in the discourse. Standard algorithms of denotation and assignment take care of this.

7.5.2 The syntax of reference

But this is not the whole story. If we take the Y-modular architecture of the grammar seriously, where LF and PF follow syntax, and if we further want to maintain the standard idea that binding of anaphors must happen late, at LF, we need a separate account for what happens in the syntax. I.e. what is the "narrow" syntactic correlate of reference? We propose (following the terminology in Reinhart and Reuland 1993, Landau 2004) a binary feature $[\pm R]$ that is both inherent and interpretable on DPs. Since $[\pm R]$ is a purely syntactic feature, it does not care about discourse dependencies on DPs. Thus, not only maximally independent DPs, like proper-names, but also deictic pronouns (and *pro*) — whose reference is determined in the discourse-context and not in the syntactic one — are marked $[+R]$.[41] In contrast, anaphoric DPs with a variable reference index

bound variable — are avoided by assuming that the pronounced features only get inherited post-syntactically on the PF branch. However, we for now reject this featureless account of anaphoric DPs for the following reasons. First, the morphophonological identity of bound-variable and deictic pronouns is reduced to accident, since both DPs are held to have an inherently different syntax and semantics. Second, it is unclear how an anaphoric DP which is locally c-commanded by another anaphoric DP (as e.g. *himself* is by controlled PRO in: *John tried* $[_{CP}$ PRO$_i$ *to slap himself*$_i]$) ever gets to inherit its ϕ-features, since this would be strictly counter-cyclic. Finally, we also believe that, while a ϕ-feature might constrain reference, it is not equivalent to it, thus the two must be kept separate. See Sundaresan (2009) for more detailed arguments on this topic.

[41] Here, we follow Reinhart and Reuland (1993) who also argue that deictic pronouns must be marked $[+R]$. But we differ from them in claiming that controlled PRO is marked $[-R]$.

such as SELF-anaphors, controlled PRO and classic bound-variable pronominals are marked with a feature $[-R]$. Crucially, the binary feature $[\pm R]$ does not itself indicate semantic referentiality or the lack thereof: reference is only determined at LF — instead, $[\pm R]$ serves to merely flag a DP in a certain manner, for later operations at LF and PF.[42] More precisely, we can think of $[\pm R]$ as an instruction to LF to deal with the corresponding DP in a certain manner: if the DP is marked $[-R]$, it should be variable-bound following independent rules of semantic binding due to antecedent-raising and predicate-abstraction; if the DP is marked $[+R]$, on the other hand, variable-binding does not obtain at LF. The $[\pm R]$ feature also serves as an instruction to PF: this plays out as Spell-Out rules conditioning phonological insertion which, among other things, determine whether a DP gets spelled out overtly or covertly. Specifically, a DP with the feature $[-R]$ in the syntax will be spelled out overtly as a SELF-anaphor/bound-variable pronoun or silently, as controlled PRO; a $[+R]$ DP, on the other hand, will be spelled out overtly as an R-expression or deictic pronoun or covertly as little *pro*. See Sundaresan (2009) for an analysis of the independent factors conditioning the choice between PRO and overt anaphors and discussion of how this is regulated by the systematic, cyclically-ordered interaction of the various grammatical modules.

But this is not all $[\pm R]$ does; crucially, being a feature that is both interpretable and inherent on DPs, it can also participate in purely syntactic dependencies. In particular, we can use it to encode the selectional relationships involved in the distribution of DP types laid out in Section 7.2 above. Following proposals in Adger (2007) and elsewhere, we formalize selection through Agree, the standard Minimalist operation that captures dependencies in the syntax. Simply put, the selecting head — the probe — bears an uninterpretable feature, which must Agree with a matching interpretable feature on an appropriate goal, obeying standard locality and minimality restrictions. While $[\pm R]$ is interpretable on DPs, it should be uninterpretable on functional heads, so we can use it there to encode the selectional restrictions that these heads carry (we use the notation $[uF]$ to indicate that a feature F is uninterpretable). Specifically, a head that is specified $[u + R]$ will require a $[+R]$ DP (i.e. an overt R-expression, deictic pronoun, or little *pro*) in its local c-command domain, while a head specified $[u - R]$ will require a $[-R]$ DP (i.e. controlled PRO, bound variable pronoun or SELF-anaphor) in its c-command domain.

The distribution of the clausal types discussed in Section 7.2 above is regulated by a binary feature we'll call $[\pm anaph]$. Specifically, an anaphoric clause will be specified $[+anaph]$, an independent clause $[-anaph]$, and a dependent clause underspecified for this feature. In each case, the feature will be on C, where it is plausible to think that it is interpretable. However, on a predicate that selects a clausal complement, such a feature would be uninterpretable (indicated as: $[u\pm anaph]$). We must also stipulate a dependency between $[\pm R]$ and $[\pm anaph]$. In particular, a C that is $[+anaph]$ will necessarily require a $[-R]$ subject and will thus itself additionally be marked $[u - R]$. On the other hand, a C that is marked $[-anaph]$ will require a referentially independent subject (marked $[+R]$), and will thus also be marked $[u + R]$.

That is:[43]

[42] A very similar point is, in fact, made by Reinhart and Reuland (1993) who first introduced the syntactic labels of $[\pm R]$ on DPs. They claim [p. 697]:

> "It is not the case that referential properties of NPs miraculously restrict their syntactic behavior; rather, some independent properties of NPs determine how they can be used to refer. Thus, R should be a purely syntactic property. Having this property is a necessary condition for an expression to function as an independent argument, but R itself does not have anything to do with reference."

This more or less conforms to our own view with regards to the binary feature $[\pm R]$ though, as mentioned earlier, for independent reasons we do not follow Reinhart and Reuland (1993)'s later conclusion that $[\pm R]$ refers to the ϕ-featural and Case specifications on the DP and nothing else.

[43] If we allow a three-way distinction in terms of the binary feature $[\pm anaph]$ (plus, minus and underspecified/not present) even on heads where it is interpretable, we should expect the same logical possibility with $[\pm R]$. That is, there should also be DPs which are simpy not specified for this feature. Such DPs would presumably be intermediate in their degree of potential referential dependency. In the current system, they would be syntactically restricted to positions where neither $[+R]$ nor $[-R]$ was being selected for, since they would not be able to Agree with any selecting head for this feature. As we have pointed out above, we expect that more degrees of relative dependence will need to be distinguished once more facts are taken into account, so this may be a positive result. One option to pursue

(56) a. $[+anaph] \rightarrow [u - R]$
b. $[-anaph] \rightarrow [u + R]$

Given the discussion in Section 7.2, temporal dependency should be related to the $[\pm anaph]$ feature as well. Of course, ultimately we would like to derive this relationship and the stipulation in (56) in principled fashion, but this will require a deeper understanding of the dependencies involved and must be left for future work.

7.6 Sample structures

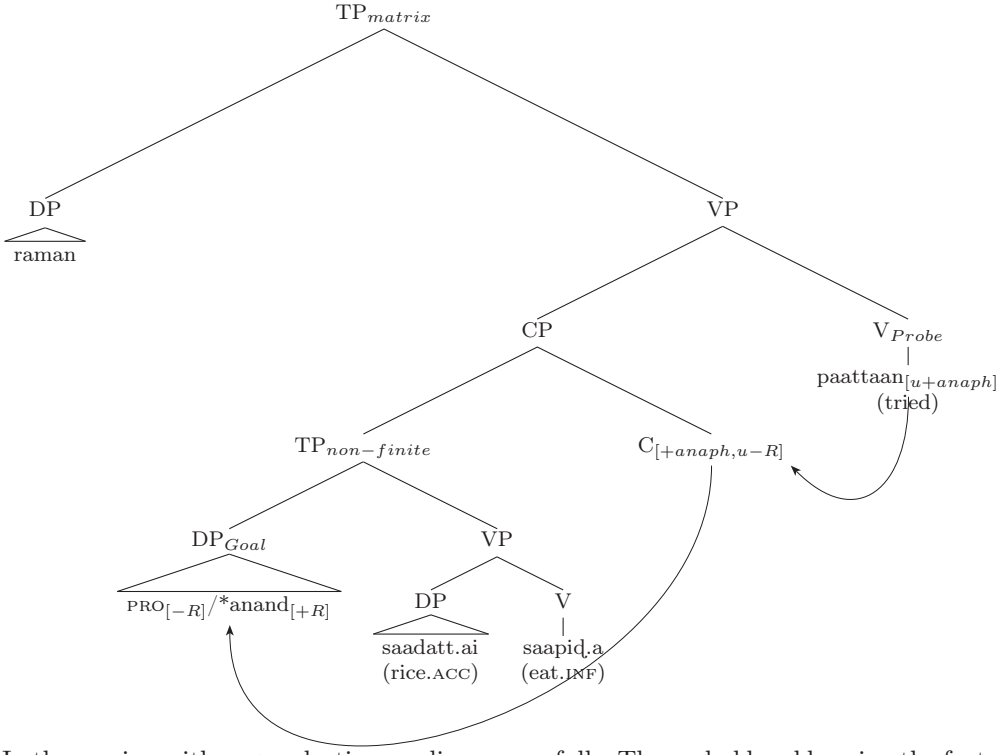
The Tamil and other data presented in Sections 3–5 can now be acounted for as follows.

7.6.1 Anaphoric clauses

With predicates that take obligatory control complement infinitives, like Tamil *paar-* 'try' and English *try*, the selection proceeds in a two-step process. The matrix predicate selects directly for a temporally and referentially anaphoric clausal complement. This is expressed by the feature $[+anaph]$ which is uninterpretable on the selecting predicate and interpretable on C. This C, by virtue of being anaphoric, in turn selects directly for a $[-R]$ subject (PRO).

The tree below shows a simplified version of this kind of selection. Note that arrows on all trees do not represent movement, but Agree relationships.

(57) raman [PRO/*anand saadatt-ai saappiḍ-a] paa-tt-aan
raman.NOM PRO/*anand.NOM rice-ACC eat-INF try-PST-3m.sg
'Raman$_i$ tried [$_{CP}$ PRO$_i$/*Anand to eat rice].



In the version with PRO, selection applies successfully. The verbal head bearing the feature $[u+anaph]$ c-commands C bearing $[+anaph, u - R]$, which in turn c-commands PRO bearing $[-R]$. In neither

would be treating at least some long distance/SE-anaphors in this fashion. However, it should be stressed that our central concern here is showing that a syntactic account of the alternation betwen PRO and other DP types in terms of selection is possible. What the precise details of that analysis should be, remains a matter of ongoing research. The specific featural machinery proposed here has been chosen primarily on this basis of simplicity, and will presumably need to be revised as our understanding improves.

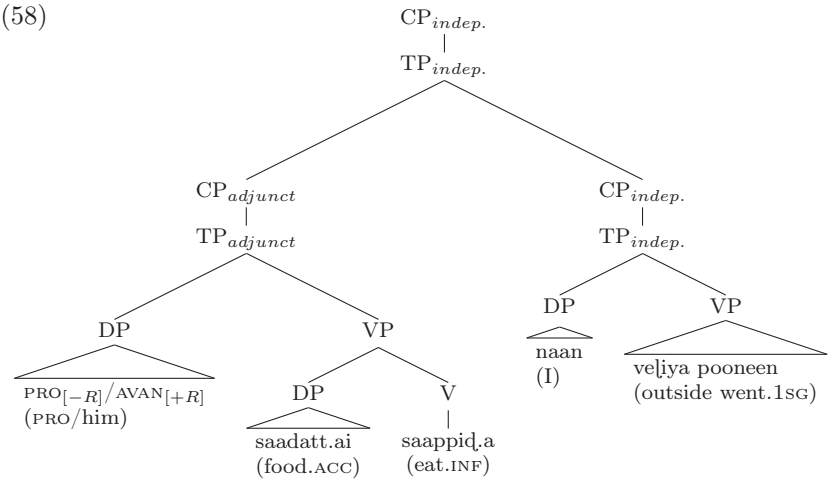
instance does a (phasal) locality boundary intervene, nor are there potential interveners which would create minimality problems. I.e. all of the syntactic requirements for Agree are fulfilled, and the uninterpretable features can be checked off by matching with their interpretable counterparts.

In the version with an overt subject (*Anand*), however, there is crucially a mismatch of $[\pm R]$ features between probe and goal. The C is marked with an uninterpretable $[u - R]$ feature, which means that it needs to Agree with a DP that is marked $[-R]$ — in other words, a PRO (or an overt anaphor). However, the closest goal, the embedded subject *Anand*, as an overt R-expression, is marked $[+R]$. As such, it is not able to Agree with the probe and the derivation crashes.

7.6.2 Dependent clauses

The next structure, corresponding to example (13a), shows the default scenario which is one where both PRO and overt subject DPs alternate in free variation. This is the situation in alternating non-finite complements such as those of *want*-class predicates in Tamil, English, Malayalam and other languages, as well as in Tamil and ME adjunct infinitives.

(58)



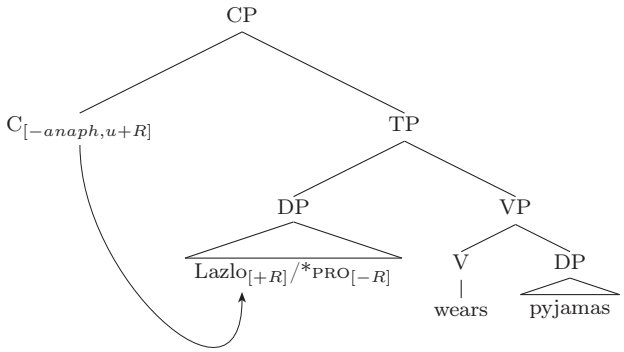
Here, the matrix predicate (*pooneen* 'went') is not inherently specified for $[u \pm R]$. Additionally, the structural conditions necessary for Agree between Probe and Goal simply do not obtain since the non-finite clause is an adjunct, not a complement of the matrix predicate. In complements of *want*-class predicates which allow both overt subjects and PRO, the structural conditions for Agree between a matrix functional head and the embedded DP subject do obtain but selection nevertheless does not occur because the matrix predicate simply does not bear a selectional feature for $[\pm R]$.

7.6.3 Independent clauses

Independent clauses show the opposite kind of selection from *try*-class infinitives. Such clauses are temporally and referentially independent and thus the C head is marked $[-anaph]$. As such, it is also specified $[u + R]$, and hence requires a referentially independent subject.[44] The locality and minimality conditions are of course the same as we saw for selection of $[-R]$ by anaphoric C in the obligatory control complements above.

[44] We assume that all independent clauses are CPs. Alternatively, if we say that (some) independent clauses are TPs, $[u + R]$ would be located in finite T.

(59)



In contrast, PRO subjects in finite clauses are predicted to be impossible because PRO, being inherently $[-R]$, cannot satisfy the selection requirements of independent C in the matrix clause. Here again, we have a mismatch between the $[u+R]$ feature of the selecting C and the $[-R]$ feature of the closest potential goal PRO.

8 Summary

We have presented extensive data from Tamil, Malayalam, Sinhala, Latin, Irish, Middle English and Present-Day English and shown that they are highly problematic to standard Case-theoretic approaches to the distribution of overt DPs and PRO. For one thing, we find that overt DPs are licensed in a number of these languages in contexts where Case theory predicts that only PRO should be licensed. For another, we find overt DPs and PRO in alternation in a number of contexts, with no independently observable factors differentiating the variants with the two types of DP. We have argued that this state of affairs is inconsistent with the Case-theoretic premise that overt DPs and PRO have distinct and complementary licensing requirements.

We have considered a number of options for handling the problematic data in such a way as to salvage Case theory but concluded, however, that none of these possibilities yields a satisfactory analysis of the range of data considered. In particular, none of the Case-theoretic strategies can accommodate the alternations between PRO and overt DPs without unattractive stipulations.

We have thus proposed an alternative analysis, which is in a sense the reverse of the standard Case-theoretic approach. Where Case theory accounts for the distribution of DPs in terms of their own need for licensing, we have argued that it is the selectional requirements of c-commanding verbs and lexical heads that are responsible. Specifically, in certain contexts — like the complement of the Tamil verb *paar-* 'try' — there is selection for a DP bearing the interpretable feature $[-R]$ like PRO. If a $[+R]$-bearing overt DP occurs in the relevant position instead, the result is ungrammatical — not because the DP is not licensed, but because the selectional requirements are not met. A similar selectional requirement for $[+R]$ on C in independent clauses ensures that they will have overt DP or little *pro* subjects rather than PRO. In many clauses, however, no such selectional pressures are at work — the embedding verb is not specified for a particular clause type, or the clause is embedded in such a way that selection is ruled out for syntactic reasons. In such clauses, the default scenario emerges, which is free variation between PRO and overt DPs controlled by the intended interpretation.

Our proposal provides a consistent account of the data that were problematic for Case theory. Not only can it accommodate the attested alternations between PRO and overt DPs, it also makes testable (and thus far confirmed) predictions about where such alternations should and should not be found. Thus it achieves a level of explanation where Case theory allowed only stipulation. Crucially, it can still handle the data that were not problematic for Case theory. Finally, while we have motivated our selection-based approach on the basis of data from Tamil and a handful of other languages, we do not intend it as an account of those languages alone. Rather, we believe that it can serve as the basis for a more general theory of DP distribution, and intend to expand it to a broader selection of languages in future research.

Acknowledgments

We are grateful to R. Amritavalli, Jim Gair, and K.A. Jayaseelan for discussion as well as to Artemis Alexiadou, Idan Landau, and two anonymous reviewers for detailed written comments on earlier versions of this paper. We also thank audiences at the SALA Roundtable in Madison, EFLU in Hyderabad, Generative Grammatik des Südens (GGS) in Leipzig, the Syntax Workshop of the Roots conference in Stuttgart, and the research seminar on syntax/morphology in Stuttgart — where variants of this paper were presented — for helpful feedback. Thanks also to Rosmin Mathew for Malayalam judgments. All errors, of course, are due to us alone.

References

Abusch, Dorit. 1997. Sequence of tense and temporal de re. *Linguistics and Philosophy* 20(1):1–50.

Adger, David. 2007. Three domains of finiteness: a minimalist perspective. In *Finiteness: theoretical and empirical foundations*, pages 23–58. Oxford: Oxford University Press.

Amritavalli, R. and K.A. Jayaseelan. 2005. Finiteness and negation in Dravidian. In G. Cinque and R. Kayne, eds., *The Oxford Handbook of Comparative Syntax*, pages 178–220. Oxford: Oxford University Press.

Barbosa, Pilar P. 2009. A Case for an Agree-based theory of control. Lingbuzz: http://lingbuzz.auf.net.

Borer, Hagit. 1989. Anaphoric AGR. In O. Jaeggli and K. Safir, eds., *The Null Subject Parameter*, pages 69–109. Dordrecht: Kluwer Academic Publishers.

Bošković, Željko. 1997. *The Syntax of Nonfinite Complementation*. Cambridge, MA: The MIT Press.

Cann, Ronald. 1983. Raising in Latin: a phrase structure analysis of the accusative and infinitive. In G. Gazdar, E. Klein, and G. Pullum, eds., *Order, Concord and Constituency*, pages 113–137. Dordrecht: Foris.

Chierchia, Gennaro. 1989. Structured meanings, thematic roles and control. In G. Chierchia, B. Partee, and R. Turner, eds., *Properties, Types and Meaning*, vol. II: Semantic Issues of *Studies in Linguistics and Philosophy*, pages 131–166. Dordrecht: Kluwer Academic Publishers.

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

Chomsky, Noam. 2001. Derivation by phase. In M. Kenstowicz, ed., *Ken Hale: A Life in Language*. Cambridge, MA: The MIT Press.

Chomsky, Noam and Howard Lasnik. 1993. The theory of principles and parameters. In J. Jacobs et al., eds., *Syntax: An international handbook of comtemporary research*. Berlin: Walter de Gruyter.

Culicover, Peter and Ray Jackendoff. 2006. Turn over control to the semantics! *Syntax* 9:131–152.

Darzi, Ali. 2008. On the vP analysis of Persian finite control constructions. *Linguistic Inquiry* 39(1):103–116.

Enç, Mürvet. 1987. Anchoring conditions for tense. *Linguistic Inquiry* 18(4):633–657.

Fischer, Olga. 1988. The rise of the *for NP to V* construction: An explanation. In G. Nixon and J. Honey, eds., *An Historic Tongue: Studies in English Linguistics in Memory of Barbara Strang*, pages 67–88. London: Routledge.

Gair, James W. 2005. Some aspects of finiteness and control in Sinhala. In R. Singh and T. Bhattacharya, eds., *The Yearbook of South Asian Languages and Linguistics*, pages 117–143. Berlin and New York: Mouton De Gruyter.

Garrett, Andrew. in press. The historical syntax problem: Reanalysis and directionality. In D. Jonas, J. Whitman, and A. Garrett, eds., *Proceedings of the Eighth Diachronic Generative Syntax Conference*.

Gildersleeve, Basil L. 1895. *Gildersleeve's Latin Grammar*. London: MacMillan, 3rd edn.

Heim, Irene. 2008. Features on bound pronouns. In D. Harbour, D. Adger, and S. Béjar, eds., *Phi Theory: phi-features across modules and interfaces*, Oxford Studies in Theoretical Linguistics, pages 35–56. Oxford: Oxford University Press.

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

Henry, Alison. 1992. Infinitives in a *for-to* dialect. *Natural Language and Linguistic Theory* 10:279–301.

Higginbotham, James. 1980. Pronouns and bound variables. *Linguistic Inquiry* 11:679–708.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30:69–96.

Jaeggli, Osvaldo and Ken Safir. 1989. Introduction. In O. Jaeggli and K. Safir, eds., *The Null Subject parameter*. Dordrecht: Kluwer Academic Publishers.

Jayaseelan, K.A. 1984. Control in some sentential adjuncts of Malayalam. In *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*.

Jayaseelan, K.A. 1985. Infinitivals in Malayalam: Some preliminary observations. *CIEFL Working Papers in Linguistics* 2.

Kratzer, Angelika. 1998. More structural analogies between pronouns and tense. In *Proceedings of SALT VII*.

Kratzer, Angelika. 2009. Making a pronoun: fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40(2):187–237.

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

Landau, Idan. 2006. Severing the distribution of PRO from Case. *Syntax* 9:32–66.

Marantz, Alec. 1991. Case and licensing. In *ESCOL '91: Proceedings of the Eighth Eastern States Conference on Linguistics*, pages 234–253.

Martin, Roger. 2001. Null case and the distribution of PRO. *Linguistic Inquiry* 32:141–166.

McCloskey, James. 1985. Case, movement and raising in Modern Irish. In J. Goldberg, S. MacKaye, and M. Wescoat, eds., *Proceedings of the Fourth West Coast Conference on Formal Linguistics (WCCFL)*, vol. 4, pages 190–205. Stanford Linguistics Association.

McFadden, Thomas. 2004. *The position of morphological case in the derivation: a study on the syntax-morphology interface*. Ph.D. thesis, University of Pennsylvania.

McFadden, Thomas. 2008. Overt subjects of infinitives and *for-to* in the history of English. Talk presented at the 10th Diachronic Generative Syntax Conference (DiGS X), Cornell.

Mohanan, K.P. 1982. Infinitival subjects, government and abstract Case. *Linguistic Inquiry* 13(2):323–327.

Nikolaeva, Irina, ed. 2007. *Finiteness: Theoretical and Empirical Foundations*. Oxford: Oxford University Press.

Pak, Marjorie. 2006. Infinitive marking with *for*: a diachronic account. In *Proceedings of the 29th Annual Penn Linguistics Colloquium*, pages 293–306.

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

Pillinger, O.S. 1980. The accusative and infinitive in Latin: a refractory complement clause. *Journal of Linguistics* 16:55–83.

Pires, Acriso. 2007. The derivation of clausal gerunds. *Syntax* 10:165–203.

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

Rizzi, Luigi. 1986. Null objects in italian and the theory of pro. *Linguistic Inquiry* 17:501–557.

Sarma, Vaijayanthi M. 1999. *Case, Agreement, and Word Order: Issues in the Syntax and Acquisition of Tamil*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge.

Sigurðsson, Halldór Ármann. 1991. Icelandic Case-marked PRO and the licensing of lexical arguments. *Natural Language and Linguistic Theory* 9:327–363.

Sigurðsson, Halldór Ármann. 1992. The case of quirky subjects. *Working Papers in Scandanavian Syntax* 49:1–26.

Sigurðsson, Halldór Ármann. 2008. The case of PRO. *Natural Language and Linguistic Theory* 26:403–450.

Sigurðsson, Halldór Ármann. 2009. The No Case generalization. In A. Alexiadou, J. Hankamer, T. McFadden, J. Nuger, and F. Schäfer, eds., *Advances in Comparative Germanic Syntax*, pages 249–280. Amsterdam: John Benjamins.

Sundaresan, Sandhya. 2009. A unified account of control and local anaphora. Talk presented at the South Asian Language Analysis Roundtable (SALA) XVIII, University of North Texas.

Szabolcsi, Anna. 2009. Hidden in plain sight: overt subjects in infinitival control and raising complements. In *Papers in syntax*, vol. 2. NYU Working Papers in linguistics.

vanden Wyngaerd, Guido. 1994. *PRO-legomena: Distribution and Reference of Infinitival Subjects*. New York: Mouton de Gruyter.

Wurmbrand, Susanne. 2001. *Infinitives: restructuring and clause structure*. Berlin: Mouton de Gruyter.

Yip, Moira, Joan Maling, and Ray Jackendoff. 1987. Case in tiers. *Language* 63:217–250.

Zaenen, Annie, Joan Maling, and Höskuldur Thráinsson. 1985. Case and grammatical functions: The Icelandic passive. *Natural Language and Linguistic Theory* 3:441–83.

Adjunct Control in Telugu: Exceptions as Non-Exceptions

Youssef A. Haddad, *University of Florida*

Received December 2008; Revised December 2009

Abstract

South Asian languages license control into adjuncts known as conjunctive participle clauses. At the same time, these languages allow exceptions to adjunct control. These exceptions have received very few, mainly semantic, analyses in the literature. This paper focuses on one South Asian language, Telugu, and offers a syntactic analysis. It shows that the so-called exceptions to adjunct control are non-exceptions and that they are instances of Expletive Control that involve two unaccusative predicates. The proposal is not without challenges. One challenge comes from English that does not allow Expletive Control. The article spells out the English details and shows that they do not create a problem for the Telugu data.

1 Introduction

Conjunctive participle (CNP) clauses are non-finite adverbial clauses. They are a crosslinguistic category that exists in many languages, such as Modern Greek, Korean, and Diyari (Haspelmath 1995). In the Indian Subcontinent, they are a defining characteristic that South Asian languages inherited from Sanskrit (Dwarikesh 1971, Masica 2005).

Structures with CNP clauses in South Asian languages obey what is known as the Same-Subject Condition (Klaiman 1981, 88) or the Common-Subject Requirement (Lindholm 1975, 30). That is, the subject of the CNP clause and the subject of the matrix clause are obligatorily coreferential, and a sentence with a CNP clause is an instance of Obligatory Control. Sentences (1)–(4) are examples from selected South Asian languages.[1] The symbol Δ stands for the unpronounced subject. The English translations show that the CNP clauses, although subordinate, have a conjunctive meaning; thus, the name conjunctive participle.[2]

(1) madhuu=ne$_i$ [$\Delta_{i/*k}$ pustaka utsl-uun] kapaaṭ=aat ṭhewlii
Madhu=Erg [books pick up-CNP] cupboard=Loc put
'Having picked up the books, Madhu put them in the cupboard.'
'Madhu picked up the books and put them in the cupboard.' Marathi
(from Pandharipande 1997: 106, ex.(266))

[1] Abbreviations used are: 3=3rd person, Acc=accusative, Ag=agent, CNP=conjunctive participle, Dat=dative, Erg=ergative, Gen=genitive, Loc=locative, M=masculine, N=neuter, Nom=nominative, Pl=plural, Sg=singular.

[2] For the purpose of economy, the examples in the rest of the article will include either the subordinate or the coordinate translation.

JSAL Volume 2, Issue 1, December 2009.

(2) avan$_i$ [$\Delta_{i/*k}$ keṇattule viZund-u] nõaamaana nelele keḍakkaran
he [well fall-CNP] bad state lie
'Having fallen in a well, he is in a critical condition.'
'He fell in a well and is in a critical condition.' Tamil
(from Lindholm 1975: 65, ex.(3.32))

(3) [$\Delta_{i/*k}$ ghar aa-kar] raaj=ne$_i$ kapṛe badle
[home come-CNP] Raj=Ag clothes changed
'Having come home, Raj changed.' Hindi
'Raj came home and changed.'
(from Kachru 1981: 36, ex.(3))

(4) [$\Delta_{i/*k}$ phal per-e] judo$_i$ bikri korlo
[fruit pick-CNP] Jodu sale did
'Having picked the fruit, Jodu sold it.' Bengali
'Jodu picked the fruit and sold it'
(from Klaiman 1981: 108, ex.(4.49))

Although the phenomenon of Obligatory Control into CNP clauses is robust, exceptions do exist, an observation that was first made by Lindholm (1975). For example, whereas the references of the CNP and matrix subjects in (1)–(4) above have to coincide, sentences (5)–(8) provide counterexamples in which disjoint subjects are allowed.

(5) [paauus paḍ-uun] dhaanya pikla
[rain fall-CNP] crops grew
The rain fell, and the crop grew.' Marathi
(from Pandharipande 1997: 446, ex.(1277))

(6) [maze penj-u] aatu=le taṇṇi ooduccu
[rain fall-CNP] river=Loc water ran
'Having rained, the water flowed in the river.' Tamil
(from Lindholm 1975: 81, ex.(3.38))

(7) [diwaar gir-kar] patthar gir gaee
[wall fall-CNP] stones fell went
'The wall having fallen, stones fell.' Hindi
from Davison 1981: 122, fn.5, ex.(i))

(8) [ceaar bhẽge giy-e] modhu pore gaelo
[chair break down-CNP] Modhu fell down
'The chair broke, and Modhu fell off.' Bengali
(from Klaiman 1981 114, ex.(4.57e))

Telugu is a Dravidian language of South Asia. Like the aforementioned languages, Telugu licenses control into CNP clauses, (9a–c).

(9) a. [$\Delta_{i/*k}$ juttu pooy-i] Kumaar$_i$ picciwaaḍu ayyaa-ḍu
[hair lose-CNP] Kumar.Nom a crazy man became-3.M.Sg
'Having lost his hair, Kumar went crazy.' Telugu

b. kumaar$_i$ [$\Delta_{i/*k}$ laybrarii=ki weḷḷ-i] pustakam cadiwaa-ḍu
Kumar.Nom [library=Dat go-CNP] book read-3.M.
'Kumar went to the library and read a book.' Telugu

c. kumaar [$\Delta_{i/*k}$ daggu=u jalubu=u wacc-i] mandulu waaḍaa-ḍu
Kumar.Nom [cough=and cold=and come-CNP] medicines used-3.M.Sg
'Having caught a cough and a cold, Kumar took medication.' Telugu

At the same, exceptions to adjunct control are attested, as (10a)–(10i) illustrate.[3] In these struc-

[3]I thank an anonymous *JSAL* reviewer for examples (10b–c).

tures, the Same-Subject Condition is violated by the lack of coreference between the CNP and matrix subjects.

(10) a. [tufaanu wacc-i] naa=illu kuulin-di
[flood come-CNP] my=house.Nom collapsed-3.N.Sg
'The flood came, and my house collapsed.' Telugu

b. [simla-loo mancu paḍ-i] ḍhillii-loo calla paḍin-di
[Simla-Loc snow fall-CNP] Delhi-Loc cool became.3.N.Sg
'The snow fell in Simla, and it became cool in Delhi.' Telugu

c. [bayaṭa baagaa calla-paḍ-i] inṭi-loo callagaa undi
[outside well cold-fall-CNP] house-Loc cool is
'Having become cool outside, it is cool in the house.' Telugu

d. [warʃam paḍ-i] ceṭlu/mokkalu perigaa-yi
[rain.Nom fall-CNP] trees/plants.Nom grew-3.N.Pl
'The rain fell, and the trees/plants grew.' Telugu

e. [cali-kaalam wacc-i] aakulu raalipooyaa-yi
[cold-weather.Nom come-CNP] leaves fell-3.N.Pl
'Winter arrived/came, and the leaves fell down.' Telugu

f. [eṇḍa-kaalam wacc-i] ceruwulu eṇḍipooyaa-yi
[hot-weather.Nom come-CNP] tanks dried up-3.N.Pl
'Summer came, and the water tanks dried up.' Telugu

g. [kurcii wirig-i] kumaar kinda paḍḍaa-ḍu
[chair.Nom break-CNP] Kumar.Nom down fell-3.M.Sg
'The chair broke, and Kumar fell down.' Telugu

h. [bhuukampam wacc-i] caala kaṭṭaḍaalu kuulipooyaa-yi
[earthquake.Nom came-CNP] many buildings.Nom collapsed-3.N.Pl
'An earthquake came, and many buildings collapsed.' Telugu

i. [baambu peel-i] caala mandi canipooyaa-ru
[bomb.Nom explode-CNP] many people.Nom died-3.M.Pl
'A bomb exploded, and many people died.' Telugu

These exceptions are not uncommon in South Asian languages and, thus, warrant an explanation.[4] In this paper, I limit the discussion to Telugu. I try to show that the so-called exceptions to Telugu adjunct control are non-exceptions. In other words, they too obey the Same-Subject Condition.

The following sections are organized as follows. Section 2 surveys the literature for available, mainly semantic, analyses and shows that they do not adequately account for the phenomenon in question. Section 3 presents an alternative syntactic account, analyzing the structures in question as instances of Expletive Control. Section 4 presents data from English that pose a challenge to the syntactic account; the data demonstrate that Expletive Control is banned in English. Section 5 shows that the lack of Expletive Control in English does not undermine the analysis of the Telugu exceptions as Expletive Control structures. Section 6 revisits the English data and explains why Expletive Control does not exist in English. Section 7 is a summary and a conclusion.

2 Existing Analyses

Researchers tend to analyze exceptions to adjunct control from a purely semantic perspective without any reference to syntax. For example, in her book on Marathi, Pandharipande (1997, 445–446) briefly indicates that such structures are allowed when there is a cause-effect relationship between the CNP

[4] My observation is that these exceptions tend to be more common in one language than in another. For example, they seem to be less common in Hindi than they are in Tamil. This observation, however, is not based on any statistics, and it calls for further investigation.

clause and the matrix clause. In this case, "the agents of the matrix and the participial [CNP] clauses can be different."

Unfortunately, at least in Telugu, a cause-effect relation between the CNP and matrix clauses is not sufficient for disjoint subjects to be allowed, as (11) illustrates. Compare to (9c) above.

(11) *[kumaar=ki daggu=u jalubu=u wacc-i] sarita mandulu iccin-di
[Kumar=Dat cough=and cold=and come-CNP] Sarita.Nom medicines gave-3.N.Sg
'Kumar having caught a cough and a cold, Sarita gave him medication.' Telugu

Similarly, Lindholm (1975) attributes the occurrence of such exceptions in Tamil to a cause-effect relation between the matrix and the subordinate clauses, and he adds another factor which he calls "natural relevance". According to natural relevance, it is not enough to have a cause-effect relation between the CNP and matrix clauses; the relations must also follow naturally — or, as I understand it, the relation must belong to the natural world. For example, the CNP and matrix clauses in (12) exhibit a cause-effect relation, but the sentence is ungrammatical because the relation lacks natural relevance (Lindholm 1975, 80 (3.37)). Compare to (6) in which the relation between rain and the flowing of the river is a cause-effect relation that is naturally relevant (Lindholm 1975, 75–83).

(12) [maze penj-u] kaḍe=le ellaam koḍe vittu pooccu
[rain fall-CNP] shop=Loc all umbrella sell went
'It rained and umbrellas got sold out at all the shops.' Tamil

Lindholm's analysis works for the Telugu examples (10a–h) above in which the cause-effect relations seem to be naturally relevant. In (10i), however, repeated here as (13a), the cause-effect relation is between a bomb explosion and the death of many people. Let us assume that the cause-effect relation between the two incidents is 'naturally relevant' — although the idea that bomb explosions are 'natural' is suspect. As expected, sentence (13a) is grammatical. If this analysis is correct — or sufficient — one would expect (13b) to be grammatical as well, for it also indicates that some disaster happened leading to a sad outcome. The only difference between (13a) and (13b) is that the latter mentions the agent behind the disaster, while the former does not. Apparently, mentioning the agent is the reason why (13b) is ungrammatical.

(13) a. [baambu peel-i] caala mandi canipooyaa-ru
[bomb.Nom explode-CNP] many people.Nom died-3.M.Pl
'A bomb exploded, and many people died.' Telugu

b. *[kumaar baambu=ni peelc-i] caala mandi canipooyaa-ru
[Kumar.Nom bomb=Acc explode-CNP] many people.Nom died-3.M.Pl
'Kumar exploded a bomb, and many people died.' Telugu

A more adequate explanation is offered by Klaiman (1981). She holds that exceptions to adjunct control in Bengali are allowed only when both the CNP and matrix clauses express a non-volitional activity. If one of the clauses expresses a volitional activity, disjoint subjects result in ungrammaticality. This is exactly the case of (13b) above; the CNP clause expresses a volition activity, which seems to be the reason why the sentence is unacceptable. The same is true if the matrix clause expresses a volitional activity, as (14) illustrates.

(14) *[baambu peel-i] caala mandi poliis=ni pilicaa-ru
[bomb.Nom explode-CNP] many people police=Acc called-3.M.Pl
'A bomb exploded, and many people called the police. Telugu

Klaiman's analysis is purely semantic. She explicitly rules out syntax and the possibility that "any existing theoretical model can handle the facts" (Klaiman 1981, 126). Nevertheless, her analysis may be translated into syntactic terms without undermining the semantic nature of the account. The following sections set out to do this and to show that what appears to be an exception to Adjunct Control in Telugu is not an exception.

3 Syntactic Analysis

A closer look at the Telugu sentences in (10) shows that what Klaiman describes as non-volitional activities correspond in syntactic terms to unaccusative structures. Each of the grammatical sentences in (10) contains two unaccusative predicates, one in the CNP clause and one in the matrix clause. By comparison, the ungrammatical structures (13b) and (14) contain at least one clause that is not unaccusative.

By "not unaccusative" I refer to, not only transitive and unergative, but also experiential predicates. For example, the sentences in (15) are ungrammatical because each contains one experiential predicate.[5]

(15) a. *[baambu peel-i] kumaar=ki koopam waccin-di
[bomb.Nom explode-CNP] Kumar=DAT anger came-3.N.Sg
'A bomb exploded, and Kumar got angry.' Telugu

b. *[ammaayi putt-i] andaru santooʃincaa-ru
[girl.Nom born-CNP] family.Nom became happy-3.M.Pl
'A girl was born; the family was happy.' Telugu

At first blush, the exclusion of experiential predicates from the category of unaccusative predicates sounds suspect. This exclusion, however, follows from the locus of the subjects of these predicates and how low they may be in their corresponding clauses. The standard assumption is that unaccusative predicates license themes that are base-generated low in the structure. Themes, along with goals and patients, are considered the lowest of all arguments. They are generated below causers, which in turn are generated below experiencers (Landau 2001, 120 and works within). This implies that the non-volitional subjects in (10a–i) are themes that are generated low in the structure, probably as complements of V^0. Subjects of transitive/unergative and experiential predicates, on the other hand, are generated in higher positions.

It is desirable to have independent evidence that the unaccusative predicates under investigation contain themes that are realized low in the structure, probably in the locus of their first merge. Evidence comes from unaccusative structures that contain a theme and a locative expression. Although Telugu is an SOV language, with the subject canonically occupying a sentence initial position, (16a), if an unaccusative predicate is involved, the locative expression is realized sentence-initially, (16b–c). These examples are not unexpected, given that the locus of locative expressions is higher than the locus of themes (Grimshaw 1990, 24).

(16) a. kumaar maa=uuri=loo baambu=ni pelcaa-ḍu
Kumar.Nom my=town=Loc bomb=Acc exploded-3.M.Sg
'Kumar exploded a bomb in my town.' Telugu

b. maa=uuri=loo caala mandi canipooyaa-ru
my=town=Loc many people died-3.M.Pl
'In my town many people died.' Telugu

[5]This is an important point because Klaiman's (Klaiman 1981) definition of non-volitional predicates seems to include experiential predicates. She presents the two examples in (i) which include one experiential predicate in the matrix clause and two disjoint subjects (Klaiman 1981, 113, (4.55a–b)). As far as I know, none of the grammatical examples in her study includes a CNP experiential predicate and two disjoint subjects.

((i)) a. [ṭaeks bere giy-e] aneke=r kasṭo hoyece
[tax increase-CNP] many=Gen difficulty became
'Taxes increaded and many people had difficulties' Bengali

b. [bristi por-e] caaside-r laabh holo
[rain fall-CNP] farmers-Gen profit became
'It rained and the farmers profited.' Bengali

The analysis offered in this article tries to account for the Telugu data. Concerning the Bengali examples in (i), I do not have an explanation.

c. pollalu=loo ceṭlu/mokkalu perigaa-yi
field=Loc trees/plants.Nom grew-3.N.Pl
'In the field the trees/plants grew.' Telugu

This observation extends to the exceptions to adjunct control in (10) above, some of which are repeated here with locative expressions.

(17) a. [pollalu=loo warʃam paḍ-i] ceṭlu/mokkalu perigaa-yi
[[fields=Loc rain.Nom fall-CNP] trees/plants.Nom grew-3.N.Pl
'The rain fell on the fields, and the trees/plants grew.' Telugu

b. [kolkata=loo baambu peel-i] caala mandi canipooyaa-ru
[Calcutta=Loc bomb.Nom explode-CNP] many people.Nom died-3.M.Pl]
'A bomb exploded in Calcutta, and many people died.' Telugu

c. [naa=uuri=loo tufaanu wacc-i] naa=illu kuulin-di
[my=town=Loc flood come-CNP] my=house.Nom collapsed-3.N.Sg
'The flood came to my town, and my house collapsed.' Telugu

This said, it is important to note that, owing to the free word order in Telugu, unaccusative structures with a sentence-initial theme followed by a locative expression are also acceptable, (18a–b). Nevertheless, sentences (16b–c) are the unmarked situation.

(18) a. caala mandi maa=uuri=loo canipooyaa-ru
many people my=town=Loc died-3.M.Pl
'Many people, in my town, died.' Telugu

b. ceṭlu/mokkalu pollalu=loo perigaa-yi
trees/plants.Nom field=Loc grew-3.N.Pl
'The trees/plants, in the field, grew.' Telugu

Based on the above discussion, I consider structures that involve unaccusative predicates in the CNP and matrix clauses as having coreferential null expletives pro^{EXP} in the subject positions, while the themes maintain their positions low in the clauses. In other words, the sentences in (10) have the structures in (19).[6]

(19) a. [$\Delta_{i/*k}$ tufaanu wacc-i] ***pro***$^{\text{EXP}_i}$ naa=illu kuulin-di
[flood come-CNP] my=house.Nom collapsed-3.N.Sg
'The flood came, and my house collapsed.' Telugu

b. [$\Delta_{i/*k}$ simla-loo mancu paḍ-i] ***pro***$^{\text{EXP}_i}$ ḍhillii-loo calla paḍin-di
[Simla-Loc snow fall-CNP] Delhi-Loc cool became.3.N.Sg
'The snow fell in Simla, and it became cool in Delhi.' Telugu

c. [$\Delta_{i/*k}$ bayaṭa baagaa calla-paḍ-i] ***pro***$^{\text{EXP}_i}$ inṭi-loo callagaa undi
[outside well cold-fall-CNP] house-Loc cool is
'Having become cool outside, it is cool in the house.' Telugu

d. [$\Delta_{i/*k}$ warʃam paḍ-i] ***pro***$^{\text{EXP}_i}$ ceṭlu/mokkalu perigaa-yi
[rain.Nom fall-CNP] trees/plants.Nom grew-3.N.Pl
'The rain fell, and the trees/plants grew.' Telugu

e. [$\Delta_{i/*k}$ cali-kaalam wacc-i] ***pro***$^{\text{EXP}_i}$ aakulu raalipooyaa-yi
[cold-weather.Nom come-CNP] leaves fell-3.N.Pl
'Winter arrived/came, and the leaves fell down.' Telugu

f. [$\Delta_{i/*k}$ eṇḍa-kaalam wacc-i] ***pro***$^{\text{EXP}_i}$ ceruwulu eṇḍipooyaa-yi
[hot-weather.Nom come-CNP] tanks dried up-3.N.Pl
'Summer came, and the water tanks dried up.' Telugu

[6]The coindexation between $\Delta$and pro^{EXP} in (19) means that the CNP subject may not have an independent reference; it has to be understood as an expletive controlled by the matrix expletive. Sections 4 and 5 suggest that this control relation is established through movement.

g. [$\Delta_{i/*k}$ kurcii wirig-i] ***pro***$^{\text{EXP}_i}$ kumaar kinda paḍaa-ḍu
[chair.Nom break-CNP] Kumar.Nom down fell-3.M.Sg
'The chair broke, and Kumar fell down.' Telugu

h. [$\Delta_{i/*k}$ bhuukampam wacc-i] ***pro***$^{\text{EXP}_i}$ caala kaṭṭaḍaalu kuulipooyaa-yi
[earthquake.Nom came-CNP] many buildings.Nom died-3.N.Pl
'An earthquake came, and many buildings collapsed.' Telugu

i. [$\Delta_{i/*k}$ baambu peel-i] ***pro***$^{\text{EXP}_i}$ caala mandi canipooyaa-ru
[bomb.Nom explode-CNP] many people.Nom died-3.M.Pl
'A bomb exploded, and many people died.' Telugu

The expletive is null because Telugu does not have overt expletives, which is expected in pro-drop languages in general. This idea is confirmed by Subbarao and Murthy (1999, 217) who maintain that Telugu has "no pleonastic or expletive expressions such as *it* or *there*". Similarly, Rao (2002, 37–39) holds that "expletives in Telugu are obligatorily null".

If this observation is correct, then exceptions to adjunct control in Telugu are non-exceptions. That is, they too qualify as instances of control — more specifically, Expletive Control — into CNP clauses. This conclusion, however, is challenged on empirical and theoretical grounds. The following section spells out the details. Section 5 offers a solution.

4 Problem

The null expletive *pro*$^{\text{EXP}}$ involved in Telugu Expletive Control seems to resemble the English expletive *there*. In English, *there* does not trigger agreement on the verb. Rather, the verb agrees with another NP that is associated with the expletive. To illustrate, in (20a) the verb agrees with the singular associate *one secretary*, while in (20b) the verb shows plural agreement with *two secretaries*. Like English *there*, Telugu null expletives, the type I assume to exist in Expletive Control structures, do not enter an agreement relation with the verb. The verb agrees with a nominative NP, as (21a–b) show. In (21a), *warʃam* 'rain' is singular; the verb shows singular agreement. In (21b), *warʃaalu* 'rains' is plural; the verb shows plural agreement.

(20) a. There is one secretary in this room.
b. There are two secretaries in this rooms.

(21) a. *pro*$^{\text{EXP}}$ warʃam paḍin-di
rain.Nom fell.3.N.Sg
'It rained.' Literally: 'The rain fell.' Telugu

b. *pro*$^{\text{EXP}}$ warʃaalu paḍaa-yi
rain.Nom fell-3.N.Pl
'It rained.' Literally: 'The rains fell.' Telugu

While adjunct control is allowed in English, (22a–b), *there*-Expletive Control (hereafter, Expletive Control) is banned, (23a–b), unless the expletive is phonologically realized in the adjunct as well, (24a–b) (Lasnik 1992, 244 (51–54)).[7]

(22) a. [Δ_i having witnessed the robbery] John$_i$ aided the investigation.

[7] The structures in (23) involve an existential expletive *there*. English also has a locative *there*, (i). When the latter is part of a control structure, (ii), the outcome is control by associate. That is, the associate of the expletive, *two men*, rather than the expletive itself controls the unpronounced subject of the adjunct (Chomsky (1995, 274); Cardinaletti (1997, 524)). Although the Telugu *pro*$^{\text{EXP}}$ may resemble the locative expletive, Telugu does not license control by associate.

(i) There arrived two more politicians.

(ii) There entered two men$_i$ [without Δ_i identifying themselves]

I consider this resemblance orthogonal to the discussion in the rest of the article. What is important for our purposes is that English does not license *there*-Expletive Control; this fact poses a challenge to the claim that Telugu has Expletive Control.

b. Harry$_i$ was a witness [without Δ_i being a victim].

(23) a. *[Δ_i having been a robbery] there$_i$ was an investigation.
b. *There$_i$ was a crime [without Δ_i being a victim].

(24) a. [There having been a robbery] there was an investigation.
b. There was a crime [without there being a victim].

Lasnik (1992) analyzes the sentences in (23) within the PRO Theory of Control. This analysis does not work for Telugu, however. The reason is that the different versions of PRO Theory assume that the subordinate subject is obligatory silent. While this is true for Telugu Forward Control, (25), the assumption is not true for Telugu Backward and Copy Control, (26)–(27). In (26a–b), the subordinate subjects are pronounced while the matrix subjects are implied. In (27a–b), both subjects are pronounced and obligatorily coreferential. I consider these examples sufficient to avoid an analysis within the PRO Theory of Control and to find an answer elsewhere.

(25) **Forward Control**

a. kumaar$_i$ [$\Delta_{i/*k}$ aakali wees-i] saandwic tinnaa-ḍu
Kumar.Nom [Δ hunger.Nom fall-CNP] sandwich ate-3.M.Sg
'Having felt hungry, Kumar ate a sandwich.' Telugu

b. [$\Delta_{i/*k}$ jwaram wacc-i] kumaar$_i$ haaspaṭal weḷḷaa-ḍu
[Δ.Dat fever.Nom come-CNP] Kumar.Nom hospital went-3.M.Sg
'Having had a fever, Kumar went to the hospital.' Telugu

(26) **Backward Control**

a. $\Delta_{i/*k}$ [kumaar=ki$_i$ aakali wees-i] saandwic tinnaa-ḍu
Δ [Kumar=Dat hunger.Nom fall-CNP] sandwich ate-3.M.Sg
'Having felt hungry, Kumar ate a sandwich.' Telugu

b. $\Delta_{i/*k}$ [kumaar=ki$_i$ jwaram wacc-i] haaspaṭal weḷḷaa-ḍu
Δ [Kumar=Dat fever.NOM come-CNP] hospital went-3.M.Sg
'Having had a fever, Kumar went to the hospital.' Telugu

(27) **Copy Control**

a. [kumaar=ki$_i$ aakali wees-i] atanu$_i$/aa pichooḍu$_i$/kumaar saandwic tinnaa-ḍu
[Kumar=Dat hunger.Nom fall-CNP] he/that idiot/Kumar.Nom sandwich ate-3.M.Sg
'Kumar felt hungry, and he/the idiot/Kumar ate a sandwich.' Telugu

b. [kumaar=ki$_i$ jwaram wacc-i] atanu$_i$/kumaar haaspaṭal weḷḷaa-ḍu
[Kumar=Dat fever.Nom come-CNP] he/Kumar.Nom hospital went-3.M.Sg
'Kumar had a fever, and he/Kumar went to the hospital.' Telugu

c. [Kumaar=ee tappu cees-i] kumaar=ee eeḍawatam modalupeṭṭaa-ḍu
[Kumar.Nom=Emph mistake do-CNP] Kumar.Nom=Emph crying started-3.M.Sg
'Kumar started crying although he has made a mistake.' Telugu

Within generative linguistics, Backward and Copy Control structures of the type exemplified in (26)–(27) have warranted an analysis of control as movement (see Polinsky and Potsdam 2002a,b, Monahan 2003, Haddad 2007, 2009, 2010a, Potsdam 2009; among several others).[8] According to the movement approach, the subject in a control structure starts out in the subordinate clause and moves to the matrix clause, whereby movement is understood as copy-plus-merge. The result is two copies of the same element at PF, one of which may have to be deleted. If the subordinate copy is deleted, the structure is realized as an instance of Forward Control. If, on the other hand, the matrix copy is deleted, the result is Backward Control. If both copies escape deletion, the outcome is Copy

[8]It is worth mentioning that there is speaker variation with respect to Copy Control structures like (27a–b). My observation is that they are found more acceptable by younger native speakers of Telugu; these are speaker in their 20's or early 30's. Sentence (27c), on the other hand, has been judged as grammatical by all the native speakers I consulted.

Control. Given that all three types of control are licensed in Telugu, I will limit the discussion to the Movement Theory of Control (Hornstein 1999).[9]

Hornstein (2001) provides an analysis of (23a–b), repeated here as (28), within the framework of the Movement Theory of Control. He argues that the unacceptability of these structures follows from the restriction that all merge has to be cyclic.

(28) a. *[Δ_i having been a robbery] there$_i$ was was investigation.
b. *There$_i$ was a crime [without Δ_i being a victim].

To elaborate, building on Nunes (1995, 2004), Hornstein holds that adjunct control is derivationally the outcome of sideward movement. In this kind of movement, an element copies out of one syntactic object (SO1) and merges in another (SO2). SO1 and SO2 do not have to be connected. Take, for example, sentence (29). The sentence has the derivation in (30). The adjunct clause and the matrix clause form independently, as (30a) illustrates. *Harry* copies out of the adjunct and merges in Spec,vP of the matrix clause, (30b). After movement takes place, the adjunct merges with the matrix clause; Hornstein assumes that this merge takes place at vP or VP. In this case, the adjunct clause merges with matrix vP, (30c). Upon merge, the adjunct becomes an island. In (30d), the matrix clause projects as high as IP; subsequently, *Harry* moves to Spec,IP, and CP projects. The highest copy of *Harry* c-commands the lower copies and forms a chain with each of them, as the dotted arrows show. At PF, the lower copy in each chain is deleted; the result is the structure in (30e).[10] Notice that the derivation proceeds cyclically. That is, it obeys the Extension Condition (Chomsky 1995, 248) which holds that merge extends the structure by applying at the root.

(29) Harry$_i$ was a witness [without Δ_i being a victim].

(30) a. [$_{\text{Adjunct}}$ without Harry being a victim] [$_{\text{Matrix vP}}$ was a witness]

b. [$_{\text{Adjunct}}$ without **Harry** being a victim] [$_{\text{Matrix vP}}$ **Harry** was a witness]

c. [$_{\text{vP}}$ [$_{\text{vP}}$ **Harry** was a witness] [$_{\text{Adjunct}}$ without **Harry** being a victim]]

d. [$_{\text{CP}}$ [$_{\text{IP}}$ **Harry** [$_{\text{vP}}$ [$_{\text{vP}}$ **Harry** was a witness] [$_{\text{Adjunct}}$ without **Harry** being a victim]]]]

e. [$_{\text{CP}}$ [$_{\text{IP}}$ **Harry** [$_{\text{vP}}$ [$_{\text{vP}}$ ~~**Harry**~~ was a witness] [$_{\text{Adjunct}}$ without ~~**Harry**~~ being a victim]]]]

Now observe the derivation of the ungrammatical sentence (28b) above as presented in (31). The adjunct clause and the matrix clause form independently, (31a). Following, the subordinate subject — the expletive — undergoes sideward movement to the matrix clause, (31b). Notice that matrix IP has already projected; this is so because the expletive can only merge at Spec,IP (see, however, Richards 2006 and Deal 2008 for an argument that expletives enter the derivation lower as the specifiers of certain kinds of v).

[9] See Davison (2008) who argues that PRO vs. movement in control follows from case restriction. Languages that do not license a dative subject in the subordinate clause of control constructions, such as Hindi-Urdu, are more likely to include PRO. Languages that allow control structures with a subordinate dative subject are more likely to be derived via movement. Telugu belongs to the latter category.

[10] It is worth mentioning that the derivation in (30), especially as pertaining to the pronunciation/deletion of copies, is slightly different from the discussion in Hornstein (2001) and more in line with Nunes' (2004) Copy-plus-Merge Theory of Movement. According to Nunes, when a copy moves, it undergoes four independent steps: copy, merge, form chain, and chain reduction. This approach makes it possible for *Harry* in (30b) to move between the two unconnected structures: the adjunct and the matrix clause. In other words, movement may take place between two positions that are not in a c-command relationship. Nevertheless, the two copies that result from movement will still have to enter a c-command relationship and form a chain. This is so because chain reduction — or deletion of copies — only targets copies in chains. Chain reduction takes place at PF in order for the structure to be mapped into a linear order without violating Kayne's (1994) Linear Correspondence Axiom (LCA). The LCA states that at PF two overt copies that are related by movement cannot be in a precedence relation; that is, they cannot be dominated by two non-terminal nodes that are in a c-command relationship. Therefore, one of them has to be deleted.

(31) a. [$_{\text{Adjunct}}$ without **there** being a victim] [$_{\text{Matrix IP}}$ [$_{\text{VP}}$ was a witness]]

b. [$_{\text{Adjunct}}$ without **there** being a victim] [$_{\text{Matrix IP}}$ **there** [$_{\text{VP}}$ was a witness]]

After movement takes place, it is time for the adjunct clause to merge with the matrix clause, presumably at VP, (32). This is not possible, however. According to Hornstein, (32) is blocked by (31b). The reason is that matrix I^0 has already projected in order to license the movement of the expletive. Accordingly, VP of the matrix clause is no longer a root, and the adjunct cannot undergo merge at VP without violating the Extension Condition.

(32) Blocked: [$_{\text{Matrix IP}}$ **there** [$_{\text{vP}}$ [$_{\text{vP}}$ was a crime] [$_{\text{Adjunct}}$ without **there** being a victim]]]

By the same token, if the adjunct clause merges cyclically at VP, (33a), the expletive subject will not be able to move out of the subordinate clause, (33b), because the subordinate clause has already become an island upon its merge with the matrix clause.

(33) a. [$_{\text{vP}}$ [$_{\text{vP}}$ was a crime] [$_{\text{Adjunct}}$ without **there** being a victim]]

b. [$_{\text{Matrix IP}}$ **there** [$_{\text{vP}}$ [$_{\text{vP}}$ was a crime] [$_{\text{Adjunct}}$ without **there** being a victim]]]

Based on this analysis, Hornstein argues that the generalization in (34) is necessary for adjunct control to obtain.

(34) Movement from the adjunct must proceed through a theta position in the matrix.
(Hornstein 2001, 120 (119))

The generalization in (34) poses a challenge to the claim that Telugu has Expletive Control. The following section offers a possible solution. It suggests that the generalization in (34) is too restrictive. The section also shows that if the language allows late adjunction while still obeying the rules of linearization,[11] movement from the adjunct may proceed through a non-theta position in the matrix clause without violating the Extension Condition.

5 Solution

The common assumption is that adjunct clauses merge at vP or VP of the matrix clause. In Hornstein's analysis, as delineated in the previous section, this restriction seems to be the main reason why Expletive Control is banned in English. Telugu, on the other hand, allows the adjunct to merge at CP. Evidence comes from the Telugu Copy Control structure (35). In Haddad (2007, 2009), I suggest that while the CNP clause in Forward and Backward Control undergoes merge at matrix vP, in Copy Control structures like (35) it undergoes first merge at CP of the matrix clause. As (36a) shows, the subject undergoes sideward movement, copying out of the adjunct and merging at vP of the matrix clause. Matrix IP projects, allowing the subject to move to Spec,IP, (36b). Following, CP projects; this is when the adjunct undergoes merge with the matrix clause, (36c). As a result, the CNP and matrix subjects do not enter a c-command relationship, which explains why they do not induce a Condition C violation. According to Condition C, an r-expression in a given structure must be simply free (Chomsky 1986, 164-165). That is, it cannot be bound by any element, including another r-expression in the same structure. See Haddad (2010b) for a more detailed discussion.

(35) [kumaar illu pooy-i] kumaar picciwaaḍu ayyaa-ḍu
[Kumar.Nom house lose-CNP] Kumar.Nom a crazy man became-3.M.Sg
'Kumar lost his house, and Kumar went crazy.' Telugu

[11] Late adjunction as used here is different from Stepanov's (2001) late adjunction. The latter induces a violation of the Extension Condition.

(36) a. [$_{\text{Adjunct}}$ **Kumaar** illu pooy-i] [$_{\text{Matrix vP}}$ **Kumaar** picciwaaḍu ayyaa-ḍu]

b. [$_{\text{CP}}$ [$_{\text{IP}}$ Kumaar [$_{\text{vP}}$ Kumaar picciwaaḍu ayyaa-ḍu]]]

c. [$_{\text{CP}}$ [$_{\text{Adjunct}}$ **Kumaar** illu pooy-i] [$_{\text{CP}}$[$_{\text{IP}}$ **Kumaar** [$_{\text{vP}}$ **Kumaar** picciwaaḍu ayyaa-ḍu]]]]

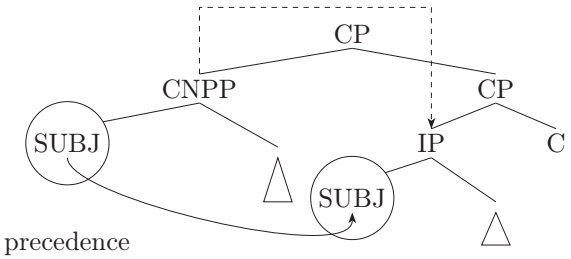
d. [$_{\text{CP}}$ [$_{\text{Adjunct}}$ **Kumaar** illu pooy-i] [$_{\text{CP}}$[$_{\text{IP}}$ **Kumaar** [$_{\text{vP}}$ ~~**Kumaar**~~ picciwaaḍu ayyaa-ḍu]]]]

Stated differently, let us assume that the CNP clause in (35) has to be base-generated at vP of the matrix clause before it moves to CP. In this case, the two copies of the subject would enter a c-command relation at some point in the derivation, as the dotted arrow in (37) shows. Thus, binding would be involved, resulting in the ungrammaticality of (35) due to a Condition C violation. But (35) is grammatical, which suggests that the derivation in (36) is on the right track.

(37) [$_{\text{CP}}$ [$_{\text{Adjunct}}$ **Kumaar illu pooy-i**] [$_{\text{CP}}$[$_{\text{Matrix}}$ **Kumaar** [$_{\text{vP}}$ [$_{\text{Adjunct}}$ **Kumaar illu pooy-i**] [$_{\text{vP}}$ Kumaar picciwaaḍu ayyaa-ḍu]]]]]

An additional remark regarding the derivation in (36) is important for the discussion of Expletive Control. Note that in (36d), the non-terminal nodes dominating the CNP and matrix subjects are in a c-command relationship, as (38) illustrates. According to Kayne (1994), the c-command relation between the non-terminal nodes is sufficient to place the terminal nodes — in this case, the subjects — in a precedence relation. This means that the two copies of 'Kumar' precede each other at PF. Kayne (1994) holds that two identical copies cannot be linearized at PF if they are in a precedence relationship, which is why one of them has to be deleted. This means that the derivation in (36) must crash, contrary to facts.

(38)



In Haddad (2009), I solve this problem by suggesting that the matrix clause in (36b) undergoes spell-out as a phase (Chomsky 2001, 2004) and is transformed into a phonological word à la Uriagereka (1999) prior to the adjunction of the CNP clause in (36c). The spell-out of matrix CP as a phase — that is, the spell-out of IP complement of C^{0} — converts the phase into a giant compound. In this sense, the outcome in (36d) looks more like (39); the box around the spelled out domain indicates that matrix IP is perceived as a phonological word at PF. The matrix subject hides inside this giant compound. Linearization cannot see into words. This is how the subject escapes deletion.

(39) [$_{\text{CP}}$ [$_{\text{Adjunct}}$ **Kumaar** illu pooy-i] [$_{\text{CP}}$ $\boxed{\text{[}_{\text{IP}}$ **Kumaar** [$_{\text{vP}}$ ~~Kumaar~~ picciwaaḍu ayyaa-ḍu]] $\boxed{\text{]}}$]]

Now we turn to Expletive Control. Consider sentence (40) and its derivation in (41). The adjunct and matrix clauses form independently in (41a). The null expletive copies out of the adjunct and merges in Spec,IP in the matrix clause, (41b). Following, the matrix clause projects as high as CP, (41c), allowing the CNP clause to undergo adjunction, (41d). Recall that, according to Hornstein, Expletive Control is banned in English because it induces a violation to the Extension Condition. No such violation is induced in (41). All merge extends the structure by applying to the root.

(40) [**pro**^{EXP} warʃam paḍ-i] **pro**^{EXP} mokkalu perigaa-yi
[rain.Nom fall-CNP] plants.Nom grew-3.N.Pl
'The rain having fallen, the plants grew.' Telugu

(41) a. [$_{Adjunct}$ **pro**^{EXP} warʃam paḍ-i] [$_{Matrix\ IP}$ [$_{vP}$ mokkalu perigaa-yi]]

b. [$_{Adjunct}$ **pro**^{EXP} warʃam paḍ-i] [$_{Matrix\ IP}$ **pro**^{EXP} [$_{vP}$ mokkalu perigaa-yi]]
c. [$_{CP}$ [$_{IP}$ **pro**^{EXP} [$_{vP}$ mokkalu perigaa-yi]]]
d. [$_{CP}$ [$_{Adjunct}$ **pro**^{EXP} warʃam paḍ-i] [$_{CP}$ [$_{IP}$ **pro**^{EXP} [$_{vP}$ mokkalu perigaa-yi]]]]

I mentioned earlier that the matrix clause in the Copy Control structure (35) is spelled out as a phase prior to the adjunction of the CNP clause. This step allows the structure to converge at PF without violating the rules of linearization. The same is not necessary when a null expletive is involved. Null expletives already lack phonological content. This means that no decisions need to be made regarding their deletion or pronunciation at PF for the purpose of linearization. This explains why (40) may also be realized as (42). Given that the matrix clause may undergo spell-out *after*, rather than *before*, the adjunction of the CNP clause, scrambling out of the matrix clause to a sentence-initial position is possible, albeit marked. The same is not possible with the Copy Control structure (35), as (43) illustrates. The reason is that the matrix IP is already a frozen compound that is opaque to all movement by the time the CNP clause adjoins to matrix CP.[12]

(42) mokkalu [***pro***^{EXP} warʃam paḍ-i] ***pro***^{EXP} perigaa-yi
plants.NOM [*pro*^{EXP} rain.Nom fall-CNP] *pro*^{EXP} grew-3.N.Pl
'The rain having fallen, the plants grew.' Telugu

(43) *picciwaaḍu [kumaar illu pooy-i] Kumaar ayyaa-ḍu
a crazy man [Kumar.Nom house lose-CNP] Kumar.Nom became-3.M.Sg
'Kumar lost his house, and Kumar went crazy.' Telugu

6 English Expletive Control Revisited

Let us assume that the solution in Section 5 is on the right track. One might wonder if this same analysis would work for English Expletive Control. In other words, could late adjunction be a solution, allowing (45) to be a possible derivation of (44)? In this case, the adjunct and matrix clauses would form independently, (45a). The subordinate expletive copies out of the adjunct and merges in Spec,IP of the matrix clause, (45b). The matrix clause projects as high as CP, (45c). The adjunct, having missed out on its chance to merge at matrix VP, undergoes late adjunction at matrix CP. The two copies of *there* do not enter a c-command relation. Accordingly, they do not form a chain, which is why neither of them is deleted.

(44) There having been a robbery, there was an investigation.

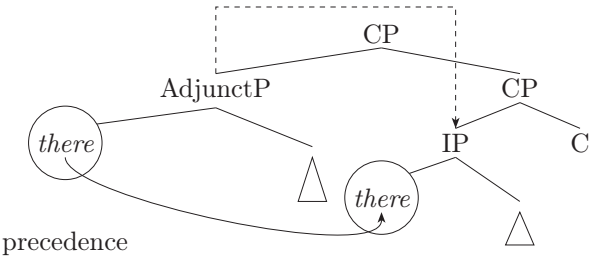
(45) a. [$_{Adjunct}$ **there** having been a robbery] [$_{Matrix\ IP}$[$_{VP}$ was an investigation]]

b. [$_{Adjunct}$ **there** having been a robbery] [$_{Matrix\ IP}$ **there** [$_{VP}$ was an investigation]]
c. [$_{CP}$ [$_{IP}$ **there** [$_{VP}$ was an investigation]]]
d. [$_{CP}$[$_{Adjunct}$ **there** having been a robbery] [$_{CP}$[$_{IP}$ **there** [$_{VP}$ was an investigation]]]]

[12] An anonymous *JSAL* reviewer asks why the expletive has to move at all. The question is based on the observation that the expletive does not move to check its Case feature or to check the theta-role feature of the target. In a forthcoming article Haddad (2010b), I suggest that the subject in Telugu Adjunct Control moves to satisfy neither its own needs nor those of its target. The subject moves to satisfy the requirements of the CNP clause and to license its merge with the matrix clause.

At first glance, the derivation in (45) seems to work. Closer examination shows that it suffers from a major flaw. Although the two copies of *there* do not enter a c-command relationship, the nodes that dominate them do, (46). This means that the expletives are in a precedence relation and one of them has to be deleted in order for the structure to be mapped into a linear order at PF. Nevertheless, deletion is not possible because it is contingent on chain formation. Given that chains require c-command and that the two copies of *there* do not c-command each other, chain formation and thus deletion fail to apply (Nunes 2004).

(46)



Sentence (44) is acceptable, however. Therefore, there must be a derivation that accounts for it. Two derivations are possible. Recall that Telugu Copy Control structures like (47) face a similar problem to the one delineated in (45)–(46) for English. To solve this problem, I suggest that the matrix clause is spelled out as a phase and transferred to the phonological component prior to the merge of the adjunct, (48). In this way, linearization is not able to detect the two subjects as non-distinct copies in a precedence relation simply because one of them hides inside a spelled-out domain. This is how both subjects escape deletion.

(47) [**kumaar** illu pooy-i] **kumaar** picciwaaḍu ayyaa-ḍu
[Kumar.Nom house lose-CNP] Kumar.Nom a crazy man became-3.M.Sg
'Kumar lost his house, and Kumar went crazy.' Telugu

(48) $[_{CP}\ [_{Adjunct}$**kumaar** illu pooy-i$]\ [_{CP}\ \boxed{[_{IP}$ **kumaar** $[_{vP}$ ~~**kumaar**~~ picciwaaḍu ayyaa-ḍu$]]}\ ]]$

It can be assumed that English Expletive Control resembles Telugu Copy Control. That is, at PF (45d) above looks like (49) in which the matrix clause has already been spelled out and transformed into a phonological word. When linearization applies, no precedence relation is detected, and the two copies of *there* escape deletion without inducing a violation. In this sense, English Expletive Control will be an instance of Copy Control.

(49) $[_{CP}\ [_{Adjunct}$**there** having been a robbery$]\ [_{CP}\ \boxed{[_{IP}$ **there** $[_{vP}$ was an investigation$]]}\ ]]$

The solution in (49) seems attractive, but it faces two major problems. First, it is more likely to be an ad hoc solution simply because, as far as I know, English does not license any other type of Copy Control. In addition, the solution is challenged by structures like (50) in which the merging site of the adjunct seems to be lower than CP. That is, the adjunct merges with the matrix clause before matrix CP is spelled out as a phase. Also, unless one assumes that the adjunct has undergone extraposition, the two copies of *there* are obviously in a c-command relationship. This means that the two copies may form a chain that is subject to chain reduction. The prediction is that one of the copies of there has to be deleted, which is not true. Such deletion leads to ungrammaticality, as (51) shows.

(50) There was a crime [without there being a victim].

(51) *There was a crime [without being a victim].

The second derivation is the one generally adopted in the literature. This derivation assumes that the two instances of *there* are copies of two different tokens selected from the numeration. That is, they are not related through movement. This means that the derivation of sentence (50) is (52). The

adjunct and matrix VP form independently, (52a). Subsequently, the adjunct merges with the matrix clause at VP, (52b). Matrix IP projects, and another copy of *there* selected from the numeration undergoes first merge in Spec,IP. Matrix CP projects, and the structure converges at PF, (52d).

(52) a. [$_{\text{Adjunct}}$ without **there** being a victim] [$_{\text{VP}}$ was a crime]
b. [$_{\text{VP}}$[$_{\text{VP}}$ was a crime] [$_{\text{Adjunct}}$ without **there** being a victim]]
c. [$_{\text{IP}}$ **there** [$_{\text{VP}}$[$_{\text{VP}}$ was a crime] [$_{\text{Adjunct}}$ without **there** being a victim]]]
d. [$_{\text{CP}}$[$_{\text{IP}}$ **there** [$_{\text{VP}}$[$_{\text{VP}}$ was a crime] [$_{\text{Adjunct}}$ without **there** being a victim]]]]

This derivation seems to be more accurate especially since Expletive Control is not enforced in English. That is, if the subject of the adjunct is an expletive, this does not necessarily entail that the subject of the matrix clause has to be an expletive, as the sentences in (53) illustrate.

(53) a. You don't get that big without there being some condition, (be it physical or mental).
b. No business shall be transacted without there being at least two officers and two ordinary members present.[13]

It is worth noting that there is a major difference between the derivation in (52) and the one assumed in the literature. Take Hornstein's (2001) account for example. According to Hornstein, the restriction on English Expletive Control and the derivation in (52) follow from the generalization in (34), repeated here as (54). The ban of Expletive Control in English as discussed in this section, however, follows from the ban on late adjunction and/or the rules of linearization. See Nunes (2004, 51–52) for a similar restriction on a derivation that employs across-the-board extraction.

(54) Movement from the adjunct must proceed through a theta position in the matrix.
(Hornstein 2001, 120: (119))

7 Conclusion

This article set out to show that Telugu structures that are normally referred to in the literature as exceptions to Adjunct Control into CNP clauses are not really exceptions. They are Expletive Control structures that are allowed only if the CNP clause and the matrix clause involve unaccusative predicates. The reason is that unaccusative predicates merge low in the structure, allowing a null expletive to fill the subject position.

The article mainly offered a syntactic account. It showed that Expletive Control is allowed in Telugu simply because the language allows late adjunction. Semantics does play a role, however. Although unaccusative CNP and matrix clauses are a prerequisite for Expletive Control to obtain, such structures seem to be limited to disasters, accidents, and natural phenomena. The reason might be because speakers look at such incidents as whole events rather than a topic and a comment. In other words, a structure like (55) does not depict a bomb or certain individuals and talks about them. Rather, it depicts two events: a bomb explosion and casualties. In this sense, the themes in (55) lack the quality of a topic. If we consider subjects to be topic-like (Rizzi 2005), then it is expected that the themes in (55) do not move to a subject position. Consequently, the subject positions are filled with expletives.

(55) [baabu peel-i] caala mandi canipooyaa-ru
[bomb.Nom explode-CNP] many people.Nom died-3.M.Pl
'A bomb exploded, and many people died.' Telugu

Once a theme acquires a topic-like status and moves to the subject position, it becomes part of the interpretation dependency in a control structure. In the Copy Control structure (56), *warʃam* 'the rain' functions as the subject of the CNP clause, determining the identity of *adi* 'it' in the matrix clause.

[13]Sentence (53a) and (53b) are from webpages (i) and (ii) respectively (last retrieved December 2009):
(i) `http://training.fitness.com/members-lounge/im-watching-show-tlc-23098.html`
(ii) `http://www.psychology.nottingham.ac.uk/bns/Constitution.htm`

(56) [warʃam pad-i] adi wiidhula=ni subram ceesin-di
[rain fall-CNP] it streets=Acc clean did-3.N.Sg
'The rain came down and cleaned the streets.' Telugu

Finally, the article limited the discussion to Expletive Control in Telugu. The hope is that the same observation and analysis would apply to other languages, such as Tamil and Hindi-Urdu, that also have exceptions to adjunct control into CNP clauses.[14] Nevertheless, the article does not make such a claim. Although similar in many ways, the languages of South Asia seem to have micro-differences regarding these exceptions; for example, Expletive Control seems to be less common in Hindi-Urdu than it is in Tamil or Bengali.[15] These micro-differences call for in-depth analysis of individual languages before arriving at any non-trivial generalization.

Acknowledgments

I would like to thank my language consultants: Suhitha Reddy Chigarapalli, Sashikiran Chowdary, Karthik Boinapally, Mahesh Tanniru, and Santhosh Kopidaka. I am also grateful to Ganga Bhavani Manthini and K.V. Subbarao for their feedback, especially as pertaining to the Copy Control data. This work was supported by the National Science Foundation under Grant No. BCS - 0131993 to Eric Potsdam.

[14]It might be worth mentioning that the analysis definitely does not work for Sinhala which allows real violations of the Same-Subject Condition; these are structures that contain accusative predicates (Gair 2003, Gair et al. 1998, Taylor 2006 among others). See Haddad (2010b) for an analysis.

[15]One factor that may be decisive in whether Expletive Control obtains in a particular language or not is the interaction between the referential quality of the themes (specific vs. non-specific, referential vs. non-referential), the strength of the EPP feature on T0, and the ability of unaccusative predicates to license case on the theme VP-internally (see Bhatt 2007 for more details). I did not have the chance to investigate these factors here, but they certainly call for further research.

References

Bhatt, Rajesh. 2007. Unaccusativity and case licensing. Handout, McGill University.

Cardinaletti, Anna. 1997. Agreement and control in expletive constructions. *Linguistics Inquiry* 28:521–533.

Chomsky, Noam. 1986. *Knowledge of Language*. New York: Praeger.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: The MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In M. Kenstowicz, ed., *Ken Hale: A life in language*, pages 1–52. Cambridge, MA: The MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In A. Belletti, ed., *Structure and beyond: The cartography of syntactic structures*, vol. 3, pages 104–131. Oxford: Oxford University Press.

Davison, Alice. 1981. Syntactic and semantic indeterminacy resolved: a mostly pragmatic analysis for the hindi conjunctive participle. In P. Cole, ed., *Radical pragmatics*, pages 101–128. New York: Academic Press.

Davison, Alice. 2008. A case restriction on control: Implications for movement. *Journal of South Asian Linguistics* 1:29–54.

Deal, Amy Rose. 2008. The origin and content of expletives: evidence from "selection.". *Syntax* 12:285–323.

Dwarikesh, Dwarika Prasad. 1971. *The historical syntax of the conjunctive participial phrase in the New Indo-Aryan Dialects of the Madhyadesha ("Midland") of Northern India*. Ph.D. thesis, University of Chicago, Illinois.

Gair, James W. 2003. Sinhala. In G. Cardona and D. Jain, eds., *The Indo-Aryan Languages*, pages 766–817. London: Routledge.

Gair, James W., Barbara Lust, Lelwala Sumangala, and Milan Rodrigo. 1998. Acquisition of null subjects and control in some Sinhala adverbial clauses. In B. Lust, ed., *Studies in South Asian linguistics: Sinhala and other South Asian languages*, pages 271–285. Oxford: Oxford University Press.

Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, MA: The MIT Press.

Haddad, Youssef A. 2007. *Adjunct Control in Telugu and Assamese*. Ph.D. thesis, University of Florida, Gainesville.

Haddad, Youssef A. 2009. Copy control in Telugu. *Journal of Linguistics* 45:69–109.

Haddad, Youssef A. 2010a. A non-stranding approach to resumption: Evidence from South Asia. *The Linguistic Review* To Appear.

Haddad, Youssef A. 2010b. Why things may move: Evidence from (circumstantial) control. *Journal of South Asian Linguistics* To Appear.

Haspelmath, Martin. 1995. The converb as a cross-linguistically valid category. In M. Haspelmath and E. König, eds., *Converbs in Cross-Linguistic Perspective*. Berlin: Mouton de Gruyter.

Hornstein, Norbert. 1999. Movement and control. *Linguistics Inquiry* 30:69–96.

Hornstein, Norbert. 2001. *Move! A Minimalist theory of construal: Generative syntax*. Oxford: Blackwell.

Kachru, Yamuna. 1981. On the syntax, semantics and pragmatics of the conjunctive participle in Hindi-Urdu. *Studies in the Linguistic Sciences* 11:39–49.

Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: The MIT Press.

Klaiman, M. H. 1981. *Volitionality and Subject in Bengali: A study of semantic parameters in grammatical processes*. Ph.D. thesis, University of Chicago, Illinois.

Landau, Idan. 2001. Control and extraposition: The case of super-equi. *Natural Languages and Linguistic Theory* 19:109–152.

Lasnik, Howard. 1992. Two notes on control and binding. In R. K. Larson, S. Iatridou, U. Lahiri, and J. Higginbotham, eds., *Control and Grammar*, pages 235–252. Dordrecht: Kluwer Academic Publishers.

Lindholm, James Milton. 1975. *The conceptual basis of the Tamil adverbial participle*. Ph.D. thesis, University of Chicago, Illinois.

Masica, Colin P. 2005. *Defining a Linguistic Area*. New Delhi: Chronicle Books.

Monahan, Philip. 2003. Backward object control in Korean. In *The Proceedings of the 22nd West Coast Conference on Formal Linguistics*, pages 356–369. Somerville, MA: Cascadilla.

Nunes, Jairo. 1995. *The Copy Theory of Movement and Linearization of Chains in the Minimalist Program*. Ph.D. thesis, University of Maryland, College Park.

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

Pandharipande, Rajeshwari V. 1997. *Marathi*. New York: Routledge.

Polinsky, Maria and Eric Potsdam. 2002a. Backward control. *Linguistics Inquiry* 33:245–282.

Polinsky, Maria and Eric Potsdam. 2002b. Backward control: Evidence from Malagasy. In A. Rackowski and N. Richards, eds., *MIT Working Papers in Linguistics 44: The Proceedings of the 8th Austronesian Formal Linguistics Association*, pages 257–272. Cambridge, MA: MIT Working Papers in Linguistics.

Potsdam, Eric. 2009. Malagasy backward object control. *Language* To appear.

Rao, Anuradha. 2002. *Null objects in Telugu and English: Merge, Binding Theory, and pro-drop*. Ph.D. thesis, Center Institute of English and Foreign Languages, Hyderabad, India.

Richards, Marc. 2006. On object shift, phases, and transitive expletive constructions in Germanic. *Linguistics Variation Year Book* 6:136–159.

Rizzi, Luigi. 2005. On some properties of subjects and topics. *Incontro di Grammatica Generativa* XXX.

Stephanov, Arthur. 2001. Late adjunction and minimalist phrase structure. *Syntax* 4:94–125.

Subbarao, KV and B. Lalitha Murthy. 1999. Lexical anaphors and pronouns in Telugu. In B. Lust, K. Wali, J. Gair, and K. Subbarao, eds., *Lexical anaphors and pronouns in selected South Asian languages*, pages 217–276. Berlin: Mouton de Gruyter.

Taylor, Chris. 2006. Perfect, sequence, recapitulation: common construal and the Sinhala conjunctive participle. *Santa Barbara Papers in Linguistics* 17:150–162.

Uriagereka, Juan. 1999. Multiple spell-out. In S. D. Epstein and N. Hornstein, eds., *Working Minimalism*, pages 251–282. Cambridge, MA: The MIT Press.

First Phase Syntax of Persian Complex Predicates: Argument Structure and Telicity

Marina Pantcheva, *CASTL, University of Tromsø*

Received November 21, 2008; Revised October 17, 2009

Abstract

In this paper, I propose an analysis of Persian complex predicates, based on the First Phase Verbal syntax developed by Ramchand (2008). I suggest that the light verbs lexicalize the subevent heads into which the verbal phrase is decomposed, while the preverbal element occupies the Rheme position and semantically unifies with the light verb to build one joint predication. Further, I propose a feature specification for some of the most productive light verbs. I discuss the role of the light verb and the preverb in determining the argument structure of the entire predicate and show how the aspectual properties of the complex predicate depend on the interaction between the preverb and the light verb.

1 Introduction

Persian is a language that makes extensive use of the so called complex predicates (CPr) — a predicate which consists of a non-verbal part, often referred to as *preverb* (Lazard 1957) and a semantically bleached verb, called *light verb*. The preverb and the light verb together build one predicate.[1]

(1) mina reza-ro[2] dust dare.
Mina Reza-om friend has
'Mina loves Reza.'

The preverb can represent different syntactic categories: noun, adjective, adverb, preposition, or prepositional phrase. Interestingly, certain light verbs tend to take preverbs belonging to certain categories. In Table 1, I present some of the most common light verbs and the preverb categories they productively combine with.[3]

An issue that has been the cause of much debate in the literature relates to the role of the two elements in the complex construction with respect to the aspectual properties of the complex predicate and its argument structure (Karimi-Doostan 1997, Karimi-Doostan 2005, Megerdoomian 2001, Megerdoomian 2002a, Folli et al. 2005). A common view is that the light verb is responsible for the projection of the external argument and, according to Karimi-Doostan, it also determines

[1] Abbreviations in glosses used in this paper are as follows: 1, 2, 3 – first, second and third person; cl – clitic; class – classifier; ez – Ezafe linker; om – object marker; pl – plural; pp – past participle.

[2] The clitic *–ro* (*-ra* is the formal/written form), commonly termed *object marker* attaches to all direct objects that are construed as specific.

[3] At the end of the paper, I have given examples for the complex predicates included in the charts.

JSAL Volume 2, Issue 1, December 2009.

Light Verb		N	P/PP	Adj/Adv
kærdæn	'do'	ok		
aværdæn	'bring'	ok	ok	ok
amædæn	'come'	ok	ok	ok
gereftæn	'take'	ok	ok	
dadæn	'give'	ok	ok	
keshidæn	'pull'	ok		ok
xordæn	'collide'	ok		
zædæn	'hit'	ok		
kærdæn	'make'	ok		ok
shodæn	'become'	ok		ok
oftadæn	'fall'	ok	ok	
ændaxtæn	'throw'	ok	ok	

TABLE 1: Preverb and light verbs combinations

the aspect of the complex predicate. Folli et al. (2005), however, claim that the (un)boundedness of the event is dependent entirely on the type of preverb the light verb combines with. The goal of the present paper is to discuss this issue and provide insight into the ways telicity arises in complex predicates. More specifically, I am going to show how each of the two elements contributes to the telicity of the entire predicate and will investigate the ways in which they interact.

The analysis of Persian complex predicates I propose is based on the First Phase Syntax research program developed in Ramchand (2008). According to her theory, events are decomposed into three subevents (*init*, *proc* and *res*), each corresponding to a distinct head in the verbal projection and introducing an event participant. Applying this system to the Persian data, I will investigate the question of what the contribution of the two components of the complex predicate is when it comes to its argument structure and telicity.

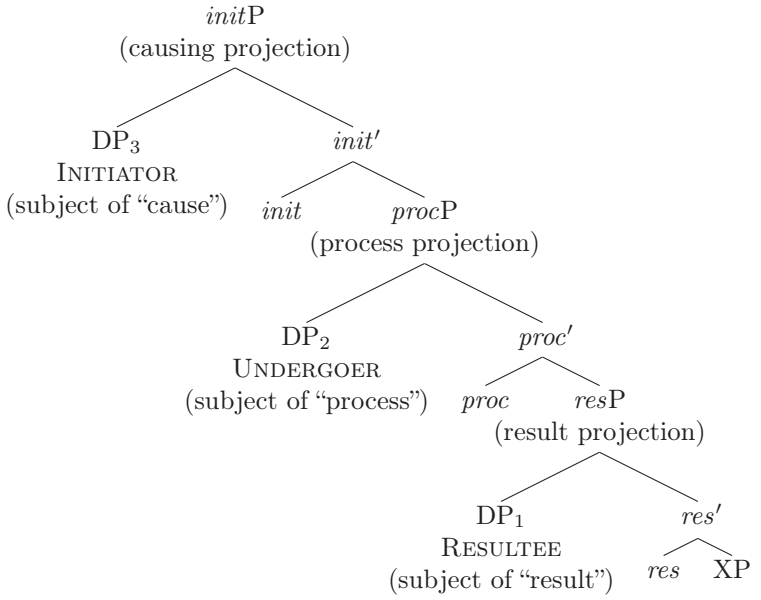
The paper is organized as follows. In Section 2, I briefly introduce the First Phase Syntax system. In Section 3, I apply the system to Persian complex predicates and lay out the proposal concerning the syntactic structure that underlies them. Section 4 deals with the feature specification of the light verbs according to the model described in Section 2. In Section 5, I handle the question of telicity of events by means of the tools provided by the system. Section 6 summarizes and concludes.

2 First Phase Syntax of Persian complex predicates

2.1 A quick guide to the Verbal First Phase Syntax

Ramchand's (2008) First Phase Syntax is characterized by the decomposition of the verbal domain into three distinct heads, each corresponding to a primitive element of events. The internal structure of the verbal phrase contains the following three subevent projections: *init*[iation]P, *proc*[ess]P, and *res*[ult]P. The first (*init*) and the third (*res*) are stative heads, while the second – *proc* – is the hallmark of dynamicity. Every dynamic verb, then, contains the *proc* head in its decomposition. The stative *init* and *res* heads, however, can be missing in the case of dynamic verbs. Each subevent head enters in a predicational relation with its specifier position, where we find the "subject" of the event. In (2), I present the maximal decomposition of the verb phrase.

(2)



As can be seen, the three core projections are:

- InitP: introduces the causation event and licenses the external argument (the Initiator)
- ProcP: specifies the process or the nature of the change and licenses the internal argument (the Undergoer)
- ResP: introduces the result state and licenses the holder of the result state (the Resultee)

Apart from the three thematic roles above, there exist *composite* roles which arise when the same DP argument occupies two (or more) specifier positions. This happens when a DP raises from the specifier of a lower subevent head to the specifier of a higher subevent head. In such cases, we have the roles of Initiator-Undergoer, Undergoer-Resultee, and Initiator-Undergoer-Resultee. The first one arises when the same argument is the holder of the initiational stage and undergoes the process/change (e.g. the sole argument of the verb *run*). The second one arises when the same argument undergoes the process/change specified by the *proc* head and holds the result state (e.g. the direct object of *break*). The third one arises when the same argument initiates the event, undergoes the process/change and is the holder of the result state (e.g. the argument of *arrive*). The composite thematic roles of the participants in the event are encoded in the lexical entry of the verb, that is, the verb determines whether a certain DP will raise from one specifier to another or not.

Crucially, a verb can lexicalize more than one head in the verbal phrase. Thus, in this model, verbs come in the lexicon with a categorial feature specification which determines which subevents they lexicalize. For example, a verb specified as $< init, proc >$ will spell out both the *init* and the *proc* head simultaneously. Depending on which subevent heads a verb lexicalizes, it belongs to a particular verb class. Thus, there is the class of $< init, proc >$ verbs, the class of $< init, proc, res >$ verbs, the class of $< proc, res >$ verbs, etc. If we are to connect these classes to the traditional aspectual classes, then activities are characterized by the features $< init, proc >$ or only $< proc >$, achievement verbs are specified as $< init, proc, res >$ or $< proc, res >$, statives have only the feature $< init >$, etc. When it comes to argument structure, unergatives are verb that are specified with the feature $< init >$, while unaccusatives lack this feature.

Concerning the semantic interpretation of the verbal phrase, the system employs compositional semantic rules that interpret the embedded predication via a causational semantics. Thus, in the verbal decomposition, we have two stative heads (*init* and *res*), the first one "leads to" the process subevent and hence is interpreted as initiation, the second one is "caused" by the process, and is therefore interpreted as result.

Clearly, the advantage of this system is that it allows for many different types of verbs to be put together by means of a fairly impoverished set of primitives, some general principles of lexical association and a compositional semantic rule based on the relation "leads-to."

A subevent descriptor is not restricted to taking another subevent phrase as a complement. An event head can also have non-verbal material (DP, AP, PP, etc.) occupying its complement position. Such non-verbal complements are called RHEMES (e.g. the XP in (2)). RHEMES are not subjects of events but part of the description of the predicate. Hence, there is an important difference between a DP in the RHEME position and a DP occupying the specifier of a subevent head. Namely, the first one builds one joint predication with the verb, while the latter is a verbal argument.

It is important to note that the (un)boundedness of the macro-event does not necessarily entail that there is a *res*P in the stucrure. The RHEME plays an important role in determining the telicity of *proc* verbs that do not instantiate *res*. As the material in the RHEME and the verb unify, a bounded RHEME makes the entire predication bounded. Examples for such bounded RHEMES are closed scale gradable adjectives, bounded Path PPs, and quantized nouns. Hence, whenever a $< proc >$ verb has such a bounded RHEME, a telic interpretation will arise for the entire macro-event, despite the fact that there is no *res* head in the structure.

3 Assembling the complex predicate

3.1 The role of the light verb

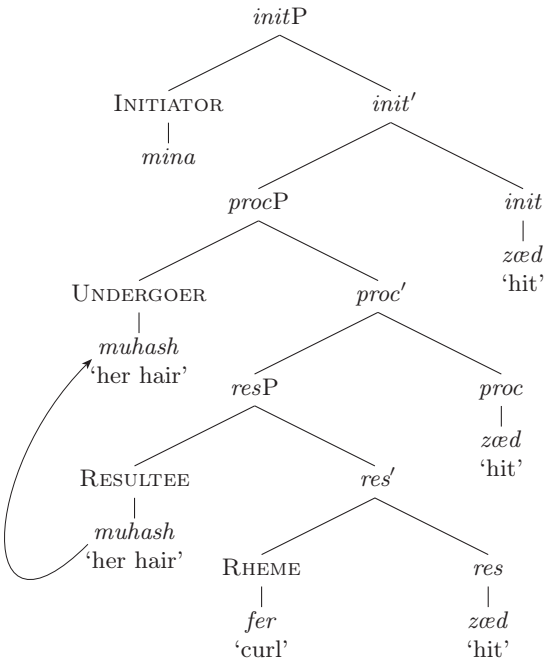
Butt (2003) argues that light verbs always have a main verb counterpart in the language. I take this to mean that there is no syntactic difference between light and heavy verbs. It is then logical to assume that light verbs lexicalize the subevent heads in the decomposed VP, just like heavy verbs do. The distinction between light and heavy verbs can be then due to the fact that the former have a very abstract semantics, while the latter have full lexical meaning. Take, for example, the verb *zædæn*, which, as a heavy verb, is agentive and punctual and means roughly "cause x to come into contact with y, quickly and forcefully," and can be best rendered by the English verb *hit* (see (3a)). The light verb *zædæn*, according to Family (2006, 60), also participates in agentive complex predicates that, in general, denote instantaneous actions, with the possibility of being iterated. The action usually involves change of state either of the agent herself, or of another entity. Thus, the light verb *zædæn* is impoverished semantically, however, it is not totally deprived of content. The semantic content of the heavy verb and the light verb *zædæn* share some meaning components, but the heavy verb carries a richer conceptual content. Note that the meaning of "hitting" is not preserved in the complex predicate in (3b), where there is no notion of impact whatsoever. Still, in both examples, the event is bounded and we have an agent.

(3) a. mina sæng-ro be divar zæd.
Mina stone-OM to wall hit
'Mina hit the stone at the wall.' (heavy verb *zædæn*)

b. mina mu-ha-sh-ro fer zæd.
Mina hairs-PL-3CL-OM curl hit
'Mina curled her hair.' (CPr with light verb *zædæn*)

As the reader can observe, the meaning of the complex predicate *fer zædæn* (curl hit) in (3b) is very specific, although, as I just argued, the light verb contributes to the predicate only a very abstract meaning. Hence, it is logical to conclude that the main conceptual-intentional content of the CPr comes from the preverb. With respect to the syntactic position of the preverb, I suggest that it occupies the RHEME position and semantically unifies with the light verb to build one joint predicate. The syntactic structure of the complex predicate *fer zædæn* (curl hit) 'to curl' in (3b) will be then as in the tree diagram in (4).

(4)



In the tree structure above, the light verb *zædæn* spells out all three subevent heads, thus projecting all three specifier positions in the VP. The direct object *muhash* 'her hair' is first merged in the lowest one as a Resultee and subsequently moves to Spec,*proc*P. As a consequence, the argument *muhash* acquires the composite role of Undergoer-Resultee, that is, it undergoes the process and holds the result state of having curls.[4] The subject *Mina* is merged directly in the specifier of *init* where it is interpreted as the Initiator of the event. The third noun element — the preverb *fer* 'curl' in the Rheme — is interpreted as part of the entire predicate.

3.2 The role of the preverb

Given that it is the light verb that lexicalizes the verbal heads, the argument structure of the whole complex predicate will depend on the feature specification of the light verb. By argument structure I mean the projection of the specifier positions of subevent heads, or, put in other words, the presence of the "subjects" of the subevents: Initiator, Undergoer, and Resultee. Thus, if we want to have an agentive complex predicate, we need to choose a light verb that has the feature $< init >$, so that the *init* head is spelled out and the Initiator position is projected. This is very much in accordance with the complex predicate analysis of Megerdoomian (2001) and Folli et al. (2005), who convincingly show that the light verbs in Persian determine the agentivity/causativity of the predicates they form, regardless of the preverb. Further, the presence of Undergoer and Resultee positions is also dependent on the light verb, and these are the positions occupied by the internal argument. Thus, in a sense, the presence of a direct object depends on the light verb. This suggestion is in line with Megerdoomian's claim that the light verb projects the internal argument of the complex predicate (Megerdoomian 2001, 2002a). However, this goes against some analyses of Persian complex predicates, according to which it is the preverb that contributes the internal argument (see, for instance, Karimi-Doostan 1997, 2005). This disagreement can be, however, resolved, as the system

[4] The *-ro* marker on the direct object *muhash* 'her hair' is due to the fact that it is construed as specific. I assume that specific direct objects (i.e., specific Undergoers, Resultees, and Undergoer-Resultees) undergo a movement to a position higher in the tree. This is in line with analyses proposed by various researchers, according to whom specific direct objects appear in a higher position than their non-specific counterpart and argue that this is the result of a syntactic movement. For instance, Browning and Karimi (1994) propose that specific DPs move to a VP-external position for case reasons. Karimi (2005) also shares the view that all direct objects are merged in the same position in the verbal phrase, but the specific objects move to the specifier of *v*P to receive interpretation.

proposed here provides a way to unify the two approaches. Consider the following examples:

(5) a. mina gusht-ro næmæk zæd.
Mina meat-OM salt hit
'Mina salted the meat.'
b. mina chærx zæd.
Mina turn hit
'Mina turned (around).'

The light verb in both sentences remains constant, still, the a-example features an external and an internal argument, while the b-example appears to have just an external argument. Since the element that varies in the two sentences is the preverb (*næmæk* 'salt' versus *chærx* 'turn'), an option is to conclude that the preverb *næmæk* contributes the internal argument in (5a). The logical question is how the preverb in the RHEME position can add an argument, given that it is the light verb that projects the subevent head and consequently also the specifiers thereof, where we find the arguments of the predicate.

Recall that under the approach assumed here, one DP can raise through multiple specifiers of subevents, thus acquiring a composite thematic role. Maintaining the proposal that the INITIATOR, UNDERGOER and RESULTEE positions are contributed by the light verb, a solution of this problem will be if we assume that certain preverbs require the DP to undergo movement to a particular specifier (or specifiers), while other preverbs do not. Thus, the preverb in a complex predicate conditions the raising of a DP argument in the same way as a heavy verb does. Applied to the data in (5), this would mean that the preverb *næmæk* 'salt' prohibits the DP *gusht* 'meat' to raise to the INITIATOR position, thus enforcing the merge of a distinct DP, *Mina*, in Spec,*init*P. The preverb *chærx*, on the contrary, requires the DP *Mina* to go through all specifiers, which results in there being just one argument but with the composite role or INITIATOR-UNDERGOER-RESULTEE. Hence, in a way, both the light verb and the preverb play a role in determining the presence of an internal argument: the light verb contributes the syntactic position for it, and the preverb says whether it is going to be a distinct DP from the external argument or not. A consequence of this proposal is that light verbs differ from heavy verbs in that the latter determine whether the DPs occupying the specifiers of subevent heads are distinct or not, while the former do not have this information encoded in their lexical entry.

The proposal that preverbs indirectly affect the argument structure of the complex predicate by determining the raising of DPs from one specifier position to another leads to a prediction. The prediction is that when a given preverb combines with two distinct light verbs with the same feature specification (but different abstract semantic contents), the argument structure of the complex predicate will be the same. Take as an example the light verbs *zædæn* 'hit' and *kærdæn* 'make,' which participate in resultative complex predicates (i.e., both have the feature *res*) with agents (i.e., both light verbs have the feature *init*). We expect the same number and thematic roles of the arguments of the complex predicates constructed by combining the same preverb with one of these two light verbs. This is illustrated by the example in (6).

(6) a. mina mu-ha-sh-ro ræng zæd.
Mina hair-PL-3CL-OM paint hit
'Mina dyed her hair.'
b. mina mu-ha-sh-ro ræng kærd.
Mina hair-PL-3CL-OM paint made
'Mina dyed her hair.'

Thus, in (6) we have the same preverb *ræng* and two distincts $< init, proc, res >$ light verbs — *zædæn* in (6a) and *kærdæn* in (6b). The argument structure of the complex predicates is the same in the (a) and (b) example: there is an external argument (*Mina*) and a distinct internal argument

(*muhash* 'her hair'). What differs is the semantic interpretation because of the different abstract semantic contents of the two light verbs. The nuances in the meaning are somewhat difficult to define in a precise way, but they doubtlessly exist. The complex predicates in the minimal pair *ræng zædæn* (paint hit) and *ræng kærdæn* (paint make) are synonymous in that they both express the transitive event of painting/dyeing something. When we use the light verb *zædæn*, however, the implication is that the result state of the hair being dyed persist longer, and the change inflicted on the patient is more accentuated. In other words, the choice of the light verb *zædæn* leads to a complex predicate that focuses on the change of state and the following result state. With *kærdæn*, on the other hand, the main stress falls on the activity of dyeing, that is, on the process part of the event, and the result state is less emphasized.

3.3 Preverb modification

Preverb modification lends support to the hypothesis that preverbs occupy the rhematic position. The point is that modification of a preverbal noun element differs from the cases when a direct object is modified. Compare the (a) and (b) example in the data set below.

(7) a. mina do-ta shune gereft.
Mina two-CLASS comb got
'Mina received two combs.'
b. mina mu-ha-sh-ro do-ta shune zæd.
mina hair-PL-3CL-OM two-CLASS comb hit
'Mina combed her hair twice.'
(modified from Megerdoomian 2006)

In (7a), the noun *shune* 'comb' is an UNDERGOER-RESULTEE direct object of the verb *gereftæn* and thus the number of combs received by Mina is two. In example (7b), featuring a complex predicate, however, the numeral does not scope over the noun but is interpreted as modifying the whole event. The reason for this is that the numeral is inside the RHEME, where it gets interpreted as part of the whole predication.[5]

A similar generalization extends to adjectival modification of noun preverbs. Consider, for instance, the data in (8).

(8) reza kotak-e bædi xord.
Reza beating-EZ bad collided
'Reza was beaten badly.'

The adjective in this example is interpreted as modifying the whole event. Still, it clearly is part of the RHEME, as evidenced by the presence of the Ezafe linking morpheme. Ezafe appears on a noun whenever it is modified by an adjective. Thus, we can conclude that the adjective in (8) indeed modifies the noun preverb and is therefore inside the RHEME.

3.4 The distinction between direct objects and noun preverbs

Before concluding this section, it is worth investigating how the approach advocated here can handle one widely discussed issue in the literature on Persian complex predicate — the status of the noun preverb or, more precisely, the question of how noun preverbs differ from bare direct objects. The reason for this interest in noun preverbs is that they, like all other preverbs, invariably precede the light verb. As Persian is an SOV language, direct objects, too, are placed before the verb. Very often then it is not easy to decide whether a given noun is a preverb or an internal argument

[5] An anonymous reviewer suggested the possibility that the numeral is adjoined and scopes over *res*P. However, the presence of the classifier *-ta* indicates that the numeral is part of the extended projection of the noun *shune* 'comb' (in the sense of Cinque 2005). In addition, if *do-ta* really were an adjunct, nothing would prevent it from adjoining to the VP in (7a) thus giving rise to a reading where Mina receives a comb twice. This reading is, however, unavailable for (7a).

of the verb. This question has triggered much debate in the literature concerning the relation of the nominal element in CPrs and the light verb. The result is a two-way split: according to some researchers, noun preverbs are just like (bare) direct objects (Samvelian 2001, 2004). According to others, most notably Megerdoomian (2006), noun preverbs differ from direct objects and occupy a different position in the syntactic structure. This is also the hypothesis maintained in this paper.

In the syntactic decomposition of verbs adopted here, the arguments of the verb occupy one or more specifiers of the subevent heads. Hence, this is where we find the noun *sæng* 'stone' in the example in (9a), where *zædæn* is a heavy verb. Preverbs, as suggested above, are in the rhematic positions of the VP, therefore, the preverb *ræng* 'paint' in (9b) is hosted by the RHEME.

(9) a. mina be divar sæng zæd.
Mina to wall stone hit
'Mina hit a stone/stones at the wall.'
b. mina be divar ræng zæd.
mina to wall paint hit
'Mina painted the wall.'

Complex predicates with noun preverbs are then structurally different from direct object+verb constructions: in direct object+verb construction the noun is in the UNDERGOER and/or RESULTEE position, while in complex predicates, the noun is in the RHEME. Thus, in the case of complex predicates, there can be an internal argument occupying the specifiers of *proc* and *res*. This will result in a direct object+complex predicate combination, like the one in (10). (10) is a different way to say (9b), where instead of the PP *be divar* 'to the wall,' we have *divar* 'wall' as direct object, as evidenced by the object marker *-ro*.[6]

(10) mina divar-ro ræng zæd.
Mina wall-OM paint hit
'Mina painted the wall.'

If we try to introduce *divar* 'wall' as a direct object in (9a), the result will be ungrammaticality (see (11)). The reason is that the noun *sæng* 'stone' already occupies Spec,*proc*P and Spec,*res*P and these positions are not available for the intended direct object *divar* 'wall.'

(11) *mina divar-ro sæng zæd.
Mina wall-OM stone hit
Intended: 'Mina hit the wall with stones.'

Thus, the proposed position of preverbs in the RHEME allows us to draw a distinction between noun preverbs and direct objects, which arises due to the different structural position they occupy.

3.5 Summing up

To recapitulate this section, I proposed that the light verb in Persian complex predicates lexicalize the subevent heads in a decomposed verbal phrase. The preverbal element is hosted by the RHEME and semantically unifies with the light verb to form one predicate.

Light verbs are like heavy verbs in that they are specified for the same verbal features in the lexicon. Still, there are two differences: (i) light verbs have a bleached and abstract semantics, and (ii) light verbs do not determine how high a DP can raise from one specifier of a subevent head to another. That is, light verbs have no bearing as to whether a given DP will have a composite thematic role or not.

I suggested that it is the preverb which determines the raising of argument DPs. This led to the prediction that a given preverb should form complex predicates with identical argument structure,

[6]There are syntactic and semantic differences between the construction in (9b) and (10), which are discussed in Pantcheva (2008).

provided it combines with light verbs with the same feature specifications. Proving or disproving this hypothesis requires an extensive corpus study, which I leave for future research.

Finally, the syntactic structure of complex predicates proposed in this paper allows us to take a stand on the issue of whether bare objects and noun preverbs are the same thing or not. I suggested that they occupy different positions in the decomposed VP, which accounts for their different properties and behavior (see also Pantcheva 2008).

4 Light verb classes

In the preceding section, I proposed that the light verbs in Persian complex predicates lexicalize the subevent heads in the verbal phrase. Hence, light verbs can be classified into types according to their feature specification just like ordinary "heavy" verbs. Given that all light verbs examined in this paper are dynamic, they will all be specified for the feature $< proc >$.[7] The two feature that are left to investigate, then, are $< init >$ and $< res >$. In the subsections to follow, I focus on these two subevent heads and propose a feature specification of some of the most commonly used light verbs in Persian, thus grouping them into classes.

4.1 Light verbs & init

Let us start with the *init*[iation] subevent and see which light verbs are endowed with this feature. Consider first the example below.

(12) mina gul xord.
Mina deceit collided
'Mina got deceived.'

In this example, *Mina* experiences a deceit and carries the role of a proto-Patient, or, put in the terminology of the Verbal First Phase, the role of Undergoer. Crucially, *Mina* cannot be seen as the person initiating the deceit, hence, she is not the Initiator. This is further evidenced by the fact that the complex predicate in (12) is incompatible with agentive adverbials such as *æmdæn* 'intentionally.'

(13) #mina æmdæn gul xord.
Mina intentionally deceit collided
('Mina got deceived intentionally.')

Moreover, the unavailability of an Initiator position in the sentence in (12) gains support from the impossibility to add a Causer (i.e., an Initiator).

(14) *reza mina-ro gul xord.
Reza Mina-om deceit collided
('Reza deceived Mina.')

Accordingly, we can conclude that the Initiator position is not projected in the examples just discussed, which in turn implies that there is no *init* head in the structure. The lack of the *init* head can be straightforwardly explained if we assume that the light verb *xordæn* is not endowed with the feature *init*.

In order to express the Causer of Mina's deception, one need to substitute the light verb *xordæn* 'collide' for the light verb *zædæn* 'hit.'

(15) reza mina-ro gul zæd.
Reza Mina-om deceit hit
'Reza deceived Mina.'

[7] In this paper, I will abstract away from the stative light verb *dashtæn* 'to have,' which forms stative complex predicates.

In (15), the light verb *zædæn* 'hit' enables the expression of a INITIATOR, which was impossible with *xordæn*. This leads to the conclusion that *zædæn* has the feature *init* and thus projects the necessary specifier position, while *xordæn* does not have the feature *init*, hence the non-agentive interpretation of CPrs with *xordæn*. In other words, I suggest that the two light verbs *zædæn* 'hit' and *xordæn* 'collide' have roughly the same abstract semantic content, expressing a (rather quick) change of state. They are also specified for the same syntactic features, modulo the feature $< init >$. When they appear with the same preverb, the difference in the meanings of the two complex predicates thus derived is due to the different underlying syntactic structures and the entailments they have for the interpretation of the predicate. More specifically, complex predicates with *zædæn* will have an external argument, most commonly a causer (in the case of a transitive predicate – cf. (15)), or a volitional agent (in the case of an intransitive predicate – cf. (16a)).

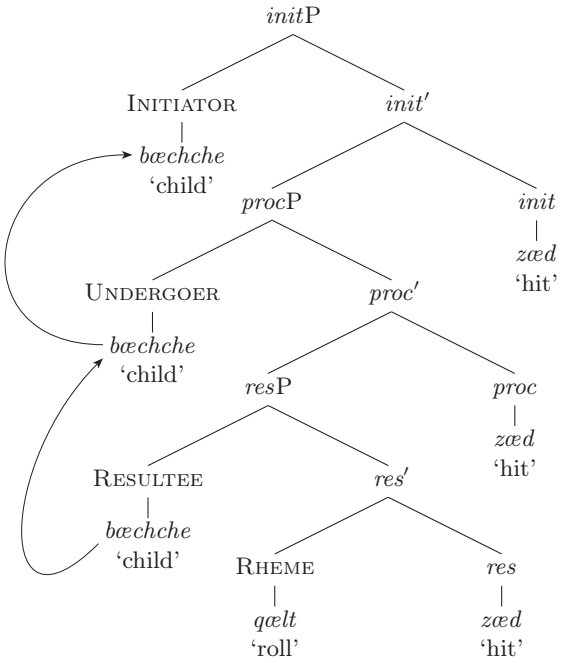
(16) Data from Samvelian (2004)

a. bæchche qælt zæd.
child roll hit
'The child rolled.' (intentionally)

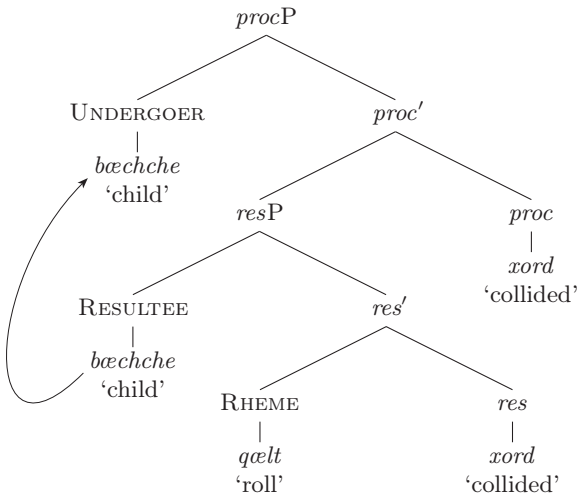
b. bæchche qælt xord.
child roll collided
'The child rolled.' (unintentionally)

The complex predicate formed by the verb *xordæn* will lack a causer/volitional agent because of the absence of the feature $< init >$ (see (12) and (16b)). The tree diagrams corresponding to the each of the sentences in (16) are presented in (17).

(17) a.



b.



Persian, in fact, provides a fairly systematic way to transform a complex predicate with no external argument to one with it by simply exchanging a light verb with no *init* for a light verb that can lexicalize *init* and therefore offers a Spec,*init*P position to be occupied by the Initiator.

This fact relates directly to Karimi-Doostan's (1997) classification of the dynamic light verbs in Persian into two groups called *initiatory* and *transition* light verbs, respectively. The former allow the expression of an Agent or Causer, while the latter do not. Translated into the terminology of the First Phase Syntax, the former lexicalize *init*, while the latter are not endowed with this feature. In Table 2, I present an overview of the most common light verbs with respect to the feature $< init >$. The verbs are arranged in the rows in such a way that they reflect the most often encountered alternating light verbs to form transitive-intransitive pairs.[8]

light verbs with $< init >$		light verbs without $< init >$	
zædæn	'hit'	*xordæn*	'collide'
kærdæn	'make'	*shodæn*	'become'
aværdæn	'bring'	*amædæn*	'come'
dadæn	'give'	*gereftæn*	'get'
ændaxtæn	'throw'	*oftadæn*	'fall'

table 2: Classification of Persian Light Verbs with respect to *init*

Given this pairing, the pattern of preverb distribution, as presented in Table 1 does not seem surprising. The table is repeated below as Table 3 and rearranged so that the symmetry of light verb+preverb combinations becomes clearer. Thus, if we regard the light verbs in the left column simply as the causative versions of the verbs in the right column, it is reasonable that they will combine with the same type of preverbs.

So, causativization of complex predicates in Persian consist of replacing the light verb of an inchoative light verb by its causative peer (the one specified for *init*), as further illustrated for

[8]The pairs given in Table 2 represent the most often encountered alternations. The relation between alternating light verbs is in reality many-to-many. For instance, the causative CPr *atish zædæn* (fire hit) 'set on fire' forms its inchoative counterpart by the verb *gereftæn* 'catch': *atish gereftæn* (fire catch) 'catch fire,' arguably because the light verb *gereftæn* has an additional meaning component of inception, which lacks in *xordæn*. Likewise, the inchoative light verb in the CPr *shekæst xordæn* (defeat collide) 'to be defeated' alternates with the causative light verb *dadæn* 'give': *shekæst dadæn* (defeat give) 'defeat' and the form **shekæst zædæn* (defeat hit) is ungrammatical. Another example of a verb that has more than one possible inchoative counterparts is the light verb *dadæn* 'give.' It alternates with the verbs *gereftæn* 'get', *ræftæn* 'go', and *shodæn* 'become'.

Light verb				N	P/PP	Adj/Adv
zædæn	'hit'	*xordæn*	'collide'	ok		
kærdæn	'make'	*shodæn*	'become'	ok		ok
aværdæn	'bring'	*amædæn*	'come'	ok	ok	ok
dadæn	'give'	*gereftæn*	'get'	ok	ok	
ændaxtæn	'throw'	*oftadæn*	'fall'	ok	ok	

TABLE 3: Preverb and light verb combinations (modified and repeated from Table 1)

amædæn–aværdæn 'come–bring' and *oftadæn–ændaxtæn* 'fall–throw' in (18) and (19), respectively (examples from Megerdoomian 2002b).

(18) a. ab be jush amæd
water to boil came
'The water boiled.'
b. nima ab-ro be jush aværd
Nima water-OM to boil brought
'Nima boiled the water.'

(19) a. homa be gerye oftad
Homa to crying fell
'Homa started to cry.'
b. nima homa-ro be gerye ændaxt
Nima Homa-OM to crying threw
'Nima made Homa (start to) cry.'

Further support comes from the passive. In the First Phase Syntax system, only verbs that project and identify *init* can passivize. In Persian, deriving passive from complex predicates is quite rare but it can be done with *init* light verbs, (20), while with *init*-less light verbs this leads to ungrammaticality, (21).

(20) a. reza xunæ-ro atish zæd.
Reza house-OM fire hit
'Reza set the house on fire.'
b. xune atish zæd-e shod.
house fire hit-PP PASS
'The house was set on fire.' (adapted from Mace 2003)

(21) a. xune atish gereft.
house fire caught
'The house caught fire.'
b. *xane atish gereft-e shod.
house fire catch-PP PASS

Summing up, the light verbs called "initiatory" by Karimi-Doostan can all be characterized by the presence of the feature $< init >$ in their specification. The "transition" light verbs lack this feature and lead to non-agentive complex predicates. Thus, the conclusion in this section is very much in line with the claim made in Folli et al. (2005) concerning the role of light verbs in complex predicates in determining agentivity.

4.2 Light verbs & res

Now that I have established that some light verbs have the feature $< init >$, while others lack it, in this subsection I will try to determine which light verbs are to be endowed by the feature $< res >$.

Since, in the First Phase Syntax, telicity arises as the result of complex interaction between different factors and, crucially, does not depend solely on the presence of a *res*P in the verbal decomposition, I will not make use of telicity tests in order to diagnose a *res*P. However, telicity is an important property of events and I will take up this discussion in Section 5.

The diagnostic I will be using in order to determine whether a certain light verb is endowed with $< res >$ is the availability of a punctual reading for a complex predicate which it is part of. Here, I follow Ramchand's (2008) suggestion that an event is punctual when a verb identifies both *proc* and *res*.

I will start out with an observation made by Megerdoomian (2002b) concerning different types of events expressed by the complex predicates. Consider the verbs in (22).

(22)

a.	dad zædæn	dad keshidæn	'to shout'
	cry hit	*cry pull*	
b.	næfæs zædæn	næfæs keshidæn	'to breathe'
	breath hit	*breath pull*	

Megerdoomian notes that the verbs in the first column have a punctual reading, whereas the verbs in the second column have a durative reading. Hence, the difference between *dad zædæn* and *dad keshidæn* is that the former denotes an event of one (sudden) uttering of a cry, while the latter denotes a prolonged production of a shout.[9] Similarly, for *næfæs zædæn* and *næfæs keshidæn*, the first one means roughly "to take a breath," while the second denotes a prolonged event of taking breath.

This distinction allows me to draw two conclusions. First, it is the light verb that carries the $< res >$ feature, since the noun in the pairs remains the same. Second, *zædæn* is endowed with it, whereas *keshidæn* lacks it.

However, this cannot be the whole story for *zædæn*, since complex predicates with this verb (shown in (23)) can give rise to durative (atelic) readings.

(23)

a. chækosh zædæn
   hammer hit
   'to hammer'

b. lægæd zædæn
   kick hit
   'to kick'

c. dad zædæn
   shout hit
   'to shout'

The behavior of the verbs in (23) very much resembles the behavior of semelfactives, which are punctual, on one hand, but systematically give rise to a durative (indefinitely iterated) reading, on the other. Since, this is presumably what happens with the verbs in (23) above, I believe that it is not incorrect to ascribe the $< res >$ feature to *zædæn*. I further believe that, just like all semelfactives in the First Phase Syntax, *zædæn* can be seen to be ambiguous between $< init, proc, res >$ and $< init, proc >$, in the former case, giving rise to punctual events and in the latter case – to durative events. Thus, I directly adopt the way semelfactives are treated by Ramchand, namely, as being specified in the lexicon as $< init, proc, (res) >$.[10]

[9] To help the reader understand the (untranslatable in English) distinction between the two Persian verbs, I provide a context where the different uses become clear: imagine children having a "shouting competition" with a prize for the one who can shout the loudest and the one who can shout the longest. In the first case, the verb *dad zædæn* will be used, in the second — the verb *dad keshidæn*, since here it is implied that the shout should last long time.

[10] A proposal along these lines is made by Megerdoomian (2005), who derives the different properties of *zædæn* by decomposing the predicate into different sets of primitive units of meaning.

4.3 Classes of light verbs

In this subsection, I present the lexical types of some of the light verbs in Persian. An important assumption is that the transitive-intransitive pairs, as shown in Table 2, differ only with respect to the availability of the *init* subevent. In other words, the feature specification of a verb from the left column will be identical, modulo $< init >$, to its peer in the right column.

In Table 4 below, I present my proposal regarding the feature specification of some of the light verbs in Persian.

kærdæn	'make'	$< init, proc, res >$	*shodæn*	'become'	$< proc, res >$
ændaxtæn	'throw'	$< init, proc, res >$	*oftadæn*	'fall'	$< proc, res >$
aværdæn	'bring'	$< init, proc, res >$	*amædæn*	'come'	$< proc, res >$
zædæn	'hit, strike'	$< init, proc, (res) >$	*xordæn*	'collide'	$< proc, (res) >$
dadæn	'give'	$< init, proc >$	*gereftæn*	'get'	$< proc >$
keshidæn	'pull'	$< init, proc >$	—		
kærdæn	'do'	$< init, proc >$	—		

TABLE 4: Light verb classes

A couple of comments are due here regarding the Table 4. First, the motivation for the different treatment of the light verbs *make*, *throw* and *bring*, on the one hand, and *hit*, on the other hand, lies in the fact that the first three are not semelfactive (but still resultative) verbs, while *hit* is semelfactive, as discussed above. Second, the fact that the light verb *kærdæn* is listed twice reflects its ambiguity between an activity verb, roughly corresponding to English *do* (24) and a causative verb *make* (cf. Megerdoomian 2001, Megerdoomian 2005). It is only in the latter meaning that *kærdæn* alternates with *shodæn*, as shown in (25).

(24) a. bæchche bazi kærd.
child game did
'The child played.'
b. *bazi shod.
game became

(25) a. reza mina-ro bidar kærd.
Reza Mina-OM awake made
'Reza woke up Mina.'
b. mina bidar shod.
Mina awake bicame
'Mina woke up.'

What is to be noted concerning the ambiguity of the verb *kærdæn* is that when it is a $< init, proc, res >$ verb, there are two distinct argument: an INITIATOR and an UNDERGOER-RESULTEE, as in (25). When *kærdæn* is a $< init, proc >$ verb, there is one single argument carrying the composite role of INITIATOR-UNDERGOER, as in (24). In this latter case *kærdæn* is what is traditionally called an unergative verb and it is not suprising that it does not have an inchoative counterpart. Interestingly, the two varieties of *kærdæn* also appear in combination with different preverbs – the unergative one takes eventive nouns, while the causative one takes adjectival preverbs and non-eventive nouns. For this reason, I do not list *kærdæn* as an $< init, proc, (res) >$ verb, as it is clearly different from *zædæn*, which appears with the same preverb, no matter whether it is $< init, proc >$ or $< init, proc, res >$.

I now turn to the light verb *keshidæn* 'pull', which is marked in Table 4 as having no inchoative peer. As already discussed in Section 4.2, the light verb *keshidæn* contributes duration to the complex predicates it participates in. It is similar to the unergative *kærdæn* 'do,' as it often forms intransitive

complex predicates (e.g. *tul keshidæn* (length pull) 'take a long time,' *chopoq keshidæn* (pipe pull) 'smoke a pipe'). *Keshidæn* can also form transitive complex predicates, like *færahæm keshidæn* (together pull) 'assemble,' or *jaru keshidæn* (broom pull) 'sweep.' Even in such cases, though, *keshidæn* does not seem to have an inchoative counterpart, that is, a corresponding light verb specified for the feature $< proc >$ and with the same abstract semantic content. It is true that there exist complex predicates like *færahæm shodæn* (together become) 'be assembled,' and *jaru xordæn* (broom collide) 'be swept,' but these CPrs are more likely to be the counterpart of *færahæm kærdæn* (together make) 'assemble,' and *jaru zædæn* (broom hit) 'sweep,' as they have the same aspectual properties and, crucially, lack the durative component of *keshidæn*.

5 Deriving telicity

In this section, I will outline how the temporal (un)boundedness of the macro-event can be accounted for by using the tools made available by the system. I will apply the *in an hour/for an hour*-test to diagnose telic and atelic predicates, respectively.

In Persian, there exist numerous ways to form the corresponding temporal phrases and sometimes speaker vary with respect to their interpretation.[11] To avoid confusion, I will use the expression *dær yek sa'æt* 'in one hour' and *bemodæte yek sa'æt* 'for one hour'[12] to diagnose telic and atelic sentences, respectively.

5.1 Rhematic material

As already mentioned in Section 2, the boundedness of the macro-event does not necessarily arise from the presence of *res* in the subevent decomposition of the VP. A telic interpretation can be the result of an $< init,\ proc >$ verb combining with a Rheme complement that is a bounded path PP, a closed scale adjective, or a quantized NP (in the sense of Kennedy and Levin 2008). I argued in Section 3 that the preverb in a complex predicate occupies the Rheme position. Therefore, the system predicts that the preverb will have impact on the telic/atelic interpretation of the complex predicate. The prediction is borne out, as illustrated in the data set below, where the light verb is the same but the interpretation nevertheless differs. When the $< init, proc >$ verb *kærdæn* 'do' (noted to lack $< res >$ when combining with a noun preverb) combines with a non-quantized nominal preverb, the predicate is atelic (26a). If we exchange the preverb for a quantized noun, the predicate becomes telic (26b).

(26) a. bæchche bemodæte /*dær yek sa'æt gerye kærd.
child for / in one hour crying did
'The child cried for an hour /*in an hour.' (atelic)

b. bæchche dær /*bemodæte yek sa'æt hæme-ye geryæ-sh-ro kærd.
child in / for one hour all-ez crying-3cl-om did
'The child did all its crying in an hour /*for an hour.' (telic)

Megerdoomian (2005) presents some data which offer convincing evidence that an analysis like the one argued for in this paper might be on the right track. She discusses complex predicates which give rise to telic/atelic readings depending on the noun preverb. A sample of these verbs is presented in Table 5 below.

Folli et al. (2005) discuss this set of data and suggest that the reason the complex predicates in the first column are telic is that the noun element is bounded. The noun preverb in the second column is unbounded and therefore gives rise to an atelic reading. I will adopt this proposal without further discussion, since it is perfectly compatible with the First Phase Syntax and the facts are

[11] For example, as pointed out by Karimi-Doostan (1997), for some speakers the non-durative adverbial *zærfe yek sa'æt* 'in one hour' has a durative meaning when stressed.

[12] The expression *bemodæte yek sa'æt*, roughly translated as 'in the course of one hour,' belongs to the formal style. A much more common way to convey the same meaning is to drop the preposition. *yek sa'æt* expresses the same notion of 'for one hour.'

Telic		Atelic	
æfsar zædæn harness hit	'to harness'	*næmæk zædæn* salt hit	'to put salt'
palan zædæn blanket hit	'to saddle'	*rouqæn zædæn* oil hit	'to oil'
zæng zædæn bell hit	'to ring'	*gærd zædæn* powder hit	'to powder'

TABLE 5

exactly what the system predicts. Below, I briefly summarize the properties of *zædæn* with respect to the different event types it can give rise to, when it is an $< init, proc >$ verb.

(27) *zædæn* as $< init, proc >$
  a. telic - *palan zædæn* 'to saddle' (when the RHEME is bounded)
  b. atelic - *rouqæn zædæn* 'to oil' (when the RHEME is unbounded)

Since the intransitive counterpart of *zædæn* is *xordæn* 'collide' and I assumed that it has the same categorial specification as *zædæn* without the $< init >$ feature, it is expected that complex predicates with *xordæn* will allow atelic readings of the type in (27b). This is the case with the complex predicate *qosse xordæn* (worry collide) 'to worry,' which is atelic according to Megerdoomian (2006). Another example comes from Megerdoomian (2002a):

(28) mærdom sal-ha æz dowlæt færib xord-ænd.
  people year-PL from government fool ate-3PL
  'People have been fooled by the government for years.'

More data illustrating the fact that telicity can be due to *proc* light verbs with bounded RHEME preverbs are shown below:

(29) a. mehmani do sa'æt tul keshid.
  party two hour length pulled
  'The party lasted for two hours.' (atelic)
  b. reza xane-ro dær yek sa'æt be atish keshid.
  Reza house-OM in one hour to fire pulled
  'Reza set the house on fire in one hour.'
  (Bounded TO path → telic)

Here, we have the *proc* light verb *keshidæn* 'pull', which combines with an unbounded noun *tul* 'length' and forms an atelic predicate. When *keshidæn* appears with a bounded preverb, like in the case of the bounded PP *be atish* 'to the fire' in (29b), the entire complex predicate is telic.

5.2 More remarks on telicity

Folli et al. (2005) discuss various important issues concerning Persian complex predicates. One of the conclusions they reach is that while the light verb determines the agentivity/causativity, the eventiveness and duration of the CPr, the preveb determines the Aktionsart of eventive CPrs. In other wors, whether the CPr will be telic or atelic depends entirely on the preverb. The event structure they propose for Persian is presented in Table 6, where, they argue, there is no relation between the boundedness of the event and the light verb.

There are two observations to be made concerning the table above. First, a complex predicate with a noun as a preverb can be either telic or atelic. This is captured by the system and discussed in the beginning of this section, so it does not come as a surprise. Second, according to this table, it

preverb	telic	atelic
noun	ok (if eventive)	ok
Adj/Adv	ok	*
P/PP	ok	*

table 6: Folli et al. (2005)

is never the case that a complex predicate with an adjectival, adverbial, prepositional or PP preverb is atelic. If we now go back to Table 1 in Section 4.3, and have a look at which light verbs combine with the aforementioned preverbs, it turns out they are mainly $< res >$ verbs, with two exceptions. Therefore, the natural interpretation of these predicates is a bounded one, because $< res >$ verbs by default lead to telic predicates, no matter the Rheme (i.e., the preverb). The facts are repeated in Table 7 below.

light verb			P/PP	Adj/Adv
aværdæn	'bring'	$< init, proc, res >$	ok	ok
keshidæn	'pull'	$< init, proc >$	ok	ok
amædæn	'come'	$< proc, res >$	ok	ok
gereftæn	'take'	$< proc, res >$	ok	
oftadæn	'fall'	$< proc, res >$	ok	
ændaxtæn	'throw'	$< init, proc, res >$	ok	
dadæn	'give'	$< init, proc >$	ok	
kærdæn	'make'	$< init, proc, res >$		ok
shodæn	'become'	$< proc, res >$		ok

table 7

Let us now examine the cases when a *proc* verb combines with preverbs which are not nouns (*dadæn* 'give' and *keshidæn* 'pull'). In the system adopted in this paper, whenever the Rheme of a *proc* verb is bounded/closed scale, the predicate will be interpreted as telic. If the Rheme is unbounded/open scale, the event will be atelic. Applied to adjectival Rhemes, whenever a *proc* light verb combines with gradable, closed scale adjectival preverb in the sense of Kennedy and Levin (2008), the interpretation should be telic and whenever a *proc* light verb combines with a gradable, open scale adjectival preverb, the interpretation should come out as atelic. Thus, the system predicts that there can exist complex predicates with a *proc* light verb and an adjectival preverb that are atelic. The prediction is borne out, as shown by the sentence below with the complex predicate *deraz keshidæn* (long pull) 'to take a nap.'

(30) madær yek sa'æt deraz keshid.
mother one hour long pulled
'Mother had a nap for one hour.'

In other words, the First Phase Syntax model correctly captures the telicity facts. To a certain extent the way telicity is accounted for in the present paper and in Folli et al. (2005) overlaps in the sense that under both approaches the preverb has a role to play in detemining the boundedness of the event. However, I disagree that telicity depends exclusively on the type of the preverb.

6 Conclusion

In this paper, I presented an analysis of Persian complex predicates in the framework of the verbal First Phase Syntax, as developed in Ramchand (2008). I suggested that the subevent heads are lexicalized by the light verb and that the preverbal material occupies the rhematic position and semantically unifies with the light verb to build one joint predication. Under this account, the light verb plays a role in determining the argument structure of the entire predicate in that it projects the specifier positions where we find the participants in the event. The preverb can indirectly affect argument structure by determining how high a DP can raise from one specifier to another and thus what composite role a DP can have.

I examined some of the most productive light verbs and proposed a feature specification for them. Thus, I divided the light verbs in classes according to their feature specification. I also showed how each of the two components of the complex predicate affects the boundedness of the macro-event. Namely, light verbs with *res* feature participate in bounded complex predicates. But also *proc* light verbs can be bounded, as the preverb in the RHEME induces a telic reading when it is bounded.

Acknowledgments

I wish to thank two anonymous reviewers for their insightful comments and incisive questions. I am also grateful to Peter Svenonius and Pavel Caha. This work has profited a lot from the discussions I had with them. Finally, I thank Parvaneh Danesh for her native speaker judgments.

N	PP/P	Adj/Adv	N	PP/P	Adj/Adv
$< init, proc, res >$			$< proc, res >$		
lule kærdæn		baz kærdæn	lule shodæn		baz shodæn
tube make		*open make*	*tube become*		*open become*
'roll up' (tr.)		'open' (tr.)	'roll up' (intr.)		'get opened'
churuk ændaxtæn	æz pa ændaxtæn		churuk oftadæn	æz pa oftadæn	
wrinkle throw	*from foot throw*		*wrinkle fall*	*from foot fall*	
'wrinkle' (tr.)	'wear out'		'get wrinkled'	'run out of energy'	
yad aværdæn	be donya aværdæn	gærd aværdæn	yad amædæn	be donya amædæn	gærd amædæn
memory bring	*to world bring*	*round bring*	*memory come*	*to world come*	*round come*
'remind'	'give birth'	'assemble' (tr.)	'recall'	'be born'	'assemble' (intr.)
$< init, proc, (res) >$			$< proc, (res) >$		
gereh zædæn			gereh xordæn		
know hit			*knot collide*		
'tie in a knot'			'get tied in a knot'		
$< init, proc >$			$< proc >$		
bu dadæn	æz dæst dadæn		bu gereftæn	æz dæst ræftæn	
smell give	*from hand give*		*smell get*	*from hand go*	
'emanate a smell'	'lose'		'become smelly'	'be lost"	
qæd keshidæn		deraz keshidæn		no inchoative counterpart	
size pull		*long pull*			
'grow taller'		'take a nap'			
fekr kærdæn				no inchoative counterpart	
thought do					
'think'					

TABLE 8: Appendix: Examples of complex predicates discussed in the charts

References

Browning, Marguerite and Ezat Karimi. 1994. Scrambling to object position in Persian. In N. Corver and H. van Riemsdijk, eds., *Studies on Scrambling*, pages 61–100. Berlin: Mouton de Gruyter.

Butt, Miriam. 2003. The light verb jungle. In *Harvard Working Papers in Linguistics*, vol. 9, pages 1–49. Cambridge, MA: Harvard.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36(3):315–332.

Family, Neiloufar. 2006. *Exploration of Semantic Space: The Case of Light Verb Constructions in Persian*. Ph.D. thesis, École des Haute Études en Sciences Sociales.

Folli, Raffaella, Heidi Harley, and Simin Karimi. 2005. Determinants of event type in Persian complex predicates. *Lingua* 115(10):1365–1401.

Karimi, Simin. 2005. *Minimalist Approach to Scrambling*. Berlin: Mouton de Gruyter.

Karimi-Doostan, Gholamhossein. 1997. *Light Verb Constructions in Persian*. Ph.D. thesis, Essex University.

Karimi-Doostan, Gholamhossein. 2005. Light Verbs and Structural Case. *Lingua* 115:1737–1756.

Kennedy, Christopher and Beth Levin. 2008. Measure of change: The adjectival core of degree achievements. In L. MacNally and C. Kennedy, eds., *Adjectives and Adverbs: Syntax, Semantics, and Discourse*, pages 156–182. Oxford University Press.

Lazard, Gilbert. 1957. *Grammaire du Persan contemporain*. Paris: Librairie C. Klincksiek.

Mace, John. 2003. *Persian Grammar for Reference and Revision*. London: RoutledgeCurzon.

Megerdoomian, Karine. 2001. Event Structure and Complex Predicates in Persian. *Canadian Journal of Linguistics: Special Issue on Iranian Languages* 46(1/2):97–125.

Megerdoomian, Karine. 2002a. Aspect in Complex Predicates. Talk presented at the Workshop on Complex Predicates, Particles and Subevents, Konstanz, October 2002.

Megerdoomian, Karine. 2002b. *Beyond Words and Phrases: A Unified Theory of Predicate Composition*. Ph.D. thesis, University of Southern California.

Megerdoomian, Karine. 2005. Typology of Complex Predicates in Persian. Talk presented at the Conference of Iranian Linguistics, Leipzig, June 2005.

Megerdoomian, Karine. 2006. The Status of the Nominal in Persian Complex Predicates. available at http://www.zoorna.org/papers/preverbs.pdf.

Pantcheva, Marina. 2008. Noun preverbs in Persian complex predicates. In P. Svenonius and I. Tolskaya, eds., *Tromsø Working Papers on Language and Linguistics: Nordlyd 35,* Special issue on Complex Predication, pages 19–45. Tromsø: University of Tromsø. Available at http://www.ub.uit.no/baser/nordlyd/.

Ramchand, Gillian. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge: Cambridge University Press.

Samvelian, Pollet. 2001. Le statut syntaxique des objets "nus" en persan. *Bulletin de la Société de linguistique de Paris* XCVI(1):349–388.

Samvelian, Pollet. 2004. Le sujet, l'objet et l'inaccusativité dans les prédicats complexes nom-verbe en persan. Ms. Université de Paris 3, to appear in Cahiers de Linguistique de l'INALCO 2003/5.

Structure of Verbs in Malto

Chaithra Puttaswamy, *The School of Oriental and African Studies*

Received February, 2009; Revised October, 2009

Abstract

Malto is a North Dravidian language spoken in Eastern India. It is an agglutinating language with SOV word order and suffixing morphology. The finite verb word in Malto maximally carries information about valence adjusting operations, tense-aspect-mood, negation and gender-number-person agreement with the subject. The non-finite verbs take suffixes marking adverbialisation, complementation, relativisation, participialisation and relative tense. Syntactically, there is only one finite verb in a sentence and all the other verbs preceding it are non-finite. This paper is a descriptive analysis of the structure of Malto verbs and an outcome of a language documentation project with the intention of describing the formal structure of the Suariya Pahariya variety of Malto. This work is a follow up on grammatical accounts on Malto by Doerse (1884), Das (1973) and Mahapatra (1979).

1 Introduction

Malto is a North Dravidian language which has been poorly documented. There are about 108,000 Malto speakers living on the Rajmahal Hills in eastern India. Their language is endangered due to limited scope of use under pressure from Hindi and Santali, the dominant languages of the region and temporary and permanent migration of Malto speakers from their villages to towns in pursuit of employment and education. The language is often referred to as Pahariya since that is how the people of the Malto speaking community are recognised in the region. Malto speakers are trilingual in Malto, Hindi and Santali, often using Malto only within the confines of their villages. Since Malto speakers wish to be a part of the mainstream economic lifestyle of modern India, propagation and preservation of their language has not been their primary concern. Malto has no indigenous tradition of writing and hence no script has been associated with the language. Printed publications in Malto have been produced by religious and non-governmental organisations involved in activities related to community development. Malto is not used in the formal education system either as a medium of instruction or as a language of study.

The north Dravidian branch, to which Malto belongs, is geographically isolated from the rest of the language family and hence scholars speculate that Malto may have retained some proto-Dravidian forms. This paper is a descriptive analysis of the structure of Malto verbs. The first section in this paper introduces the minimal unit that can be considered as a verb in Malto and then discusses the formal structure of the verb (Section 2). The structure of finite verbs in Malto and the three levels of verb word formation are discussed in detail in Section 3. Non-finite verbal suffixes (Section 4) and the various functions associated with them, namely conditional (Section 4.1), causal adverbial (Section 4.2), relative past (Section 4.3), simultaneity marking (Section 4.4), conjunct participle (Section 4.5), infinitive (Section 4.6) and adnominal (Section 4.7) are explained in the third part of this paper.

JSAL Volume 2, Issue 1, December 2009.

The fourth part of this paper is dedicated to a discussion of category changing derivational processes (Section 5) involving verb roots (Section 5.1) and verb stems are derived from nouns and borrowed roots (Section 5.2).

2 Formal Structure of the Verb

The Malto verb word minimally consists of a verb stem. A stem is a form from which a word is derived by the addition of one or more affixes. A verb stem in Malto can be obtained by the addition of a stem formative suffix (Section 2.1.1) to the verb root or by the addition of a derivational or inflectional suffix. A verb root is a form from which words or parts of words are derived. A root is not itself derivable from any smaller or simpler form. All verb roots in Malto are bound forms that either undergo stem alternations in case of the past tense forms or take a suffix. However, the Malto verb in its stem form is restricted to compound verb constructions and cannot appear as the head of an independent clause. In order to appear as the head of an independent clause a verb stem in Malto has to take at least one suffix from a set of inflectional suffixes (Section 2.1.3) that includes negative suffixes, tense/mood suffixes and gender-number-person agreement suffixes. Verb stems can also appear as the head of a dependent clause by taking at least one suffix from a set of non-finite verbal suffixes (Section 4). Verbs that appear as the head of an independent clause are called finite verbs (Section 3) and verb words that appear as the head of a dependent clause are called non-finite verbs. The following example shows a verb stem that appears with another verb word to form a compound verb. This compound verb is a non-finite verb which is the head of a dependent clause and depends on a third verb word which is the head of an independent clause.[1]

(1) [os ṯunheki] hec-a:ḏ
cut.Pst collect-Ep-Rp-3Sg.Nm tie-3Sg.F
'Having cut and collected (the bamboo), she tied them up.' Story C2

In the above example *os ṯunheki* is a compound verb word where *os* is the past stem of the verb *oy* 'cut'. The second nucleus of the compound *ṯunh*, carries the relative past tense marker (Section 4.3) which has scope over the whole compound word. The finite verb *heca:ḏ* consists of the verb root *hec* 'tie' and the gender-number-person agreement suffix.

2.1 Levels of Verb Word Formation

Malto is an agglutinating language and verbs are formed by adding suffixes to the verb root with little morphophonemic change. Malto verb word formation maximally takes place at three levels. The first level is the level of stem formation (Section 2.1.1). It is possible for the verb root to act as a stem or for a stem to be derived by attaching a stem formative suffix to the verb root. The verb stem can act as a meaningful syntactic unit as the first nucleus of a compound verb. The second level of verb word formation is the concatenation of derivational suffixes to the verb stem (Section 2.1.2). It is a typological characteristic of derivational suffixes that they occur closer to the verb root than inflectional suffixes (Bybee 1985). In case of a finite verb, once the verb stem takes a derivational suffix, it is obligatory that it is followed by an inflectional suffix. The third and final level of verb word formation is the concatenation of inflectional suffixes (Section 2.1.3). Steever (1993, 12) compares verbal inflections and derivations by stating that:

> "inflection differs from derivation in that the members of an inflectional opposition are mutually implicating so that, for example, the existence of a past tense in a grammatical system always implies the existence of a non-past tense, and viceversa. Derivation, on the other hand, creates an opposition of two terms, a base and a derived form, whose members are not mutually implicating: while a derived form always implies the existence of a base form, forms that might otherwise serve as base forms need not imply the existence of a derived form."

[1] Every example in this paper carries a tag at the right end of the free translation line. This tag indicates the file name of the annotated text available at the Endangered Languages Archive, School of Oriental and African Studies.

One way of understanding the structure and function of verbs in Malto is to locate them in the larger perspective of Dravidian verbs and analyse how they conform to or differ from the typical features of the language family. Verb roots in Dravidian languages are known to be monosyllabic with the canonical shape (C) V (C) (Krishnamurti 2003) as shown in the following Malto examples.

ek	'go'
baɾ	'come'
ok	'sit'
men	'be'

TABLE 1

However, some verb roots in Malto have more than one syllable. They may have been monosyllabic historically and retained some suffix that is no longer productive.

t̪unh	'collect'
caḍg	'slip'
muluh	'drown'
cuɖup	'drop'

TABLE 2

The verb stem formation process is discussed in the following section.

2.1.1 Level One: Verb Stem Formation

The first level of verb word formation in Malto involves the addition of the stem formative suffix to a verb root. There are two types of stem formative suffixes in Malto. The first type of stem formatives create a stem that is the past tense alternate of the non-past verb root. The other kind of stem formative suffixes are the tense-transitivising suffixes. Krishnamurti (2003, 278) points out for Dravidian in general that "no meaning can be assigned to the formative suffixes. It is speculated that they represented tense and voice markers at an early stage of Proto-Dravidian and were already losing that significance within Proto-Dravidian in different subgroups" (also see Cladwell 1956). Although both the stem formative suffixes are synchronically non-productive, it is a signification stage in the process of word formation as it explains the verb stem alternations in the case of past stem formation and recurrent phonological endings of some transitive verb stems that deviate from the canonical shape of the Dravidian verb.

Based on the typical phonological structure for Dravidian verbs, which is (C)V(C), it can be deduced that the factor that explains the presence of formative suffixes in Malto is that some of the formatives are relics of Proto-Dravidian inflected verb forms. The verb roots taking past stem formatives are paired with non-past alternates. In contrast, the verb roots with the transitivising NP (nasal + plosive) formatives in Malto do not always have an intransitive correspondent and in such instances the bare root without the stem formative is no longer a meaningful unit. Hence the verb stems with the Proto-Dravidian-NP formatives are derived bases that are now part of the lexicon in Malto. However, there is a productive derivational process in the language that is explained in Section 2.1.2. Krishnamurti (2003, 182) has postulated that:

> "at a very early stage within Proto-Dravidian, sonorant suffixes of the L type (l, ɭ, z̪, r, w, y) were added to (C)V:- or (C)VC-V-stems to form extended intransitive/middle voice stems. This assumption is based on the observation that verb stems ending in sonorant suffixes tend to be intransitive in the descendent languages. At a later period, -L, -VL lost their identity as grammatical elements and became incorporated into the preceding stems. The P-suffixes signal both tense and voice."

The following Table shows how the tense and transitivity properties combine to form stem formatives in Proto-Dravidian, where the dental vs. non-dental distinction indicates past vs. non-past;

simple (N)P signals intransitive, and geminate (N)PP, transitive:

	Non-Past		Past
Intransitive	**p*	**k*	**t*
	**mp*	**nk*	**nt*
Transitive	**pp*	**kk*	**tt*
	**mpp*	**nkk*	**ntt*

TABLE 3

Based on the diachronic data on Dravidian languages presented by Krishnamurti (2003) and Subrahmanyam (1971), the following non-productive stem formatives can be reconstructed for Malto. The labial series of non-past stem formatives is missing in Malto. In all instances, Malto has replaced the geminates in the proto-form of the suffix by voiced plosives.

i) The /-d̪/ suffix is a weakened form of the proto-Dravidian transitive-causative suffix **-tt*. This suffix always attaches to a root ending in /n/.

(2) *on* 'drink' *ond̪* 'cause to drink'
pun 'wear' *pund̪* 'cause to wear'

The above example shows intransitive verbs that become derived transitive-causative verbs with the addition of the /-d̪/ suffix. However, not all transitive verbs with /-d̪/ formative suffix have intransitive counterparts.

(3) *hon-d̪* 'fetch, bring'
man-d̪ 'bury, plant'
men-d̪ 'burn'
mun-d̪ 'wrap'
pun-d̪ 'put'
nin-d̪ 'fill'

ii) The /-j/ suffix is a weakened form of the proto-Dravidian transitive causative suffix **-cc*∼**-kk*. These verbs do not have intransitive counterparts.

(4) *am-j* 'talk'
an-j 'fruit'
kun-j 'throw, give birth'
con-j 'fasten, bind'
cun-j 'pound'

iii) The /-g/ suffix is the transitive form of the proto-Dravidian paired intransitive and transitive stem with -(N)P/-(N)PP. Synchronically, all the verbs taking this suffix express telic transitive events, but not all of them are punctual.

(5) *haɾ-g* 'climb'
t̪is-g 'open'
t̪iɾ-g 'press, apply force'
is-g 'crack open'
aɖ-g 'press'
as-g 'shear'
bas-g 'peel'
nus-g 'rub off (scales of a fish)'
nuɾ-g 'drag, slide'
nun-g 'swallow'

puḏ-g ‘pluck’
kuɖ-g ‘bend’

Malto has lost the intransitive member of the pair for the above stems and instead uses the productive intransitive suffix /*-ɾ*/ to derive intransitive stems.

(6) *aɖg-ɾ-aːḏ* ‘It was pressed.’
caḏg-ɾ-aːḏ ‘It slipped.’

The productive process of stem formation is explained in the following section.

2.1.2 Level Two: Derivational Suffixes

The second level of verb word formation is the addition of derivational suffixes. A derivational suffix is not obligatory in forming either a finite or a non-finite verb word. The addition of a derivational suffix changes the argument potential of the verb and creates an idiosyncratic meaning for the resulting verb base. A Malto verb word can include a sequence of two derivational suffixes at most. The first, which is the closest to the verb root, is the suffix that determines the overall transitivity of the verb word. Alternatively, this slot can be occupied by the verbalising suffix that derives verbs from nouns (Section 5.2). Words which can be verbalised are either Malto noun roots or borrowed stems from Indo-Aryan languages. The second derivational suffix in the sequence can be chosen from a set of valence changing operators that include the causative, reciprocal and the passive suffixes.

The following Table shows the productive derivational suffixes in Malto.

Detransitivising/reflexive	*-ɾ*
Causative	*-ṯr* / *-ṯar*
Reciprocal	*-nah*
Passive	*-uhɾ*

TABLE 4

Detransitivisation

A transitive verb is changed to an intransitive by the addition of the suffix -ɾ. Krishnamurti (2003) observes that this maybe a relic of Proto-Dravidian forms since such a strategy is not found in the neighbouring Indo-Aryan and Austro-Asiatic languages and adds that “most verbs ending in formative -(V)l/-(V)ɾ in South Dravidian I and South Dravidian II tend to be intransitive.” (Krishnamurti 2003, 279). The following examples from Malto, where deriving transitive verbs by a process of suffixation is still productive, support Krishnamurti’s hypothesis about transitive verbs in Dravidian languages which was based on his observations from South Dravidian languages.

(7) a. *ṯes-po* ‘winding, kneading’
tes-ɾ-po ‘binding’

b. *ṯaɖic-po* ‘slapping’
ṯaɖic-ɾ-po ‘clapping’

c. eːn nam-i-ṯ-aːn
1Sg.Nom scold-Ep-Pst-1Sg
‘I scolded.’
eːn nam-ɾ-aːṯaːn
1Sg.Nom scold-Dtr-Ep-Pst-1Sg
‘I was scolded.’

From the above examples I have deduced that the morphological process of detransitivisation applies to all predicates that have a valence more than one.

Valence adjusting operators

Steever (1993) has proposed the concept of Compound Verb Contraction in Dravidian languages. Those syntactic constructions that were compound verbs with the V2 of the compound expressing change in valence diachronically have contracted to become simple verbs synchronically with the V2 of the compound being reduced to a suffix. In other words, the stem combined with the inflected form of the auxiliary verb to form a single verb word. The causative, reciprocal and passive suffixes in Malto are derived by this process. A diachronic account of each of these suffixes along with a reference to the corresponding entries in the Dravidian etymological dictionary is provided in the relevant sections.

It is a typological feature of agglutinative languages that they have fewer lexical causatives and that causation is manifested morphologically by attaching an affix to the verb base (Dixon 2000). Causation in Malto is productively realised by the following suffixes: */-ʈar, ʈr/*. The causative suffix is possibly an archaic form of what is synchronically the lexical verb form *tara* 'give to 1st or 2nd person recipient' in Tamil/South Dravidian. The Proto-Dravidian form of this verb is **ta/tar* [DED 3098]. Winfield (1928) has recognised this morpheme as the 'transition particle' in Kui (Central Dravidian) and Israel (1979) calls it the 'personal object' suffix for the sister language Kuvi, both of which point to the fact that **ta/tar* is used as a valence increasing morpheme in two other lesser known Dravidian languages. The causative suffix attaches itself to the intransitive verbal stem and makes it a transitive verb. Hopper and Thompson (1980) have observed that this association of causitivity and transitivity is a universal phenomenon. The tense-aspect-mood marker and then the agreement marker follow the causative suffix. The causative suffix does not affect the shape of the verb root.

/-ɖ/ is a relic of the Proto-Dravidian causative-transitive suffix and is no longer productive in Malto but for a few exceptions such as *oːn* 'drink', *oːnɖ* 'cause to drink/serve a drink' (Krishnamurti 2003, 280). The */-ɖ/* suffix acts as a causativiser in example (8a) and as transitiviser in example (8b) where it is then detransitivised by the addition of the suffix */-ɾ/*.

(8) a. hani ʈeho-ɖ man-ɖ-aːɖ aɾo cila-ɖ aːɾ giɖɖa-ɖ
then mother-Nom.Nm bury-Caus-3Sg.Nm and eagle-Nom.Nm and vulture-Nom.Nm
men-ɖ-aːɖ
burn-Caus-3Sg.Nm
'Then, the mother buried (one half) while the vulture and the eagle burnt (the other half).'
Story C2

b. ʈaːni peh-ɾ-a hon-ɖ-ɾ-aːh
self hold-Dtr-Cp bring-Sf-Dtr-3Sg.M
'He himself brought (them).' History

The following examples illustrate how a one place predicate in example (9) is converted to a two place predicate in example (10) by the addition of a causative suffix.

(9) hani haː maa giɖraː-ɖ kʰajjaːk ay-aːɖ
then Dem.Dst Clf fox-Nom.Nm lot dry-3Sg.Nm
'Then, the fox weakened a lot.' Story C4

(10) eːn ʈaʈaha-n aːy-ʈaɾ-iːn
1Sg.Nom mango-Acc dry-Caus-1Sg
'I dried mangos.' Elicitation

The impact of the reciprocal situation is equal on all the participants of the situation and hence there is no hierarchy among the participants. It is due to this phenomenon that the valence of a reciprocal situation is reduced. The reciprocal suffix in Malto is */nah, naʔ/*. It is interesting to note that the word *naqe* 'to act or be to one another' [DED 3571] exists only in the two North Dravidian

languages Malto and Kurukh. This form resembles the Hindi nouns *nakal* 'copy' and *nakli* 'duplicate' and hence I suggest that the reciprocal suffix may be a borrowed form. The following examples show how the reciprocal event is expressed by the addition of the */nah, naʔ/* suffix.

(11) giḍraːḍ havḍ-aːh
fox-Nom.Nm speak-3Sg.M
'The fox spoke.' Story C4

(12) hani eːr ma.a-ḍ havḍ-ɾ-naʔ-iːy-aːḍ
then hen children-Nom.Nm speak-Dtr-Recp-Pst-3Sg.Nom
'Then the chicks discussed (with each other).' Story C4

The transitive verb is detransitivised by the addition of the detransitivising suffix */-ɾ/* and then the reciprocal suffix is added to the verb stem as shown in the following examples.

(13) eːm-u manḍraːn aːḍ-a-ṭ-aːm
1Pl.Nom-En medicine-Acc apply-Ep-Pst-1Pl
'We applied medicine.' Elicitation

(14) eːm-u manḍraːn aːḍ-ɾ-naʔ-ṭ-aːm
1Pl.Nom-En medicine-Acc apply-Dtr-Recp-Pst-1Pl
'We applied medicine to each other.' Elicitation

Historically passivisation in Malto, just as in other Dravidian languages, might have been realised as an explicator compound. The morphological passive in Malto is marked by the suffix */-uhɾ, -hɾ/*. This suffix is derived from the verb base *uɾqe* 'come out, come forth' [DED 668] by the process of compound verb contraction.

(15) eḍ-du gaːḍiye-no mand-hɾ-aːd
leg-Nom.Nm sludge-Loc bury-Pass-3Sg.Nm
'The foot got buried in the sludge.' Elicitation

2.1.3 Level Three: Inflectional Suffixes

The third level of verb word formation is the suffixation of inflectional affixes. A Malto verb word can include up to three inflectional suffixes chosen from:

1. a set of negative suffixes
2. tense/mood suffixes
3. gender-number-person (GNP) agreement suffixes

Alternatively an affix from the set of non-finite suffixes can take the place of a tense/mood suffix. Non-finite suffixes are discussed in detail in Section 4 of this paper.

Negation

Negation in Malto is post-verbal. There are two forms that express negation in Malto: the negative verb mala (example 16a) and the negative suffix */-la/* (example 16b). */-le/* is an allomorph of the negative morpheme */-la/* and their distribution is governed by the vowel harmony rules of the language. The concatenation of negative morphemes is also governed by the vowel sandhi rules. These two negative forms undergo some modifications such as appearing with epenthetic vowel */o/* to express negation of existentials and imperatives as shown in example (16c).

(16) a. paːɾe inoɾ aː saba mala
but now Dem.Dst case Neg
'But that is not the case now.' Village

b. haːh-in ṭunḍ-k-iːḍ ḍin-su jaːgu-n lap-laː-iːḍ
3Sg.M-Acc see-Rp-3Sg.Nm day-two food-Acc eat-Neg-3Sg.Nm
'Having seen him, she did not have food for two days.' Story C3

c. holh-oma:
cry-Neg
'Don't cry.' Story C3

Tense/Mood

The prevalent strategy among Dravidian languages is to have verb stem alternation in the past tense. Subrahmanayam (1971) points out that this is one of the oldest morphological constructions in Dravidian languages. However Krishnamurti (2003) notes that not all the past stem allomorphs of the proto-language are fully recoverable since some of them have been analogically regularised by the daughter languages. The following Table maps the proto-Dravidian past stem allomorphs to the corresponding manifestations in Malto.

Proto-Dravidian Past Tense Allomorph	Malto Non-Past-Stem	Malto-Past-Stem
**-t-*	*o:n* 'drink'	*o:n-ḍ*
**-i-*	*hek* 'go'	*hek-iy*
**-cc-*	*men* 'happen'	*men-j*
	baɾ 'come'	*baɾ-c*
	key 'die'	*ke-c*
	oy 'cut'	*o-s*
	hil 'stand'	*hi-j*

TABLE 5

Malto has regularised these formatives in two ways:

1. placing phonological restrictions on them
2. incorporating them into verb paradigms.

Proto-Dravidian past stem formatives **-i-* and **-cc-* have been regularised in Malto. *-iy-* is the past tense marker in the third person and it attaches to stems ending in obstruents. *-c-* occurs as a stem formative everywhere else. *-ṯ-* as a stem formative remains as a relic in a few exceptional verbs such as *o:n-ḍ* and in most other cases it has been regularised as the past tense suffix in the first and second person. *-k-* is the fourth type of proto-Dravidian past stem formative (see Table 2) which functions as the relative past tense marker in Malto (Section 4.3).

The past tense paradigms for the verbs *o:n* 'drink', *hek* 'go' and *baɾ* 'come' are represented as follows:

- *-c* for *baɾ*, *-iy* for *hek* in the third person
- *-ṯ* everywhere else.

1SG	*o:n-ḍ-ṯa:n*	*hek-ṯa:n*	*baɾ-ṯ-a:n*
1PL	*o:n-ḍ-ṯ-a:m*	*hek-ṯ-a:m*	*baɾ-ṯ -a:m*
2SG M	*o:n-ḍ-ṯ-e*	*hek-ṯ-e*	*baɾ-ṯ-e*
2SG F	*o:n-ḍ-ṯ-i*	*hek-ṯ-i*	*baɾ-ṯ-i*
2PL	*o:n-ḍ-ṯ-aːɾ*	*hek-ṯ-aːɾ*	*baɾ-ṯ-aːɾ*
3SG M	*o:n-ḍ-iy-a:h*	*hek-iy-a:h*	*baɾ-c-a:h*
3SG NM	*o:n-ḍ-iy-a:ḍ*	*hek-iy-a:ḍ*	*baɾ-c-a:ḍ*
3PL [+HUM]	*o:n-ḍ-iy-aːɾ*	*hek-iy-aːɾ*	*baɾ-c-aːɾ*
3PL [-HUM]	*o:n-ḍ-iy-a:ḍ*	*hek-iy-a:ḍ*	*baɾ-c-a:ḍ*

TABLE 6

The present tense in Malto describes a situation that takes place simultaneously with the time of utterance. Bybee et al (1994, 126) point out that the present tense does not just have a deictic

temporal reference, but also covers various types of imperfective situations with the moment of speech as the reference point. In Malto, the habitual and the progressive aspects are both expressed using present tense. All verbs in Malto have the same inflectional pattern to mark present tense. The present tense paradigm for the verb *hek* 'go' is represented as follows:

- *-i*: in the first person singular and third person non-masculine
- *-n*: in the third person plural (human)
- *-d̪*: everywhere else.

1SG	*hek-i-iː n*
1PL	*hek-d̪-aː m*
2SG M	*hek-d̪-e*
2SG F	*hek-d̪-i*
2PL	*hek-d̪-aː ɾ*
3SG M	*hek-d̪-aː h*
3SG NM	*hek-i-iː d̪*
3PL [+human]	*hek-n-aː ɾ*
3PL [-human]	*hek-i-iː d̪*

TABLE 7

The future tense in Malto represents a situation that is predicted to occur after the time of utterance. Bybee et al (1994, 244) "regard the focal use of future as equivalent to a prediction on the part of the speaker that the situation in the proposition, which refers to an event taking place after the moment of speech, will hold." All verbs in Malto have the same inflectional pattern to mark future tense. The future tense paradigm for the verb *hek* 'go' is represented as follows:

- *-en* when it occurs before GNP markers beginning with front vowels
- *-an* everywhere else.

The future tense is haplologised in the first person singular.

1 SG	*hek-aː n*
1PL	*hek-an-aː m*
2SG M	*hek-en-e*
2SG F	*hek-en-i*
2PL	*hek-an-aː ɾ*
3SG M	*hek-an-aː h*
3SG NM	*hek-en-iː d̪*
3PL [+human]	*hek-an-aː ɾ*
3PL [-human]	*hek-en-iː d̪*

TABLE 8

Tense and the illocutionary force expressed by imperatives and permissives in Malto have scope over the entire clause. Modality, status and illocutionary force are often discussed together under the broad category of mood, as is done in this description of Malto.

Imperative constructions in Malto do not have the typical GNP agreement marking that finite verbs typically carry. The verb ends with the morpheme */-aː/*.

(17) a. led̪ɾa paːsaː paːv koɾ-aː
left side street enter-Imp
'Enter from the left side!' Directions

However, the vocative marker also acts as the imperative marker. The vocative is added to the verb stem to specify the identity of the addressee.

(17) b. koɾ-c-i lag-ɖe
enter-Pst-Pp approach-Voc.M
'Approach by entering!' Story C2

All other modals in Malto are either expressed as compound verbs.

Agreement

All finite verb words in Malto carry gender-number-person marking in agreement with the subject of the clause. The agreement markers are portmanteau morphs that simultaneously mark gender, number and person of the subject. The only exception is the finite verb word carrying imperative marking. Among the non-finite verb forms, the verb form that includes relative past tense marking also carries agreement marking with the subject. The gender-number-person agreement markers in Malto are formally derived from personal pronouns. These kind of inflectional affixes marking agreement are referred to as pronominal affixes in linguistic typology (Corbett 2006). Malto can be called a pro-drop language because it makes allowance for the possibility of omitting the pronominal subject of a clause. Omission of the pronominal subject is, however, not obligatory. The following table lists the gender-number-person suffixes against the corresponding personal pronouns.

	Personal Pronouns	GNP suffixes
1 singular	*eːn*	*-aːn*
1 plural	*eːm*	*-aːm*
2 singular masculine	*niːn*	*-e*
2 singular feminine	*niːn*	*-i*
2 plural	*niːm*	*-aːɾ*
3 singular masculine	*aːh*	*-aːh*
3 singular non-masculine	*aːɖ̲*	*-aːɖ̲*
3 plural human	*aːɾ*	*-aːɾ*
3 plural non-human	*aːɖ̲*	*-aːɖ̲*

TABLE 9

From the above table we notice that second person plural and third person plural; and third person singular non-masculine and third person plural non-human are homophonous. Sentences containing them are disambiguated by the tense marking in the case of the present tense.

The structure of the verb word in Malto can thus be summarised as follows:

Verb word = Verb root + [stem formative] + [derivational suffixes] + [inflectional suffixes]

Some possible combinations of the above formula are listed below in order of increasing complexity based on how many derivational and inflectional suffixes are attached to the verb base. However this is not an exhaustive list of possible verb word forms in Malto.

a) Verb word = verb root + inflectional suffix

(18) ɖ̲ok-aː
sit-Imp
'Sit!' Elicitation

b) Verb word = verb root + stem formative 1 + inflectional suffix

(19) baɾ-c-ah
come-Sf-3Sg.M
'He came.' Elicitation

c) Verb word = verb root + stem formative 2 + inflectional suffix

(20) ku-nj-ah
throw-Sf-3Sg.M
'He threw.' Story C4

d) Verb word = verb root + inflectional suffix + inflectional suffix

(21) ʈ̪al-ɖ̪-aːm
sacrifice-Prs-1Pl
'We sacrifice.' Rituals

e) Verb word = verb root + stem formative 1+ inflectional suffix +inflectional suffix

(22) ʈ̪al-c-aː-k-aːm ...
sacrifice-Sf-Ep-Rp-1Pl
'(We) having sacrificed... ' Rituals

f) Verb word = verb root + stem formative 2 + inflectional suffix + inflectional suffix + inflectional suffix

(23) ha-nɖ̪-laː-y-iːɖ̪
find-Sf-Neg-Pst-3Sg.Nm
'It could not find.' Story C3

g) Verb word = verb root + derivational suffix + inflectional suffix

(24) samjʰ-aːɾ-aːɾ
understand-Vrb-3Pl
'They convinced (him).' History

h) Verb word = verb root + derivational suffix+ derivational suffix + inflectional suffix + inflectional suffix

(25) avɖ̪-ɾ-nah-iy-aːɾ
talk-Dtr-Recp-Pst-3Pl
'They discussed (it) with each other.' Story C2

3 Finite Verb

The key to deciphering the inter-relation of verbs in multi-verb constructions in Dravidian languages lies in appreciating the meaning of finiteness. Morphologically, the finiteness of verbs in Malto and all other Dravidian languages depends on whether the verb is marked for tense and gender-number-person agreement. Syntactically, finite verbs can appear in independent clauses and they typically occupy the sentence final position. The gender-number-person agreement marker agrees with the subject of the sentence.

(26) sirip mak ɖaːhɖaː-n hoy-n-aːɾ
only Mak branch-Acc take-Prs-3Pl
'They only take the branch of the Mak tree.' Medicine

Miller (1993, 381) defines a complex verb as one which has undergone some sort of derivation to alter the form, meaning and argument structure of the base verb. A complex stem for a finite verb in Malto will have the derivative suffix preceding the tense marker.

(27) boh-a-ʈ̪ɾ-ɖ̪-aːn
run-Ep-Caus-Prs-1Sg
'I am caused to run. ' Elicitation

Compound verbs have two verbal bases — V1+V2. Usually only the second base V2 carries the tense and gender-number-person agreement marker. When compound verbs are positioned clause finally, only the second verb is a finite verb and the verb preceding it is a non-finite verb.

(28) uḏ-ʈaɾ-uḏ-ʈaɾ hi-j-aḏ
fix-Caus-Fix-Caus stood-Sf-3Sg.Nm
'(They) stood with the horns fixed (to each other).' Story C3

In the above example, the V1 *uḏʈaɾ-uḏʈaɾ* is a reduplicated compound verb word that forms the first base and the V2 *hijaḏ* is the fully inflected finite verb word. This example illustrates compound verb constructions with all their internal complexities in that V1 can itself be a compound verb.

The formal structure of verb-verb compounds in Malto is explained in the section on non-finite verbs (Section 4). In addition to verb-verb compounds, Malto also has noun-verb compounds. The most productive processes of noun-verb compounding in Malto involve using the verb 'to be' to encode stative predicates and the verb 'to do' to encode active predicates.

(29) haḏ-e saja: nan-iy-a:ɾ je ke-c-a:ɾ ha: bic-no saja naniya:ɾ
Dem.Dst-Dat punishment do-Pst-3Pl.H who die-Pst-3Pl.H Dem.Dst between-Loc punishment do-Pst-3Pl.H
'That, they punished, whoever died in that interval, they punished.' History

(30) ha:ḏ-inʈe salha: men-j-a:ɾ
there-Abl consultation be-Sf-3Pl.H
'From there, they consulted.' History

nan, the verb 'to do' and *men*, the verb 'to be' are usually used with a borrowed noun as in the above examples where *saja:* and *salha:* are both borrowings from Hindi. The verb 'to be' also appears with nominal predicates in copular constructions as illustrated in the following example.

(31) ha: bi:c-e-no bahuʈ sa:janga:-manga: men-j-a:ḏ
Dem.Dst between-Ep-Loc lot of confusion be-Sf-3Pl.Nm
'There was a lot of confusion during that time.' History

However noun-verb constructions are different from copular constructions on two counts. First, and most importantly, unlike the noun-verb construction, the copular construction is not a compound construction and secondly the nouns in noun-verb constructions are always borrowings from dominant languages (examples 29 and 30) whereas the nouns in copular constructions need not necessarily be borrowings (example 31).

Apart from declarative and negative sentences, finite verbs also appear in imperatives. The verb in the imperative clause carries the imperative suffix *-a:* or a vocative suffix that acts as a portmanteau morph combining both imperative mood marking and gender-number marking.

(32) guɾ-a:ɾ-e-k-e ba:li-n ʈis-g-a:
turn-Vrb-Ep-Rp-2Sg door-Acc open-Sf-Imp
'Having turned, open the door!' Directions

(33) ɖaḏ-onḏ maɾga-n ka:n-ɖ-ɖe
Clf-one horn-Acc strike-Sf-Voc.M
'Strike a horn.' Story C3

The finite verb form in Malto can stand as an independent clause. Structurally a finite verb word can be reduced or modified by introducing non-finite verbal inflections into the verb word. This involves loss or modification of verbal inflections such as the deictic tense suffix and the gender-number-person agreement marking.

4 Non-finite verbs

Steever (1993, 17) has stated for Dravidian in general that:

> "non-finite verbs are divided into two broad sets according to their combinatoric properties. The first set includes all those non-finite verbs which combine with the following verb, with or without other grammatical material intervening: the conjunctive, the infinitive, the durative, the conditional and

others. Their use implies the existence of another verb elsewhere in the sentence on which the non-finite forms depend. The second set of non-finite verbs includes those which combine with the following nominal to form a variety of structures. When, however, it combines with a following pronoun with a restrictive reading, the two combine and a verbal noun is formed."

The preconditional, temporal conditional, causal adverbial, relative past, simultaneity, infinitive, and conjunct participle forms in Malto combine with the following verb, the adjectival participles combine with the following nouns and the relativised adnominals are formed by combining with a following pronominal suffix. A non-finite verb is the syntactic head of a subordinate clause and functions as durative, perfective or conditional. Morphologically, non-finite verbs are usually differentiated from finite verbs by the absence of the TAM and GNP pronominal markers. But this is not always true in Malto. In the following sections I will discuss instances where the non-finite verb word carries relative tense marking and GNP agreement marking. All non-finite suffixes in Malto are inflectional suffixes since the addition of a non-finite suffix neither changes the meaning of the resultant verb nor does it alter the valence of the verb. The table below lists the suffixes involved in non-finite verb word formation.

Preconditional	*-ṯa*
Temporal Conditional	*-no*
Causal Adverbial	*-ko*
Relative Past	*-k-*
Simultaneity	*-i*
Infinitive	*-oṯ, -oṯi*
Conjunct Participle	*-a*
Adnominal	*-u, -ur*
Adjectival Participle	*-i*

TABLE 10

4.1 Conditionals

Conditional clauses are used to describe a situation that is a pre-requisite for another situation to occur. The verb word in the subordinate clause encoding the conditional protasis is a non-finite verb form. The main situation in this complex construction is expressed as a finite verb. The preconditional verb form appears with both the declarative and the negative forms. It is marked by the morpheme */-ṯa/*.

(34) niːn baj-y-a-ṯa eːn olh-aːn
2Sg.Nom hit-Prs-Ep-Cond 1Sg.Nom cry-1Sg
'If you beat me, I will cry.' Elicitation

(35) niːnu guran aʈ-oma-ṯ-i ṯa eːn olh-aːn
2Sg.Nom sweet give-Neg-Pst-2Sg.F Cond 1Sg.Nom cry-1Sg
'If you don't give me sweets, I will cry.' Elicitation

From the above examples it appears that conditional marker */-ṯa/* can be analysed as a clitic because *bajyaṯa* in example (34) is one intonational unit while *aʈomaṯi ṯa* has two intonational units with *ṯa* pronounced as a separate word form. Clitics are distinguished from suffixes by the nature of the forms with which they combine. Clitics combine with free forms, namely with words that can stand alone without the clitic. Steever (1993, 12) argues that "Dravidian clitics are exclusively postclitic and serve many important syntactic and pragmatic functions from conjunction and subordination to emphasis."

The second type is the temporal conditional clause that describes an entailment relation where an event y is bound to follow upon the completion of event x. The conditional clause is marked by the morpheme */-no/* and the matrix clause contains a finite verb.

(36) caːg-no haːɖ-inṯi baːd men-iːd
sow-Cond Dem.Dst-Abl after be-3Sg.Nm
'Upon sowing, after that it is available.' Rituals

Thompson et al (2007, 258) point out that the difference between the 'if' and 'when' clause is simply one of degree of expectability. Malto codes this difference by using two different morphemes to represent the two situations.

4.2 Causal Adverbial

The causal adverbial clause describes the situation where an event y can take place only if another event x occurs. Such a clause is marked by the morpheme */-ko/*. The clause describing the resulting situation y is the matrix clause and it contains a finite verb.

(37) han-no mahaɾe hi-l-aː haːn-ko, hi-j-ah
there-Loc facing stand-Sf-Imp say-Adv stand-Sf-3Sg.M
'Upon telling him to face that way and stand, he stood.' Story C4

4.3 Relative Past Tense

The verbs suffixed by the relative past tense marker denote situations that would have occurred prior to the situation described by the finite verb in the matrix clause. The suffix */-k/* marks the relative past. This maker is employed to give the sense of 'x having occurred'. Malto places no restrictions as to how many events can occur before the event described in the matrix clause. Hence this proves to be a productive process for clause chaining. The clauses occur in the sequential order which mirrors the event order in the sentence. The verb word in the matrix clause is a finite verb. The relative past marker is followed by GNP agreement marking. It also agrees with the GNP marker on the finite verb in the matrix clause.

(38) ninjaːriː nan-k-aːm bevaː-n uʈ-aː-ʈaɾ-aː-k-aːm haːɖ-inṯe baːɖu kuɾiː
invitation do-Rp-1P donation-Acc raise-Ep-Caus-Ep-Rp-1Pl Dem.Dst-Abl later worship
ʈhan-ek baɾ-c-aː-k-aːm, kuɾi kuʈi-no aɖi-ɖ-aːm
place-Dat come-Sf-Ep-Rp-1Pl worship place-Loc worship-Prs-1Pl
'Having invited (people), having collected donations, then having reached the place of worship, we worship at that place.' Ritual

The verb word containing the relative past tense marker is the only non-finite word form that takes GNP agreement marking. A plausible explanation for this exception is that relative past form was diachronically a compound word form that has contracted by reducing the V2 of the compound to a suffix. The possible contender for the V2 position in this case is *ek* 'go' since Kachru (1993, 117) generalises for South Asian languages that the verb 'go' as a vector regularly expresses the deictic meaning of completion.

4.4 Relative Present Tense/Simultaneity

Simultaneity is a relative tense marking on the verb where the speaker intends to express two events taking place at the same time. The verb bearing the simultaneity marker precedes the main verb in the matrix clause. The simultaneity marker in Malto is */-i/*.

(39) hani ɾaːjaː ʈahaːnɖi-ɖ ʈunɖ-i-ʈunɖ-i meca: haɾg-k-iː ʈunɖ-iy-aːɖ haːh-in
then king daughter-Nom see-Sim-see-Sim up climb-Rp-3Sg.Nm see-Pst-3Sg.Nm 3Sg.M-Acc
'Then the king's daughter seeing (it), climbed up and saw him.' Story C3

In example (39) the two simultaneous situations of *ʈunɖ* 'seeing' and *harg* 'climbing' are represented in two different clauses that are chained together.

4.5 Conjunct Participle

The conjunct participle in Malto is expressed by the suffix */-a/*. The clause with the verb word containing the conjunct participle precedes the finite verb in the matrix clause. This kind of construction

is also called a 'conjunct participle construction' because of the syntactic nature of the construction. The conjunct participle links two verb words that together describe one complex situation.

(40) haːɖ-en-u baːhɾe-no hoc-a kun-j-ɖ-aːm
Dem.Dst-Acc-En outside-Loc take-Cp throw-Sf-Prs-1Pl
'We take it out and throw it.' Medicine

4.6 Infinitive

/-oṭ , -oṭi/ is the infinitive marker. The verb form in an infinite clause is not inflected to agree with any subject and is understood to be co-referential with the matrix clause subject. From the temporal point of view, the infinitive suffix expresses the relative future tense in Malto. The infinitive word form is used as a purposive and as the complement of modal auxiliary verbs. The auxiliary verb expresses grammatical distinctions that are not expressed by the main verb. Example (41) shows the infinitive functioning as a purposive and example (42) shows the infinitive with the obligatory modal.

(41) oṛtoɳɖ ṭeho-ɖ aɾaː-n oy-oṭ ʔek-iy-aːɖ
one mother-Nom bamboo-Acc cut-Inf go-Pst-3Sg.Nm
'A mother went to cut bamboo.' Story C2

(42) haːɖ-en saːgoɾ moɖaː-ɖ haːn-oṭi meni
Dem.Dst-Acc Sagar turn-Nom say-Inf Oblig
'It is called the Sagar Turning.' Directions

Infinitives are also used as complements of the declarative and negative ability modals. The following example shows the infinitive with the negative ability modal.

(43) maː-oɳɖ saɾve-saɾve eːɾ mo.o lol-laː-iːɖ boh-oṭi
Clf-one small-small hen child can-Neg-3Sg.Nm run-Inf
'One little chick couldn't run.' Story C4

The default word order in Malto has the finite verb in the clause final position. However, the only exceptional case of a non-finite verb word occupying the clause final position is that of the infinitive wordform, as shown in the above example.

4.7 Adnominals

Adnominal clauses serve to modify a head noun. There are two kinds of adnominal constructions in Malto that contain verbal forms:

1. Restrictive relative clauses
2. Adjectival (relative) participle constructions

Headless relative clauses are formed by combining the relativised verb and the pronominal suffix that agrees with the head noun that it replaces. These constructions are used as restrictive relative clauses. When it is restrictive the relative clause restricts the potential reference of a head noun. */-uh, -uɾ/* are the relativising suffixes in Malto that attach to the verb stem (example 45). The suffix */-u/* is used when relativised verbal is followed by a noun (example 44). The suffix */-u/* is also the relativiser in the non-past forms and the relativised forms in the past tense take the suffix */-pa/* as shown in example (46).

(44) hoṛtu ɖahɖaː guɾ-aː-ṭaɾ-u maaː-h hok-ɖ-aːh
one branch roam-Ep-Caus-Rel child-Nom.m be-Prs-3Sg.M
'There is a boy who takes the branch around.' Medicine

(45) pahaːɖi-no ɖok-uɾ, maleɾ-in moa-n-aːɾ, han-iy-aːɾ
hill-Loc live-Rel people-Acc eat-Prs-3Pl say-Pst-3Pl
'The ones living on the hills eat people, they said.' History

(46) heng-inṯi pahle paḍ-i-pa-aːr ne jahaː bey-oː-r
1Sg-Abl before study-Ep-Rel-3Pl.H who Indef be-Neg-3Pl.H
'There is no literate person before me.' History

Krishnamurti (2003, 444) states that "all Dravidian languages change tensed finite verbs into adjectival (relative participles) by replacing the personal suffixes with adjectival markers *-a* or *-i*." The suffix */-i/* in Malto marks a verb in the attributive position that modifies the following noun. Syntactically this is an adjectival or relative participle construction.

(47) piːṯ-iː jaːnvaːraː men-no hem-e haːḍ-ki-in seromaːn hem-e kos-n-aːr
kill-Pp animal be-Cond 1Pl-Dat Dem.Dst-Gen-Acc portion 1Pl-Dat share-Prs-3Pl
'Upon finding a killed animal, they share a portion with me from that.' Medicine

5 Category Changing Derivational Processes

Malto uses suffixation to derive verbs from nouns and nouns from verbs.

5.1 Nominalisation: Deriving nouns from verbs

/-po/ is the nominalising suffix in Malto and is added to verb stems to derive nominals. They are the citation forms of verbs in Malto.

Verb Root	**Derived nominal**
hiːl 'stand'	*hiːl-po* 'standing'
ok 'sit'	*ok-po* 'sitting'
oṯ 'break'	*oṯ-po* 'breaking'

TABLE 11

Nominalised verbals function as gerunds in a dependent clause, as shown in the following example.

(48) haːḍ-inṯe saːtʰ-saːtʰ haːr-hiː joḍs-ṯiːn jen malaː guraːr-po porcaːr naːn-oːṯi
Dem.Dst-Abl together-together and-Emp pair-three Clf men wander-Nomr publicity do-Inf
'Then, together with the wandering of two or three people, to publicise (the event).' History

5.2 Verbalisation

/-aːr/ is the verbalising suffix in Malto. The verbalising suffix is used to convert non-verbal roots into intransitive verb stems in Malto.

Non-verbal root	**Derived Verb**
alsi 'sweat'	*als-aːr* 'be irritated'
kaːkli 'misery, pain, predicament'	*kaːkl-aːr* 'to be distressed'
ḍagraha: 'bad, wrong'	*ḍagr-aːr* 'do wrong'

TABLE 12

The suffix */-aːr/* is also used to verbalise borrowed words from Hindi. The borrowed root can either be a verb root or a predicate nominal. The verbalising suffix is mandatory even if the borrowed word is a verb in the source language.

Borrowed root	**Derived Verb**
samaj^{h} 'understand'	*samj*^{h}*-aːr-aːr* 'they convinced (him/her)'
bana 'make'	*ban-aːr-aːh* 'he made'

TABLE 13

Example (48) shows both the category changing derivational processes on the borrowed root guɾ. The borrowed word is first verbalised into a Malto verb stem by adding the */-aːɾ/* suffix and then nominalised by the */-po/* suffix.

The */-aːɾ/* suffix is replaced by the */-ey/* or */-es/* suffix to derive transitive verbs from nouns and borrowed stems. The following example set shows the derivation of a transitive verb from a noun.

(49) a. *alsi* 'sweat'
b. *als-aːɾ* 'to be irritated'
c. als-es-iy-aːh
worry-Tr-Pst-3Sg.M
'He irritated (it).' Elicitation

6 Conclusion

This paper has explained the formal structure of the verb word in Malto in terms of the levels of word formation and the role and place of each morpheme within a verb word. Malto has three main levels of word formation namely, stem formation, derivational suffixation and inflectional suffixation. We have had a close look at the classification of verb roots, stem formation strategies and category changing derivational processes involving verbs in Malto. Although the information provided on the formal structure can be claimed to be exhaustive, the functional aspects of the verbal forms are beyond the scope of this paper.

Acknowledgments

I thank the Wingate Trust for providing me with a post-doctoral funding that has enabled me to develop this article. I am also grateful the Endangered Languages Documentation Project, SOAS and Central Research Fund, University of London for supporting my fieldwork. Special thanks are due to Pulak Mathur, Chandu Pahariya and all the residents of Angwali village, Jharkhand for enthusiastically participating in my data collection process.

Abbreviations

Abl	Ablative	Neg	Negative
Acc	Accusative	Nm	Non Masculine
Ad	Adnominal	Nom	Nominative
Add	Additive Particle	Nomr	Nominaliser
Adv	Adverbial	Oblig	Obligatory
Caus	Causative	Opt	Optative
Clf	Classifier	Pass	Passive
Comp	Comparative	Pl	Plural
Cond	Conditional	Pp	Past Participle
Cp	Conjunct Participle	Prf	Perfective
Dtr	Detransitiviser	Prs	Present
Ep	Epenthetic	Pst	Past
Emp	Emphatic	Q	Question
En	Enunciative Vowel	Recp	Reciprocal
Fut	Future	Rel	Relativiser
Gen	Genitive	Rp	Relative Past
H	Human	Sf	Stem Formative
Imp	Imperative	Sg	Singular
Indef	Indefinite Particle	Sim	Simultaneous
Inf	Infinitive	Tr	Transitiviser
Loc	Locative	Voc	Vocative
M	Masculine	Vrb	Verbaliser

References

Bybee, Joan. 1985. *Morphology: A study of the relation beween meaning and form*. Amsterdam: John Benjamins.

Bybee, Joan, Revere Perkins, and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: The University of Chicago Press.

Cladwell, Robert. 1956/1961. *A Comparative Grammar of the Dravidian or South-Indian Family of Languages*. Madras: University of Madras.

Corbett, Grenville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Das, Shishir Kumar. 1973. *Structure of Malto*. Annamalainagar: Annamalai University.

David, Anne. 1999. A Comparative Study of Dravidian Infinitives. Unpublished Ms., Chicago University.

Dixon, Robert M. W. 2000. A Typology of Causatives: Form, Syntax and Meaning. In R. M. W. Dixon and A. Y. Aikhenvald, eds., *Changing Valency: Case Studies in Transitivity*, pages 30–83. Cambridge University Press.

Droese, Rev. Ernest. 1884. *Introduction to the Malto Language*. Agra: Secundra Orphanage Press.

Israel, M. 1979. *A Grammar of the Kuvi Language*. Tiruvananthapuram: International School of Dravidian languages.

Kachru, Yamuna. 1993. Verb Serialization in Syntax, Typology and Historical Change. In M. Verma, ed., *Complex Predicates in South Asian Languages*. New Delhi: Manohar Publishers.

Krämer, Martin. 2003. *Vowel Harmony and Correspondence Theory*. Berlin: Mouton de Gruyter.

Krishnamurti, B.H. 2003. *The Dravidian Languages*. Cambridge: Cambridge University Press.

Mahapatra, B.P. 1979. *Malto: An Etnosemantic Study*. Mysore: Central Institute of Indian Languages.

Miller, Gary D. 1993. *Complex Verb Formation*. Amsterdam: John Benjamins.

Steever, Sanford B. 1987. The Roots of the Plural Action Verb in the Dravidian Languages. *Journal of the American Oriental Society* 107(4):581–604.

Steever, Sanford B., ed. 1998. *The Dravidian Languages*. London: Routledge.

Subrahmanyam, P.S. 1971. *Dravidian Verb Morphology*. Annamalainagar: Annamalai University.

Winfield, W.W. 1928. *A Grammar of the Kui Language*. Calcutta: Asiatic Society of Bengal.