# Null arguments and the inverse problem

2nd revised manuscript

October 2020

**Abstract**. Native speakers can reconstruct null arguments from a linear string of words where they are notoriously absent. How they solve this problem remains an unsolved and largely unaddressed issue. This article argues for a unified analysis of null arguments that relies on bare sensory input. The key assumption is that antecedent recovery takes place at the syntax-semantics interface if and only if unvalued phi-features of a lexical element cannot be valued from the information available in the input. The analysis unifies the theory of finite null subject pronouns and control and further interprets null subject sentences without generating actual null pronouns. Evidence is considered from three languages with distinct null argument profiles: English (non-pro drop), Italian (consistent pro-drop) and Finnish (partial pro-drop). Finally, the model is formalized and tested deductively by a computer simulation.

**Keywords**: null arguments; control; pro-drop; computational linguistics; language comprehension; parsing

## 1 Introduction

The notion of recovery plays a role in the theory of null arguments. Being inaudible, their syntactic and semantic properties must be recovered from an “overt linguistic context,” as observed by Rizzi (1986: 520). To illustrate, consider the examples in (1)-(2).

(1)

a. *John wants to leave.*

b. *John wants Mary to leave*.

(2)

a. **John orders to leave*.

b. *John promises Mary to leave*.

A grammatical theory must explain why the thematic agent of *leave* is *John* (1a) but *Mary* in (1b), and why *John* does not constitute a possible antecedent in (1b). Furthermore, the antecedent relations change as a function of a change in the main verb (2a-b). Yet, the surface strings contain no direct cues for any of these properties. How speakers accomplish this inverse problem, as I will call it here, remains an unsolved and largely unaddressed issue that this article proposes to solve. The solution argued for in this paper and developed on the basis of Borer (1986, 1989) is that null argument recovery is caused by the presence of an unvalued phi-set (e.g., number and person features) of a lexical item that cannot be valued morphosyntactically by using the resources available at the input. They are valued at the syntax-semantics interface, which results in the phenomenon known as control.

The material is organized in the following way. Section 2 reviews and elucidates the linguistic background, considering data from three languages with distinct null subject/object properties: English, Italian and Finnish. Section 3 presents the hypothesis, keeping with core ideas, while Section 4 addresses the details of formalization and tests the analysis by means of computational simulation. Section 5 contains the conclusions.

## 2 Background

### *2.1 The inverse problem*

Let us begin by looking at some of the computational challenges associated with the inverse problem and null arguments. We consider one type of null argument, the phonologically null subject pronoun usually referred to as “pro” in the literature. At first approximation, this element occurs in the position of the finite clause subject in the presence of sufficiently rich subject-verb agreement. Thus, it is available in Italian (3a) and Finnish (3b), but not in English (3c).

(3) a. *Italian*
(Io) parl-o.
(I) speak-1SG
‘I speak.’
b. *Finnish*
(Minä) puhu-n.
(I) speak-1SG
‘I speak.’
c. *English*
*(I) speak.∅
‘I speak.’

We might therefore consider that the native speaker processing these sentences infers the existence of the first-person subject on the basis of two facts accessible from the sensory input: presence of rich agreement and absence of an overt pronoun/subject. This guess, although perhaps intuitively compelling, turns out to be insufficient. Consider (4a-b).

(4)
a. *Italian*
Sembra essere io.

seems to be I

’It seems to be me.’

b. *Finnish*

| | | | | | | |
|---|---|---|---|---|---|---|
| Nämä | kirja-t | ol-i | halunnut | ostaa | Merjalta | Jari. |
| these | book-PL.ACC | had-3SG | wanted | to.buy | from.Merja | Jari.NOM |

’Jari had wanted to buy these books from Merja.’

These sentences contain an overt grammatical subject, but it occurs in a noncanonical position. For example, while the canonical word order in Finnish is SVO, the preverbal position in (4b) is occupied by the direct object, and the grammatical subject *Jari* is in the last position of the clause, behind a sequence of finite and non-finite verbs. Because the subject occurs in the last position, speakers must analyze the whole sentence to judge if a subject is missing and thus if its properties should be inferred from agreement alone. The parser must, furthermore, provide the sentence with a real and intelligible parse in order to recognize that *nämä kirjat* ‘these books’ constitutes a direct object and ‘Jari’ could potentially fill in the role of the grammatical subject. It must do this without assuming that they are prototypical or simple DPs: ‘these books’ could be substituted with ‘the book that I discussed yesterday with Bill’ and ‘Jari’ with ‘the president of an association I am a former member of’, showing that to get these properties right the first pass parse must perform real parsing with no artificial limit imposed on complexity. Simple cues such as case marking are insufficient: Finnish has both nominative direct objects and non-nominative subjects/thematic agents (Vainikka 1989; Nelson 1998). Finally, the first pass parse must be correct. If the parser errs, conditions for the null subject could be satisfied in the wrong way. Thus, already the intuitively simple task of finding out whether a grammatical subject is missing, just to mention one small part of the inverse problem, presents a nontrivial challenge for the human language comprehension system.

### *2.2 Introduction to null subjects*

Before looking at the inverse problem specifically it is useful to gather the basic linguistic facts any model of null argument must minimally account for. The several decades of work that began with the discovery of null arguments (Rosenbaum 1967; Postal 1970; Rosenbaum 1970; Brame 1976; Perlmutter 1971) and continued during the GB-era and beyond (e.g., Chomsky 1980, 1981; Rizzi 1982; Hyams 1989; Jaeggli 1980; Chomsky 1982, to mention a few sources) have provided an overall taxonomy of null arguments that I will use as a starting point in this work.

The finite null subject pro illustrated by the example (3) constitutes a subtype of null arguments connected in some way to agreement (Perlmutter 1971; Rizzi 1982, 1986; Chomsky 1982; Taraldsen 1980; Huang 1984; Jaeggli and Safir 1989). In a language such as Italian with rich verbal agreement, the subject pronoun of a finite clause can be omitted or silenced phonologically. Some properties of the subject can then be inferred from agreement alone. In languages with meager agreement, subjects cannot be silenced. Matters are complicated by the existence of partial pro-drop languages, such as Finnish or Hebrew, in which only a subset of subject pronouns can be silenced (Vainikka and Levy 1999; Holmberg 2010). In particular, the third person pronoun can be silenced in Finnish if and only if there is both subject-verb agreement and a suitable antecedent (5a-c). The second condition distinguishes Finnish from a consistent pro-drop language such as Italian.

(5)

| | | | | | | |
|---|---|---|---|---|---|---|
| a. | *Halua-a | ostaa | kirja-n. | | | |
| | want-3SG | to-buy | book-ACC | | | |
| | 'He wants to buy a book.' | | | | | |
| b. | Halua-n | ostaa | kirja-n. | | | |
| | want-1SG | to.buy | book-ACC | | | |
| | 'I want to buy a book.' | | | | | |
| c. | Jari$_1$ | sanoi | että | [halua-a | ostaa | kirja-n.] |
| | Jari | said | that | want-3SG | to.buy | book-ACC |
| | 'Jari said that he (=Jari) wants to buy a book.' | | | | | |

Only the coreference reading, in which Jari is both the speaker and the buyer, is available in (5c). This example shows that pro requires both formal licensing, a set of grammatical properties which determine when it is available, and some form of antecedent recovery that uses the linguistic context to further infer the properties of the covert subject pronoun. The data in (5) complicates the inverse problem: the acceptability of the third person null argument depends on the larger context, not only on local morphosyntax.

Pro can be distinguished from a different type of null argument that is not licensed by agreement and exhibits (at least in appearance) different recovery behavior. It is illustrated by examples (1-2) and (6) below.

(6)

a. *John$_1$ wants* PRO$_1$ *to leave*.

b. **John orders to leave*.

c. *John wants Mary to leave*.

d. *John$_1$ promises Mary$_2$* PRO$_{1,*2}$ *to leave*.

PRO, which is often used in the literature to signify the missing argument when discussing sentences of this type, occurs in subject positions of infinitival verbs and gerundive nouns that exhibit neither morphosyntax nor impoverished morphosyntax (see, e.g., Boeckx and Hornstein 2004; Culicover and Jackendoff 2001; Landau 2000, 2003, 2004, 2013; Hornstein 2001, 1999; Manzini and Roussou 2000; Chomsky 1980, 1981; Martin 1996; Rosenbaum 1967, 1970; Postal 1970). Recovery cannot rely on agreement suffixes. Recovery, called *control* in the literature, is a matter of prototypically local dependency that selects for c-commanding subject and direct object antecedents, although both c-command and locality should be regarded as typical and not necessary properties of control (see especially Landau 2013, ch. 4.1.2).[1] This makes it possible to hypothesize that both pro and PRO share an anaphoric component (Borer 1989; Huang 1989). Italian finite null subjects are, however, not anaphoric in this sense, and in Finnish and Hebrew anaphoricity is limited to third person null subjects.

If no antecedent is present, an impersonal or arbitrary interpretation results that refers to ‘people in general’. This is illustrated by (7a). The same interpretation emerges in Finnish under certain conditions also when the object is null (7b).

(7)

a. To give up too easily would be a mistake.
‘For people in general, it would be a mistake to give up too early.’

b. *Finnish*

| | | | | |
|---|---|---|---|---|
| Epäonnistuminen | pakotta-a | __ | [harjoittelemaan | enemmän.] |
| failure | forces-3SG | | to.practise | more |

’A failure forces *(one) to practice more.’

Another similarity between PRO and pro is the fact that in Finnish also the pro-construction can create the generic or impersonal interpretation, as shown in (8).

(8)

| | | |
|---|---|---|
| Tässä | istu-u | mukavasti. |
| in.here | sit-3SG | comfortably |

‘One can sit here comfortably.’

This observation presents a further challenge to the inverse problem. Sentences like (8) involve third person agreement without a grammatical subject or antecedent, yet the sentence is not ungrammatical. It is grammatical, however, only if interpreted as generic. In addition, the sentence evaluates as ungrammatical if the preverbal position is not filled in by the locative PP or by another suitable phrase (see Holmberg and Nikanne 2002 and Huhmarniemi 2019 for discussion).

Although the verbal sequence ‘want – to leave’ can occur either with or without the infinitival agent argument in English (*John wants Mary to leave*, *John wants to leave*), in Finnish only the controlled version

---

[1] Nonlocal control is illustrated by examples such as *John thinks that it will be difficult PRO to leave*. Chomsky (1981) notices several pragmatic factors that regulate long-distance recovery and suggests that the null anaphor (PRO) “searches for a possible antecedent within its own clause, and if it can’t find one there, looks outside” (p. 78). The recovery algorithm proposed in the present work to account for control dependencies will not be sensitive in any way to the finite clause boundary, accepting the general line of thought in Landau (2013) that dispenses with the idea that control were confined to infinitival domains.

is grammatical (9). This creates *obligatory control*: the antecedent mechanism is the only mechanism pairing the predicate with its argument.

(9) *Finnish*
a. Jari halusi lähteä.
Jari wanted to.leave
'Jari wanted to leave.'
b. *Jari halusi Merja-n lähteä.
Jari wanted Merja-GEN to.leave
'Jari wanted Merja to leave.'

The pattern reverses if we use a different infinitival construction, the Finnish VA-infinitival (10).[2]

(10) *Finnish*
a. *Jari halusi lähte-vän.
Jari.NOM wanted leave-VA/INF
b. Jari halusi Merja-n lähte-vän.
Jari.NOM wanted Merja-GEN leave-VA/INF
'Jari wanted Merja to leave.'

Thus, whether the subject of an infinitival is null obligatorily or optionally depends in Finnish in some manner on the constitution of the selecting item and the selected item (Brattico 2017). Whatever the mechanism is, its properties are not trivial and subject to considerable debate in the literature.

At this stage it is important to point out that the above discussion neither assumes nor presupposes that null arguments (e.g., pro, PRO) constitute real silenced pronouns. Indeed, several grammatical theories do not make that assumption (see Janke 2008 and references mentioned therein), and I will end up rejecting it as well. How null arguments are represented in grammar is the problem a theory of null arguments must solve, not assume. The question is, instead, how to derive the syntactic and semantic properties of sentences in which we 'hear' the presence of participants that do not correspond to anything directly available in the surface string.

[2] Finnish A-infinitivals, such as *want*, are often desirative in meaning, or perhaps they have future tense, whereas the VA-infinitival is propositional and exhibits overt past-present tense alteration (but not finiteness). In both cases, the thematic agent of the infinitival is located inside the infinitival phrase by all syntactic tests. For Finnish infinitivals, see Koskinen (1998) and Vainikka (1989).

# 3 Null arguments and the inverse problem

## *3.1 Introduction*

In this and the next section I will look at the problem of null arguments from the point of view of the inverse problem. When analyzed from such vantage point, every rule or principle must be formulated by referring exclusively to overt sensory objects or to a structural interpretation or other higher-order property generated directly from such input objects. The justification for this assumption is the fact that native speakers can provide correct structural and semantic interpretation for sentences in their own language without consulting anything else expect the surface string. Moreover, the process is effortless; standard cases of null arguments do not create garden paths.

A theory of this type must contain two ingredients. It must contain (1) a formal framework implementing a *parser* (i.e. function from sensory objects into sets of semantic interpretations) and (2) an embodiment of an analysis of null arguments within that framework. I will begin by formulating the analysis of null arguments and then turn to the implementation of the parser component. Once the components have been set up, they will be formalized and tested by means of computer simulation.

## *3.2 Null subject pronoun (pro)*

A necessary first step in solving the inverse problem is the retrieval of lexical items and their features on the basis of phonological words occurring in the input. In addition, agreement suffixes must be extracted and isolated in order to handle the pro-drop phenomenon. Let us therefore begin from the assumption that phonological words occurring in the input are matched with a linear string of lexical items retrieved from the surface vocabulary, each lexical item consisting of a set $\{f_1, \ldots, f_n\}$ of features. If the phonological word contains several primitive grammatical items, say tense, causativization and a verbal stem, these elements are first decomposed and then matched with lexical items. The process is illustrated in Figure 4. Merge refers to a computational operation that builds syntactics structures from incoming lexical items and will be discussed later.

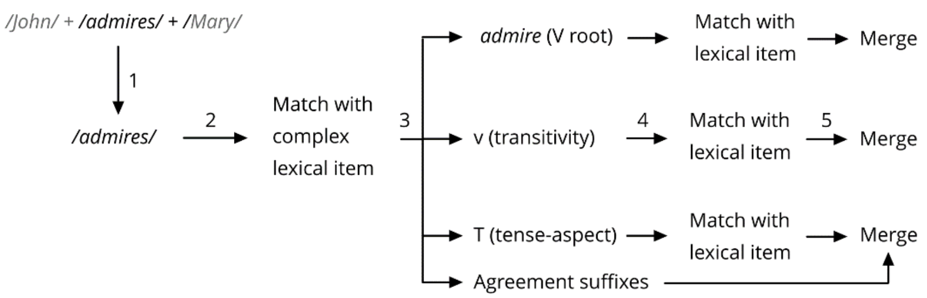


*Figure 4. A schematic overview of the structure of the surface vocabulary and the computational operations involved in lexical decomposition. The pipeline contains the following steps: (1) consume a phonological word from the input; (2) match the word with a lexical entry, or a list of lexical items if it is ambiguous; if the matched entry is complex, break it into primitive parts (3), each matched with a primitive lexical item (4); Merge (i.e. attach) the lexical item into the phrase structure. If a component of a word maps to an inflectional affix, it is interpreted as a feature. This means that a feature can appear through the "lexical route" (i.e. from the lexicon) or through the "input route" (i.e. as an element, such as a suffix, in the PF-input). The assumptions provided in Figure 4 seem to me to constitute the absolute minimum for doing anything useful in language comprehension.*

To establish a connection between pro-drop and agreement, let us assume that agreement suffixes are extracted from phonological words and stored as features of grammatical heads, as shown by the arrow "agreement suffixes" in Figure 4. There are two types of phi-features that we must distinguish from each other. Proper names or pronouns retain their phi-features no matter where they occur in the structure, whereas the same features at the finite verb covary with those of a local argument. Let us assume, therefore, that proper names and other nominals obtain their phi-features via the *lexical route*, whereas the finite verb gets them via the input route (Figure 4). The received generative view, which I will adopt here as a starting point, maintains that finite T comes with an *unvalued* phi-set (denoted by φ_ in this article) that will come to reflect the intrinsic phi-features of a local argument. Whether a functional head has an unvalued phi-set is determined lexically. We say that an element with intrinsic phi-set values an unvalued phi-set by an operation *Agree* (Chomsky 2000: 121–26). The operation is illustrated in (11). The element with unvalued features is called the *probe*, its counterparty the *goal*.

(11)

| | | |
|---|---|---|
| John | | admire-s. . . |
| 3SG | ←Agree→ | φ_ |
| goal | | probe |
| Lexical φ | | Lexical φ_ |
| 'Argument' | | 'Predicate' |

The mechanism in (11) assumes that the unvalued phi-set is valued by the phi-features of *John*. Another possibility is that it is valued by the overt verbal agreement features (*admire+s*) that were extracted from the input. That feature arrives via the input route and not by the lexical route. Let us assume that the unvalued phi-set can be valued either (i) by a local DP-argument by means of Agree or (ii) by overt phi-features of the head itself, if present in the input and thus part of the feature composition of T. The second route is illustrated by the Finnish pro-drop sentence (12).

(12) *Finnish, grammatical pro-drop construction*

Ihaile- n Merja-a.

admire- 1SG Merja-PAR
φ_ ← 1SG
‘I admire Merja.’

To capture the pro-drop signature, we can now assume that valuation by means of an overt pronoun and by means of overt agreement suffixes are syntactically equivalent with respect to the output of the operation. Notice that no null pronoun is projected. This solution therefore agrees with the style of analysis developed by Alexiadou and Anagnostopoulou (1998), Barbosa (1995, 2009), Manzini and Savoia (2002), Borer (1986, 1989) and Platzack (2004), all which differ from each other in details but are united by the assumption that agreement suffixes are ‘pronominal’ enough to assume the role of a full subject pronominal. The analysis disagrees with Holmberg (2005) and most of the literature on Finnish (Vainikka 1989; Vainikka and Levy 1999; Vilkuna 1989, 1995) and other analyses (Cardinaletti 2004; Sheehan 2006) that claim that the clause must contain a covert phrasal subject at SpecTP. I will return to this matter further below.

The overt phi-cluster at T has no anaphoric properties, which means that the analysis fails to capture the properties of the Finnish/Hebrew third person pro. There must be some condition that comes into play and is not present in Italian. Holmberg (2005) and Holmberg and Sheehan (2010) propose a solution that I will adopt here. They propose that the Finnish third person pro contains an unvalued D-feature (denoted by D_ in this article) that must be valued by means of an antecedent. To transform this idea to the comprehension perspective, we must posit a condition to the effect that the overt Finnish third person agreement suffix in the input values the person and number features but does not value D_. This would make the third person suffix to behave like a variable, in the sense that it lacks at least one referential feature.[3] Suppose, furthermore, that the presence of an unvalued D-feature triggers a *recovery mechanism* at the syntax-semantics interface that attempts to locate a suitable antecedent, providing the missing value. The idea is illustrated in (13).

(13) *Finnish*
Jari sanoo että ihaile-e Merja-a.
[Jari [says [that admire-3SG Merja-PAR]]]
D ←—— Recovery ——→ D_
‘Jari says that he (=Jari) admires Merja.’

Thus, according to this analysis the recovery mechanism constitutes an operation that deals with an uninterpretable phi-feature when it passes through the morphological and syntactic components and ends up at the syntax-semantic interface.

[3] I am assuming that D_ hosts values such as ‘definite’, ‘indefinite’ and ‘generic’, but the list is not exhaustive.

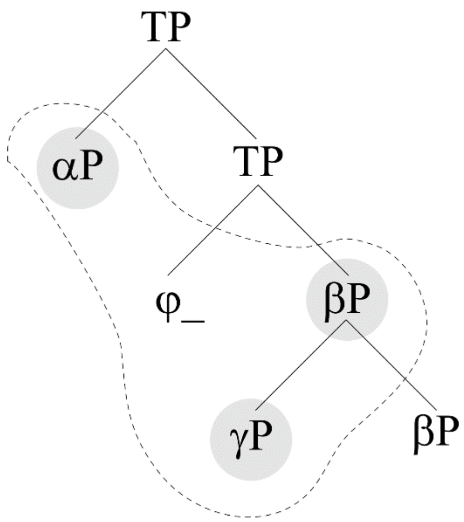
What constrains recovery? One condition is c-command: if an argument with a matching phi-set is found that c-commands the offending feature, that element will be selected as the antecedent (Vainikka and Levy 1999; Holmberg 2005, 2010; Holmberg, Nayudu, and Sheehan 2009). Thus, in the example (13), Jari must be the antecedent. An element that is c-commanded by the variable can never be selected, for example. On the other hand, very few reliable locality conditions have been reported in the literature, so I will assume that the antecedent search for D_ is in principle unlimited in upward distance, as argued by Holmberg and Sheehan (2010). Any c-commanding third person antecedent can be selected as a legitimate target.[4]

The pro-drop phenomenon is absent in English. Adopting the basic idea of Jaeggli and Safir (1989), I rely on the fact that the English agreement suffixes are ambiguous: *admire* is consistent with several pronouns, *admires* with two gender features (*he* or *she*). If the agreement features at a head cannot be associated with an unambiguous pronoun, then the null argument reconstruction fails. In other words, conflicting or incoherent phi-sets block pro-reconstruction, as they would lead into an uninterpretable or impossible pronoun; otherwise the mechanism is available. If a predicate contains no valued phi-features, no conflict arises and null arguments are automatically licensed. I will use this later to capture agreementless radical pro-drop languages such as Chinese, Japanese and Korean (Rizzi 1986; Huang 1984).

The analysis presupposes a theory of Agree and its inverse version, call it Agree-1, that handles agreement in the presence of an input. I assume that an unvalued phi-set can agree with (i) a sister DP, (ii) DP inside its sister, and with (iii) the specifier DP, in this order and with successful match blocking further operations. These options are illustrated in (14), with the possible goals being αP, βP and γP. The operation searches the sister first (βP, γP), then SPEC (αP).

[4] Holmberg and Sheehan (2010) report several examples in which an antecedent is selected that does not c-command the variable, but in these examples no structural c-commanding antecedent is present. I will assume that such antecedents are accessed only if no c-commanding antecedent is available. The matter will be discussed in Section 4.6. See also Brattico (2017).

(14) [SPEC$_{\alpha}$ [T(φ_) [$_{\beta}$. . .DP$_{\gamma}$. . .]]]




*Locality domain for Agree-1*

While the theory of Agree is usually formulated so that it only allows (i-ii), concrete simulations showed that condition (iii) constitutes a useful addition when working with input objects that will often have arguments at specifier positions in the surface string. I will also assume the phase impenetrability condition (PIC), which prevents Agree-1 from searching below/above vP and CP (Chomsky 2008, 2000: 108). This prevents T from agreeing with an argument over an arbitrary distance (e.g., **we think-s that John admires Mary*, with Agree$^{-1}$ (*thinks*, *John*)). To rule out sentences in which the verbal agreement suffixes and the phi-features of a local DP argument do not match (e.g., **John leave*), I assume that the final phi-set at any given head, resulting from Agree-1 and local phi-features, must not involve a feature conflict. A situation in which a third person subject is combined with first person agreement will be filtered out.

To make the system completely water-tight in anticipation of the formalization and eventual computer simulation, we have to solve one additional technical issue. Not all heads trigger agreement, less so carry overt agreement suffixes. An agreeing head must have some property allowing it to host and express phi-features. I assume that this property is represented by a lexical feature ±VAL(uation). A head marked for –VAL will not be able/will not to value φ_ and Agree-1 is not triggered. Agreementless particles, connectives and other such items are also marked for –VAL. The feature also distinguishes object-verb agreement languages from languages that only allow subject-verb agreement. The former, and not the latter, have VAL-marking in some of the verbal components (v, V). Similarly, Finnish, like Hungarian, exhibits a wide variety of infinitival finite agreement phenomena (discussed briefly later) which will be captured by assuming that these infinitival heads are marked by +VAL. In English and Italian, most or all infinitival heads are marked for –VAL. For present purposes, we can think of ±VAL as a lexical feature controlling the application of Agree-1.

To summarize, lexical items may enter the derivation with an unvalued phi-set that requires valuation. Whether a lexical item has an unvalued phi-set and whether it can be valued by Agree-1 is decided in the lexicon. The former property is controlled by the presence/absence of unvalued phi-features (abbreviated as ±PHI), the latter by feature ±VAL. An unvalued phi-set in the category (+PHI, +VAL) may get valued either by overt argument or by verbal agreement suffixes, the latter option creating a necessary condition for pro-drop. If a feature arrives to the syntax-semantics interface unvalued, recovery attempts to value the features by searching for an antecedent.

**Table 1. Four types of predicates generated in this study.**

| | OVERT AGREEMENT POSSIBLE | |
|---|---|---|
| UNVALUED PHI-FEATURES | –VAL (NO) | + VAL (YES) |
| –PHI (NO) | Frozen particles that do not introduce or are not linked with arguments (e.g., *and*, *but*, *that*). These elements can be thought of as not being 'predicates'. | Words that agree by concord? Concord was not modelled in the present study. |
| +PHI (YES) | Predicates linked with arguments, but which do not exhibit Agree-1. This category will play a key role in control. | Predicates linked with arguments by valuation or phi-features and which exhibit Agree-1 (e.g., finite verbs in Finnish, Italian, English). Partial pro-drop will be captured by relying on incomplete or deficient valuation. |

Crucially, the whole analysis has been set up in such a way that it works *in principle* by assuming only the bare sensory input. There is no stage at which the existence of an invisible/inaudible pro is assumed. Features ±ARG and ±VAL can be retrieved from the lexicon on the basis of the sensory input, whereas Agree-1 is formulated so that it applies to a phrase structure reconstructed directly form the input.

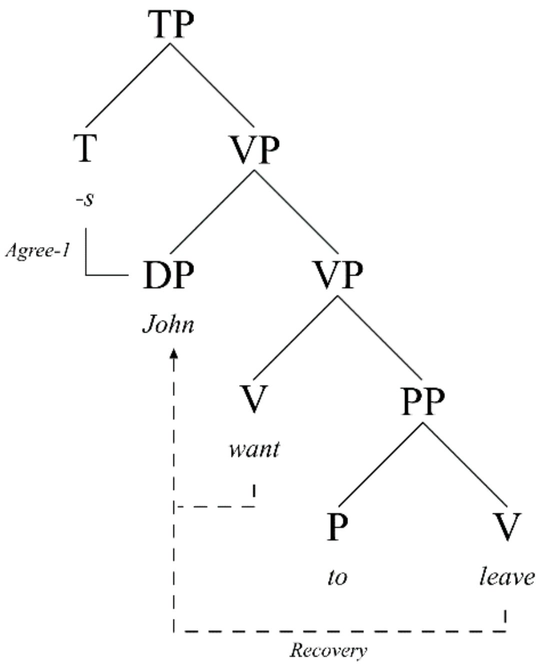
### *3.3 Control*

An unvalued feature at the syntax-semantic interface was assumed to trigger recovery. In the case of Finnish/Hebrew third person pro, the triggering feature was D_. Let us generalize this proposal and say that if all phi-features remain without value at the syntax-semantics interface, then the closest c-commanding argument that can value these features and does not have conflicting valued phi-features is selected as an antecedent. The proposed mechanism is based on the Minimal Distance Principle (MDP) originally proposed by Rosenbaum (1967, 1970; see also Lasnik 1991), well-known for its shortcomings but useful as a starting point and attractive in its conceptual and algorithmic simplicity. Consider what happens when the system is provided (15) as an input. I will assume that *leave* contains an unvalued phi-set as a lexical feature.

(15) John$_{\varphi}$ wants$_{\varphi_}$ to leave$_{\varphi_}$.

The infinitival verb *(to) leave* does not exhibit overt agreement and there is no local argument here that Agree-1 could detect. Feature φ_ will arrive to the syntax-semantic interface unvalued. This triggers recovery, which selects the closest antecedent, *John*. I notate antecedents in this article by writing "φ_=John", with the left side containing the unvalued features and the right the antecedent phrase.

(16) John [wants to leave$_{\varphi_=\text{John}}$.]




If an argument occurs between the agent of the main verb and the infinitival, then, all else being equal, it will be selected as an antecedent, deriving *John wants Mary to leave*$_{\varphi\_=Mary}$. This will provide an antecedent for the infinitival verb and link it with an argument missing in the input. No null argument (PRO) is projected; we keep within the framework in which everything is reconstructed form the overt input. The analysis is crucially based on earlier work by Borer (1986, 1989), who proposed that null pronouns themselves are not anaphoric; instead, the anaphoric behavior emerges from features inside functional heads. Borer assumes, furthermore, that each functional head with agreement features comes with the property that it must be "linked" with an argument (that she calls "I-subject") in its "accessible domain." The agreement reconstruction process (Agree-1, (14)) proposed earlier in the present work can be thought of an inverse variation of Borer's linking principle. The idea is that if and only if a predicate cannot be paired with an argument by a morphosyntactic agreement operation working directly with the input representation, the situation will get 'repaired' at the syntax-semantic interface by means of recovery. Another precursor that has influenced the present approach is that of Janke (2008), who eliminates PRO by unifying control with anaphora resolution and proposes that thematic roles of predicates can be saturated by a upward looking percolation mechanism resembling the recovery algorithm proposed in the present study.[5] I find both approaches particularly useful in solving the inverse problem.

The analysis presupposes that we distinguish antecedent recovery for D_ (referentiality, definiteness) and φ_ (person, number): the latter will create standard control, the former a more liberal antecedent recovery observed in connection with Finnish partial pro-drop. The fact that Finnish third person null subject antecedent differs from obligatory control (OC) and non-obligatory control (NOC) PRO constructions was argued convincingly by Holmberg and Sheehan (2010). Landau (2013: 93–94) argues for the same conclusion. I assume, following these works, that the nature of recovery depends on the nature of unvalued phi-features arriving to the syntax-semantics interface: D_ alone creates the Vainikka-Holmberg-Sheehan signature of the Finnish partial third person control, whereas additional features, such as number and person, require strictly local antecedents. Further distinctions are possible (e.g., logophoric and/or topic-based antecedents), but not required for deriving null argument behavior in the present dataset. What happens if an unvalued phi-set cannot be valued even at the syntax-semantics interface? This situation occurs if morphosyntactic valuation fails and recovery finds nothing. I will assume that if a phi-feature remains unvalued, a generic interpretation, referring to people in general, is created as a last resort strategy.

An obligatory control structure (OC) arises under this analysis if a lexical item has an unvalued phi-set but neither thematic subject argument nor agreement suffixes can be projected. This results in a sentence that cannot host a subject argument (a property still to be captured) or generate agreement suffices but still requires an antecedent. Sentence (17) provides two examples of obligatory control structures in English.

(17)

a. John began (*Mary) to leave$_{\varphi_}$.

b. John tried (*Mary) to leave$_{\varphi_}$.

The class is syntactic: even when a thematic subject argument is available, it can co-refer with the main clause agent (compare *John wanted himself to resign* vs. *John wanted to resign*). It cannot be, then, that only *began* (and not *want*) is compatible with a 'reflexive meaning' in which the main clause subject and the embedded subject denote the same thing. On the other hand, the meaning of *began* does imply that the 'agent of beginning' and the 'agent of the event that thereby begins' are connected conceptually: it is not possible to begin something if the 'agent of doing' is separated conceptually from the 'agent of beginning' (see Farkas 1988 for a similar proposal that has inspired the present approach). Furthermore, it is not required that the

[5] I do not unify antecedent recovery with anaphora resolution in the current work; the proposed algorithm does not process reflexives. That hypothesis constitutes an interesting idea worth exploring in future work.

two agents are the same; only that they cannot be separated conceptually. This is due the existence of partial control, which requires internal conceptual connection without identity (Landau 2000; Wilkinson 1971).[6]

We still have to formalize these assumptions, again to anticipate the computational work. The following formalization was adopted. Let us assume that a verb such as *begin* has a lexical feature 'SEM:internal', and a head with the opposite profile (e.g. *persuade*) will have the feature 'SEM:external'. Feature 'SEM:internal' means that the agent of the verb and the selected infinitival must be linked conceptually; 'SEM:external' triggers the opposite behavior, forcing non-conceptual, external linking. Thus, the intuition that *begin* requires that its agent is connected conceptually with the agent of the complement clause is formalized by assuming that *begin* has a lexical feature 'SEM:internal', whereas *persuade* has the opposite property 'SEM:external'. The verb *want* has neither feature and will be compatible with both interpretations (*John wants to leave*, *John wants Mary to leave*). A formal mechanism is then required for connecting the presence of this lexical feature to the absence of an independent thematic role inside the vP/VP of the selected infinitival. Let us assume that the relevant feature is ±ARG such that −ARG renders the selected VP unable to project a separate thematic argument to its specifier while +ARG forces projection of a thematic agent. For example, if the infinitival *to* is marked for +ARG, then the infinitival verb it selects is able to project an independent thematic agent; if the specification is −ARG, we get a truncated structure 'to+V' with no argument between. We can then assume that 'SEM:internal' selects for −ARG (18).

(18)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| a. | John | tries | | | to | | leave. | |
| | | SEM:internal | → | | −ARG | → | V (−SPEC) | |
| b. | John | persuades | | $\text{Mary}_1$ | to | | $\_\__1$ | leave. |
| | | SEM:external | → | | +ARG | → | V (+SPEC) | |

It follows that the unvalued phi-features at the embedded infinitival *leave* in the example (18a) will have to be valued by the main clause subject by recovery. This derives properties of the OC-signature.

To summarize, the SEM-feature creates three classes of control predicates: SEM:internal (*John tried (*Mary) to leave*), SEM:external (*John persuaded *(Mary) to leave*) and neither (*John wanted (Mary) to leave*). Because 'SEM:external' requires that the argument it projects and the argument below are not connected conceptually, the presence of this feature must limit upward antecedent search at the syntax-semantic interface. This restriction was added as a formal condition to recovery: the operation is blocked by the

[6] When the agent of the infinitival is reflexive (*John wanted himself to resign*), conceptual separation is possible, and thus it feels as if the agent targets himself 'externally', in a way in which the identity between the agent of wanting and the agent of resigning is accidental.

presence of ‘SEM:external’ at a head. Symbol v* is used in this study and in the lexical resources for a transitivizer with this feature (e.g. *order* = v* vs. *want* = v).

### *3.4 Mapping sensory inputs into phrase structure objects: left-to-right architecture*

An analysis was proposed that could, at least in principle, work within the parameters set by the inverse framework. Everything that the analysis requires can be put together from overt components available in the input. To show that the proposed system suffices to solve the inverse problem, we must show that the correct antecedent properties are derived from nothing but such inputs and that, furthermore, the system rules out all ungrammatical variations and unattested interpretations. To construct an argument of this kind, certain further requirements must be satisfied. First, we need a function that maps the incoming words into plausible phrase structure objects that contain lexical items with the features and structural configurations presupposed by lexical decomposition, Agree-1 and recovery. To this end, a left-to-right architecture of Phillips (1996, 2003) was used as a starting point. According to this system, Merge can operate also in tandem with reading words from the input in a left to right order. The process is illustrated in (19).

(19) *John* * *admires* * *Mary* (Input sentence read from left to right)
| | |
[John admires] | (First Merge)
[John [ admires Mary ]] (Second Merge…)

The system generates phrase structure objects incrementally from the sensory input. This architecture was then enriched with parsing recursion developed and implemented computationally in Brattico (2019), which reads words from the input in the manner elucidated in (19), creates a search space based on the possible and plausible merge sites in the existing phrase structure, explores the search space in a well-defined order, and consumes the next word until all words have been processed. The parser backtracks if the final output is not well-formed, resulting in a reanalysis phase caused by a garden path. The input is judged ungrammatical if no solution if found, as assumed in the Dynamic Syntax framework of Cann et al. (2005). For present purposes, the important point is that the system will explore all phrase structure interpretations compatible with any input sentence (Brattico 2019: 16-23). It is important to point out, however, that the analysis of null arguments does not force one to use this specific parsing architecture. The analysis could be added to anything that maps sensory inputs into phrase structure objects or other types of syntactic objects sufficiently rich to sustain Agree-1 and recovery.

# 4 Formalization and testing

## *4.1 Introduction*

The analysis elucidated in the previous sections was next formalized in order to verify that it can solve the inverse problem. Specifically, the analysis was added to an existing parsing toolkit written in Python (Brattico 2019), which I then used to automatize all calculations. The existing algorithm did not process agreement and phi-features and had no mechanism for antecedent recovery, so it failed to understand all sentences that contained null arguments.

Formalization of a linguistic theory by means of a machine-readable language is perhaps not common in the field, but it does not differ from regular linguistic formalization; instead, it provides several advantages over purely symbolic formalization. Running calculations becomes a matter of starting one script. Thus, it is possible to test the logical consequences of an analysis automatically over a potentially huge number of test sentences. The second advantage is that a machine-readable formalization requires the researcher to be completely explicit about every assumption, principle, and computational step in the analysis. Since there is no ambiguity in the analysis, verification and replication becomes trivial. Third, the calculations can be done efficiently, with the speed of ~70ms per sentence in the present study. Finally, the whole formalization can be deposited into public domain where it can be shared, examined, criticized, and developed. Thus, the formalization together with all the required files are available online.[7]

The deductions were done in the following way. A *test corpus* was crafted by hand that contained 2512 null argument sentences from Italian, Finnish and English that covered the properties relevant to the proposed analysis and discussed in the preceding sections. The test corpus was a simple text file containing a list of sentences, written without any annotation, together with comment lines explicating their properties for the researcher but ignored by the algorithm. In order to avoid any bias in the selection of the test materials, the relevant properties (pro-drop, agreement, word order, embedding and control) were crossed mechanically to result in 2 x 2 x 2 x 2 x 2 = 32 experimental categories. These categories are summarized in Table 2 and

---

[7] The source code is maintained at www.github.com/pajubrat/praser-grammar. The latest version that adopts most of the analytic solutions proposed in the present article plus many others is contained in the master branch. The version that was used in this study, and which should be used only for replication purposes, is in the branch *null-arguments-and-control-2020* (https://github.com/pajubrat/parser-grammar/tree/Null-arguments-and-control-2020). External files containing the input parameters and output results discussed in this article can be found from the folder *language data working directory/study-3_2020-control*. The derivational log file is too large to be hosted at Github and can be obtained from the author upon request. For a general description of the algorithm and the organization of the materials in this repository, see the documentation in the /*Documentation* folder.

discussed in more detail in the supplementary document. The important point is that these experimental materials cover all empirical phenomena targeted for an analysis.

**Table 2. Structure of the test corpus as classified by a construction type**

| Sentences | Pro-drop | Agreement | Word order | Embedding | Control | Comment |
|---|---|---|---|---|---|---|
| 1-82 | n/a | n/a | n/a | n/a | n/a | Sentences cited in the main article |
| 83-103 | No | Grammatical | Canonical | No | No | Canonical agreement |
| 104-130 | No | Grammatical | Canonical | No | Yes | Canonical control |
| 131-144 | No | Grammatical | Canonical | Yes | No | Canonical embedding |
| 145-171 | No | Grammatical | Canonical | Yes | Yes | Control under embedding |
| 172-246 | No | Grammatical | Noncanonical | No | No | Noncanonical word order |
| 247-1186 | No | Grammatical | Noncanonical | No | Yes | Control with noncanonical order |
| 1186-1261 | No | Grammatical | Noncanonical | Yes | No | Embedding and noncanonical word order |
| 1262-1503 | No | Grammatical | Noncanonical | Yes | Yes | Embedding, control and order |
| 1504-1541 | No | Ungrammatical | Canonical | No | No | Agreement errors |
| 1542-1568 | No | Ungrammatical | Canonical | No | Yes | Control with agreement errors |
| 1569-1600 | No | Ungrammatical | Canonical | Yes | No | Agreement errors and embedding |
| 1601-1627 | No | Ungrammatical | Canonical | Yes | Yes | Control with agreement errors and embedding |
| 1628-1807 | No | Ungrammatical | Noncanonical | No | No | Noncanonical order with agreement errors |
| 1808-2051 | No | Ungrammatical | Noncanonical | No | Yes | Control, agreement and noncanonical order |
| 2052-2081 | No | Ungrammatical | Noncanonical | Yes | No | Embedding, order and agreement errors |
| 2082-2099 | No | Ungrammatical | Noncanonical | Yes | Yes | Control, agreement, order and embedding |
| 2100-2116 | Yes | Grammatical | Canonical | No | No | Basic pro-drop |
| 2117-2143 | Yes | Grammatical | Canonical | No | Yes | Pro-drop with control |
| 2144-2151 | Yes | Grammatical | Canonical | Yes | No | Pro-drop with embedding (partial pro-drop) |
| 2152-2178 | Yes | Grammatical | Canonical | Yes | Yes | Pro-drop with embedding (partial pro-drop) |
| 2179-2194 | Yes | Grammatical | Noncanonical | No | No | Pro-drop with noncanonical word order |
| 2195-2260 | Yes | Grammatical | Noncanonical | No | Yes | Pro-drop, word order, and control |
| 2261-2266 | Yes | Grammatical | Noncanonical | Yes | No | Embedding, word order and pro-drop |
| 2267-2510 | Yes | Grammatical | Noncanonical | Yes | Yes | Pro-drop, order, control and embedding |
| - | Yes | Ungrammatical | Canonical | No | No | n/a (i.e. pro-drop and agreement error) |
| 2511-2512 | n/a | n/a | n/a | n/a | n/a | Miscellaneous items (two adverb tests) |

The algorithm processed the whole corpus when the main script was executed and produced two output files, one which contains the analytical solutions together with null arguments and their antecedents, and another which contains detailed derivational logs (Brattico 2019: 66-69). The correctness of the output was verified by the author; more detailed comments on the verification procedure are available in the supplementary document. Each sentence processed by the model was numbered consecutively by the algorithm during the execution of the main script. Thus, it is possible to find a step-by-step derivation for any sentence by searching for its identifier (e.g. "# 5", "# 2313") from the log file. These numbers and sometimes even the line numbers in the log files are referred to in the main text, so that a reader can find the corresponding entries in the case further details of the computations are needed. Finally, the test corpus is organized so that all sentences discussed in the main article are listed as separate entries in the order of their presentation in the beginning of the test corpus and can thus be found more easily (#1-82). All lexical knowledge, including the relevant features ±VAL, ±PHI and ±ARG, was provided in external files as independent parameters. Lexical elements were decomposed and retrieved on the basis of phonological words in the input, as elucidated in Section 3.2. The methodological framework is illustrated in Figure 1.

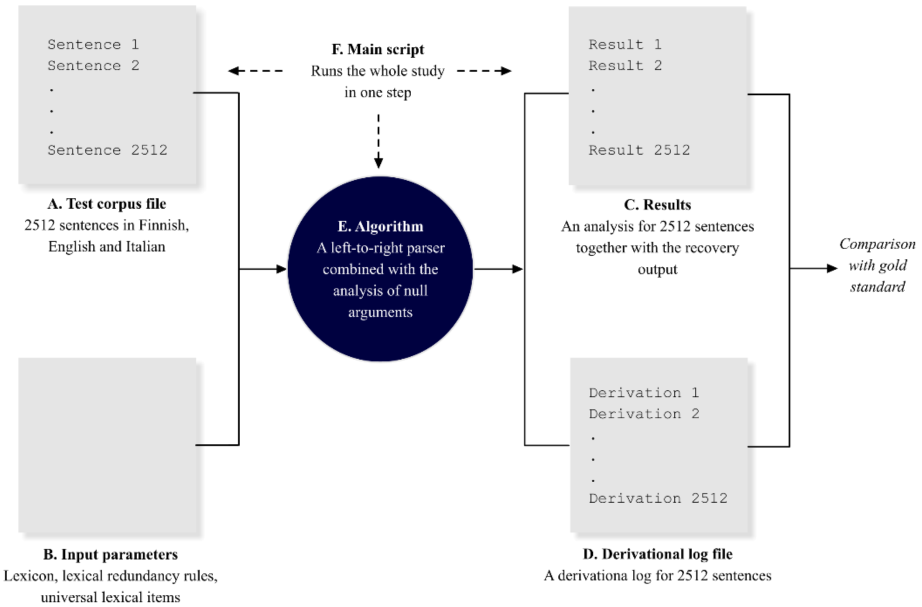



*Figure 1. The general methodological framework. Components A-D constitute the ultimate content of the analysis proposed in this article: the analysis (F), input parameters (A, B), and the raw output data (C, D).*

### *4.2 Testing the model without null subjects*

The computational mechanisms were first tested without the presence of null arguments. Example (20) shows how the formalization deduces subject-verb agreement patterns in Italian when the subject is overt. Both matches (*John admires*) and mismatches (**John admire*) were tested. The same tests were run for all person and number combinations and for all three languages (items #83-103, #1506-1541 in the output). Notice that all the output configurations presented here were generated by the algorithm; they should be viewed as logical consequences of the analysis, not independent analyses proposed by the author.

(20) *Italian*

| | | |
|---|---|---|
| a. | Noi ador-iamo Luisa. | (Input, #101) |
| | Noi$_{1,\varphi}$ [T$_{\varphi}$ [ ___$_{1}$ [ v$_{\varphi_=noi}$ [adora$_{\varphi_=Luisa}$ Luisa]]]]] | (Output) |
| | ‘We admire Luisa.’ | |
| b. | *Io ador-ate Luisa | (Input, #1540) |
| | I admire-2PL Luisa | (Judged ungrammatical) |

The left-to-right algorithm (Section 3.4) produces standard bare phrase structure interpretations for the input sentences. After creating the subject for the input (20a), Steps 1-7, lines 45285-45310 in the derivational log file, the input word *adoriamo* ‘admire-v-T-3pl’ (Step 8) is decomposed into a complex head ‘V-v-T’ by the lexical-morphological component, with T containing features {φ_, φ:1pl}(Step 8). φ_ is provided by the lexicon, {φ:1pl} by the suffix *iamo* extracted from the input and arriving through the morphosyntactic route, as assumed in the analysis (Step 8, line 45312).[8] These grammatical heads are fed individually to the algorithm (Steps 9, 10, 11), where they are ‘repackaged’ into a complex head TvV (see line 45348). The direct object *Luisa* is treated in the same way (Steps 12-18). The result is (21), with “DN” and “φTvV” denoting complex polymorphemic words (D = determiner, N = noun head, T = tense, v = transitive verbal head, V = verb root). This provides a rigorous formalization of the lexical decomposition procedure illustrated in Figure 4.

(21) We admire Luisa. (#101, Step 18, line 45408)
[DN [φTvV DN]]

Complex heads are reverse-engineered by head reconstruction (lines 45411-45415).[9] DP arguments are spread into [D N] (=DP) structures, while the TvV complex generated the familiar head chain. The result is (22).

(22) [We [T [v [admire(V) Luisa]]]] (line 45415)

The algorithm reconstructs the preverbal subject *noi* ‘we’ into the canonical thematic position at SpecvP, as indicated by the reconstructed position $\_\__1$ (lines 45423-4). Agreement reconstruction (valuation by Agree-1 and/or by means of input suffixes) will then be applied to the structure (lines 45426-45433). Heads with unvalued phi-features and +VAL trigger morphosyntactic valuation. Valuation is sought from DP arguments according to (14). The result is (23)(line 45435).

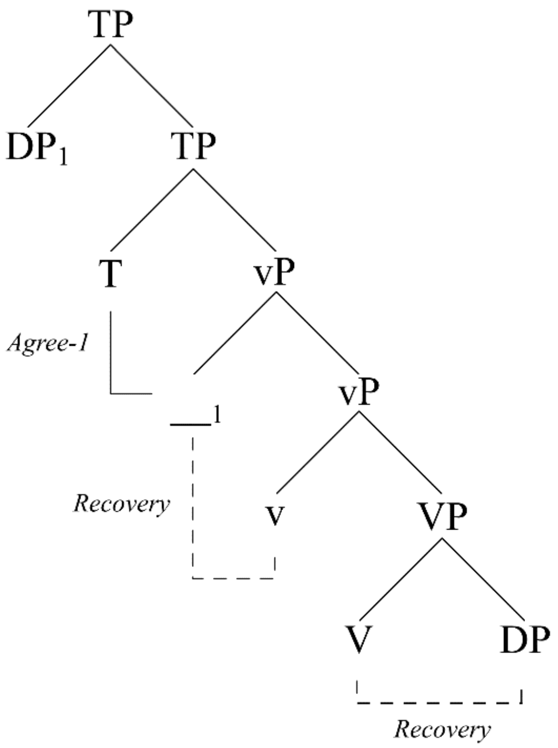
(23) $(\text{Noi}_1)_{\varphi}$ [ $\text{T}_{\varphi\_=1\text{pl}}$ [ $\_\__1$ [ $\text{v}_{\varphi\_}$ [ $\text{adora}_{\varphi\_}$ Luisa]]]]]

Notice that T, v and V are assumed to contain unvalued phi-features, but only T seeks morphosyntactic valuation due to the +VAL feature. Unvalued features at v and V are valued by recovery at the syntax-semantic interface: the antecedent for v is *noi* ‘we’ (lines 45438-9), whereas the closest antecedent for the verb is *Luisa* (lines 45440-1). This provides (24)(line 45448). The antecedent relations can be found also from the results file (entry #101, lines 1125-1131), where they are expressed in a nontechnical way.

[8] Unvalued phi-features are represented as features PHI:NUM:_, PHI:PER:_, PHI:GEN:_ (when applicable) and PHI:DET:_, not as a single feature. They are referred as phi/φ in the main text.

[9] I have described the head reconstruction algorithm in detail in a currently unpublished manuscript “Predicate clefting and long head movement in Finnish,” submitted for publication. See also Brattico (2019:41-43, 86-88).

(24) $(\text{Noi}_1)_{1pl}$ $[T_{\varphi=1pl}$ $[\ \_\__1$ $[v_{\varphi\_=noi}$ $[\text{adora}_{\varphi\_=Luisa}$ Luisa]]]]]




Recovery begins from the element triggering the operation and by examining if its sister can provide an antecedent; if not, the operation is applied iteratively to the mother. It tries to establish an ‘upward path’ from the triggering element into an antecedent. The agent argument for v is identified and reconstructed by means of control: φ_ links with the closest DP at SpecvP, if present. If V contains an unvalued phi-set and does not trigger Agree-1, recovery will associate it potentially with the complement DP, as is the case here. Thus, one consequence of the present analysis is that there occurs a control relation between the verb and its complement in (24). If this control relation were ignored, for example by means of a formal stipulation, the verb would target a nonlocal antecedent and create a reflexive meaning ‘we admire ourselves’.[10]

Sentences containing subject-verb agreement were processed in all three languages, both grammatical (#83-103) and ungrammatical (#1506-1541) combinations. In the actual testing, agreement sentences were also crossed with other structural variables, such as embedding, word order and control. For example, it was verified that the agreement mechanism did not interact in an incorrect way with (*wh*-) operator movement. This was checked by feeding the algorithm with Finnish sentences (25-26), which it handled correctly.

[10] It is interesting to speculate in this connection if the generalized recovery algorithm could be used to eliminate some standard structural locality conditions such as CompVP and SpecVP from the syntax-semantics interface. I leave this question for future research.

(25) *Finnish*

| | | | |
|---|---|---|---|
| Kuka | ihaile-e | Merja-a? | (#36) |
| who.NOM | admire-s | Merja-PAR | |

‘who admires Merja?’

(26) *Finnish*

| | | | | |
|---|---|---|---|---|
| Ketä | hän | ihaile-e | __? | (#37) |
| who.PAR | he.NOM | admire-s | | |

‘Who does he admire?’

These matters are discussed in the supplementary document. Overall, however, these tests provided that the basic mechanisms were operating correctly in connection with sentences that did not involve null arguments. The system is able to process input sentences in a meaningful way, and both Agree-1 and recovery were working as intended. These tests verified also that the parser component (Section 3.4) was doing what it was supposed to do.

### *4.3 Pro-drop*

The model was next examined in connection with subjectless sentences in each of the three languages and in each person and number combination (items #2100-2116 in the output). The key results are summarized in (27). The first line contains the input sentence, second is an English gloss provided by the author, and the third is the output provided by the model and simplified by the author. The output should be interpreted as representing logical consequences of the analysis, not independent and freely modifiable solutions generated by the author.

(27)

a. *Finnish, pro-drop*

| | | |
|---|---|---|
| Ihaile-n | Merja-a. | (#2100) |
| admire-1SG | Merja-PAR | (Gloss) |
| [T$_{1sg}$[v$_{\varphi_=1sg}$ [admire$_{\varphi_=Merja}$ Merja]]] | | (Simplified output) |

‘I admire Merja.’

b. *Finnish, third person pro-drop*

| | | |
|---|---|---|
| *Ihaile-e | Merja-a. | (#2102) |
| admire-3SG | Merja-PAR | |

(No parsing solution found.)

c. *Italian, third person pro-drop*

| | | |
|---|---|---|
| Ador-a | Luisa. | (#2112) |
| admire-3SG | Luisa | |

$[T_{3sg}[v_{\varphi\_=3sg}$ $[adora_{\varphi\_=Luisa}$ Luisa]]]    (Simplified output)
'He admires Luisa.'

d. *Admire Mary    (#2106)
(No parsing solution found.)

e. *Admires Mary    (#2107)
(No parsing solution found.)

The model is able to deduce pro-drop clauses correctly in Italian and Finnish, whereas they are correctly rejected in English.

It is not self-evident that the syntactic analysis of the Finnish and Italian sentences generated by the algorithm is the correct one. It has been argued, convincingly, that Finnish finite verbs have an EPP feature that requires their specifier positions to be filled in by a syntactic phrase (Holmberg and Nikanne 2002; Vilkuna 1989; Vainikka 1989, Huhmarniemi 2019). There is no such phrase at the preverbal subject position in the pro-drop sentences above. To understand how the analysis solves this issue, consider the English examples (27d-e) first. If neither the agreement affix nor an overt subject argument is present, as is the case here, then, all being equal, the unvalued phi-set should remain unvalued and create a generic interpretation. The algorithm judges these sentences ungrammatical instead. The reason is because we have assumed that *only non-conflicting phi-features* can generate null pronouns and thus check the EPP. The morphosyntactically impoverished verb form is unable to determine an unambiguous conflict-free pronoun: *admire* creates conflicts in two types of phi-features, matching with the two person features (first and second: *I/we/you admire*) and two number features (singular and plural: e.g., *I/we admire*), whereas *admires* conflicts with gender features (*she/he admires*). These clauses are therefore judged ungrammatical by the algorithm because there is nothing to check the EPP. The same reasoning applies to Finnish, but the outcome differs: the model does impose an EPP requirement to T but satisfies it by constructing an unambiguous pronominal element from the agreement suffixes extracted from the input. Thus, no EPP violation is marked in the derivational log. For example, *ihaile-n* 'admire-1sg' reconstructs an unambiguous first-person singular pronoun (gender being not grammaticalized in this language).[11]

[11] The EPP condition is usually defined as requiring that some nominal feature, such as D, N, φ or Case, must be checked at SpecTP (see, for example, Chomsky 1995). Holmberg & Nikanne (2002) assumes that in Finnish T checks a discourse feature [non-focus] from its specifier position. The algorithm uses a specifier selection feature forcing a phrase with some feature (e.g., D, non-focus) to occur at the specifier position (Brattico 2019: 23-26, 63-65). This alone is insufficient because agreement features at the head can also satisfy the same condition. Therefore, the notion of "specifier of head H" was generalized so that it refers both to the specifier and (consistent) pro-elements inside the head. In later iterations of the algorithm, the notion of specifier was replaced with the notion of edge.

The Finnish third person pro-drop construction (28b) is correctly deduced as ungrammatical. The third person suffix present in the input values the number and person features at T (lines 1446497-9), but D_ remains unvalued and triggers recovery (line 1446506). No antecedent is found, and the input is judged ungrammatical (line 1446507). The mechanism was further tested with a complex clause in which a nonlocal antecedent was available. The model handles these correctly, as shown in (28).

(28) *Finnish*

| | | | | | | |
|---|---|---|---|---|---|---|
| a. | Pekka | sano-o | että | ihaile-e | Merja-a. (#25, 43) | |
| | Pekka | say-3SG | that | $T_{D\_=Pekka}$-3Sg | Merja-PAR | (D_ valued by nonlocal antecedent) |
| | ’Pekka says that he (=Pekka) admires Merja.’ | | | | | |
| b. | *Minä | sanon | että | ihaile-e | Merjaa. (#44) | |
| | I | say-1SG | that | $T_{D\_}$-3SG | Merja-PAR | (D_ remains unvalued) |
| | ‘I say that (he) admires Merja.’ | | | | | |

Example (28b) is ruled out because the phi-features of the antecedent do not match with the phi-features at the probe. Notice that in a sentence in which there are two potential c-commanding antecedents, this version of the model can only select the local one:

(29) *Finnish*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Pekka | sanoo | että | Jukka | sanoo | että | ihaile-e | Merja-a. (#45) |
| Pekka | says | that | Jukka | says | that | admire-$s_{D\_=Jukka}$ | Merja-par |

‘Pekka says that Jukka says that he (=Jukka) admires Merja.’

To me, the non-local argument is marginally possible when the semantic interpretation leans towards such interpretation.[12] Thus, nonlocal antecedents are registered and shown in the logs, but they are not selected.

Let us consider Finnish generic null pronouns next. The phenomenon is illustrated in (30) and has been discussed especially by Holmberg (2010), whose work I rely on here.

(30) *Finnish*

| | | | | | | |
|---|---|---|---|---|---|---|
| a. | Pekka | sanoo | että | [istu-u | mukavasti.] | (#46) |
| | Pekka | says | that | sit-3SG | comfortable | |
| | ’Pekka says that he (=Pekka) sits here comfortably.’ | | | | | |
| b. | *Istu-u | tässä | mukavasti. (Holmberg 2010, ex. 1a) | | | (#47) |

[12] ?*Murhaaja*$_1$ *uskoi yhä että poliisi*$_2$ *luulee että pro*$_{1,2}$ *ei ole syyllinen* ’murderer.nom believes still that police.nom thinks that _ not be guilty’, i.e. the murderer claimed that the police thinks that he (=murderer, not the police) is not guilty.

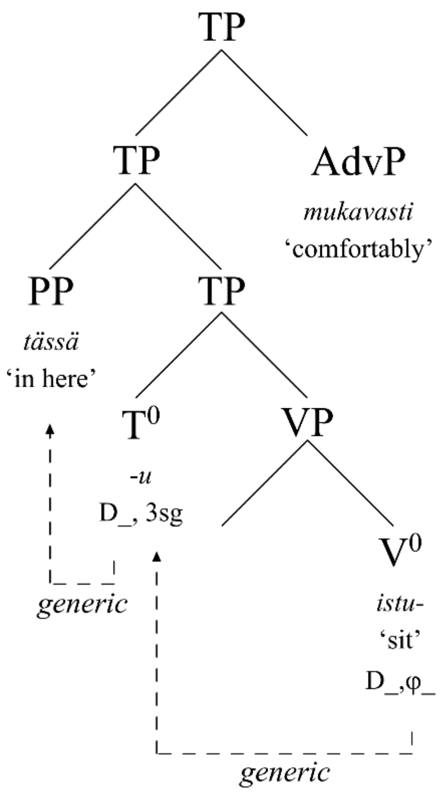
sit-3SG here comfortably

c. Tässä istu-u mukavasti. (#48)
here sit-3SG comfortably
'People sit here (e.g., in this chair) comfortably.'

d. Pekka sanoo että tässä istu-u mukavasti. (#49)
Pekka says that here sit-3SG comfortably
'Pekka says that one can sit here comfortably.'
*'Pekka says that he (=Pekka) sits here comfortably.'

Example (30a) exhibits recovery: D_ remains unvalued, triggers antecedent search, which then finds *Pekka* from the main clause. Example (30b) shows that a third person pro-drop without antecedent leads into ungrammaticality: D_ remains without an antecedent. The interesting example is (30c). The third person pro-drop is not ungrammatical when the preverbal subject position is filled in by a locative PP, but the interpretation comes out as generic. Control is not possible; PP intervenes recovery (30d). The model deduces these properties correctly. Virtually any phrase can (as long as it constitutes the topic) satisfy the EPP condition of Finnish, hence the locative PP will do as well (see footnote 11). This PP cannot, however, value D_ at T by Agree-1 because it is not a DP. The D-feature therefore enters the syntax-semantics interface unvalued, triggers recovery and finds the PP, which leads to the generic interpretation. The analysis is shown in (31).

(31)



This deduction is not uncontroversial. No generic null pronoun fills in an argument position; the generic interpretation is created at the syntax-semantic interface as a last resort. Holmberg (2010) considers a number of arguments suggesting that a generic pronoun must be present. These arguments rely on the fact that anaphoric elements, such as reflexives, possessive suffix, or adverbial null arguments can take the generic agent as their antecedent. The author then immediately notices, however, that whether this argument goes through depends on many independent assumptions concerning anaphor binding. The fact that an ordinary null subject in Finnish can function as a regular antecedent means that under the present analysis any consistent pronominal phi-set inside a head must be able to function as an antecedent. Indeed, heads with valued phi-features are accepted as antecedents by the recovery algorithm.

The analysis has no derivational path for interpreting an object pro construction. A sentence such as (32) will be classified as ungrammatical because the necessary complement selection feature of the main verb *want* is not satisfied.

(32) *Finnish*

*Minä halua-n. (#50)
I.nom want-1SG
'I want.'

There is no deductive path for generating a pronominal complement from the phi-set of a head, and even if there were, there are no such features at verb due to the lack of object-verb agreement.

Radical pro-drop languages license null arguments despite exhibiting no agreement morphology. The present analysis licenses a null argument (pro) when a consistent pronominal element constructed from phi-features residing in a head satisfies conditions involved in checking the presence of a phrasal subject. In the examples analyzed so far, the mechanism has relied on phi-features extracted *from the input*. On the other hand, if a head has phi-feature(s) as a lexical property, then a null argument with those features will be generated. The crucial assumption is that complete loss of agreement must eliminate conflicting phi-features from the lexicon (contrary to what is the case in Swedish, English and French).[13] Holmberg (2005) seems to propose something similar when he claims that radical pro-drop languages have lost unvalued phi-features and have "unspecified" (null) subjects that check syntactic conditions. Another possibility is that radical pro-drop languages project unvalued phi-features that are linked to arguments via recovery that accesses discourse, perhaps in a way comparable to Finnish third person null argument. To show that the analysis works in principle I constructed an imaginary 'radical pro-drop English' in which all agreement has been lost, and

[13] Conflicting or non-conflicting phi-features are required as long as some subject-verb combinations are ruled out, otherwise there is no way of pairing specific subjects with specific verb forms. Once all agreement is lost, all subject-verb combinations are possible and selective phi-features can be eliminated.

lexical elements have no conflicting phi-features. I tested both hypotheses, one in which the imagined finite verbs had only valued person feature, corresponding to the first hypothesis, and another in which they had an unvalued person feature, corresponding to the second (person feature was selected only for testing purposes). Since no agreement was extracted from the input, no phi-feature conflict could arise and thus both constructions admitted pro-drop (see #79-82 in the output, with imaginary verbs *admire´* (with unvalued person feature) and *admire´´* (with lexically valued person feature), both found from the lexicon file). The four possible routes for licensing null arguments predicted to exist on the basis of the present analysis are summarized in Table 2.

**Table 2. Four routes for generating pro-drop sentences.**

| | OVERT AGREEMENT | |
|---|---|---|
| LEXICALLY VALUED PHI-FEATURES | No (−VAL) | Yes (+ VAL) |
| No | Null argument by recovery (control, discourse antecedent, depending on which phi-features remain unvalued). If no antecedent is found, interpretation is generic. | Null argument by rich agreement (Italian), with partial recovery possible (Finnish, Hebrew) and conflicting phi-features blocking the null argument (Swedish, English, French). |
| Yes | Null argument by lexically valued phi-feature(s), interpretation depends on the nature of φ (Chinese, Korean, Japanese?). | Null argument licensed by both overt agreement and lexical φ-features, which must of course match with each other. Corresponds best to ‘frozen agreement’ that does not and cannot trigger recovery and control. |

Notice that no radical pro-drop languages were included into the test corpus, and hence these remarks should be considered as speculative or tentative at most.

### *4.4 Control*

Let us next consider how the model deduces the properties of control. Consider the following examples of standard control (33). The second line presents the (simplified) output generated by the model, not by the author.

(33)

a. John wants to leave.

[John$_1$ [T$_{\varphi_:3sg}$ [ ___$_1$ [V$_{\varphi_=John}$ [to leave$_{\varphi_=John}$ ]]]]] (#51)

b. John wants Mary to leave.

[John$_1$ [T$_{\varphi_:3sg}$ [ ___$_1$ [v$_{\varphi_=John}$ [V$_{\varphi_=John}$ [Mary$_2$ [to [___$_2$ leave$_{\varphi_=Mary}$]]]]]]]] (#52)

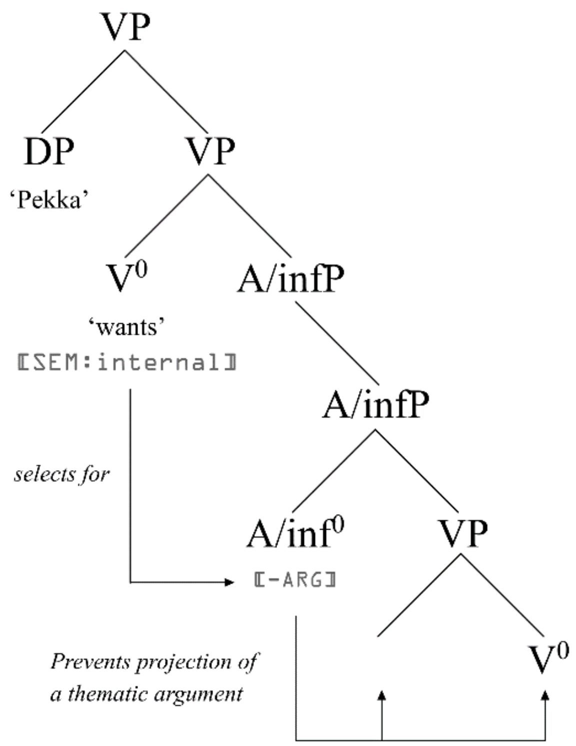
T values its phi-features from the sensory input. If we leave the subject out, the input evaluates as ungrammatical due to an EPP failure at T. The small verb v finds an antecedent at SpecvP into which the model reconstructs the subject (the chain marked by subscript 1). The lower verb *leave* is linked with *Mary* in (b). The algorithm, therefore, deduces the correct antecedent properties. Moving to more complex examples, (34) represent the core cases of Finnish control and cover the slightly more complex selection and selectee dependencies discussed in Section 3. The grammaticality judgments and analyses are deduced by the model, not by the author; the model judgment however matches with native speaker judgment. This dataset and the results are further explained below.

(34) *Finnish*

| | | | | | | |
|---|---|---|---|---|---|---|
| a. | Pekka halusi | | [lähte-ä.] | | | (#57) |
| | Pekka wanted leave-A/INF | | | | | |
| | 'Pekka wanted to leave.' | | | | | |
| b. | *Pekka | halusi | [Merja-n | lähte-ä.] | | (#58) |
| | Pekka | wanted | Merja-GEN | lähte-A/INF | | |
| c. | Pekka | käski | [Merja-n | lähte-ä.] | | (#59) |
| | Pekka | ordered | Merja-GEN | leave-A/INF | | |
| | 'Pekka ordered Merja to leave.' | | | | | |
| d. | *Pekka | käski | [lähte-ä.] | | | (#60) |
| | Pekka | ordered | leave-A/INF | | | |
| e. | Pekka | halusi | [Merja-n | lähte-vän.] | | (#61) |
| | Pekka | wanted | Merja-GEN | leave-VA/INF | | |
| | 'Pekka wanted Merja to leave.' | | | | | |
| f. | *Pekka | yritti | [Merja-n | lähte-vän.] | | (#62) |
| | Pekka | tried | Merja-GEN | leave-VA/INF | | |
| g. | *Pekka | yritti | [lähte-vän.] | | | (# 63) |
| | Pekka | tried | leave-VA/INF | | | |
| h. | *Pekka | yritti. | | | | (#64) |
| | Pekka | tried | | | | |
| j. | Pekka | uskoo | lähte-vä-nsä. | | | (#65) |
| | Pekka | believes | leave-VA/INF-PX/3SG | | | |
| | 'Pekka$_1$ believes that he$_{1,*2}$ will leave.' | | | | | |
| k. | *Pekka | uskoo | lähte-vän. | | | (#66) |
| | Pekka | believes | leave-VA/INF | | | |

The Finnish verb *haluta* 'want' is marked for 'SEM:internal' (in contrast to English *want*), hence it disambiguates the A-infinitival into –ARG (34a-b). This prevents the underlying VP from projecting an external argument, as shown in (35).

(35)




Sentence (34b) is judged as ungrammatical, as the argument *Merja-n* ‘Merja-gen’ reconstructs into SpecVP but will be left without thematic role due to the selecting –AGR head. The derivation crashes at the syntax-semantics interface without interpretation. The verb ‘order’ (34c-d) is marked for ‘SEM:external’ and disambiguates the selected infinitival into +ARG. The embedded verb projects a thematic role to its specifier and a DP argument may appear there and is interpreted at the syntax-semantics interface as the agent of the infinitival event. The VA-infinitival (34d-e) is marked lexically for +ARG, hence a separate subject occurs inside the infinitival. Examples (34f-h) are ungrammatical because ‘try’ requires an obligatory A-infinitival complement and is marked for SEM:internal. Examples (34i-j) illustrate the effects of infinitival (possessive) agreement, glossed as px/3sg. The VA-infinitival can be used as a subject control verb if it exhibits overt agreement (34i); without agreement it requires an overt argument (34j). The explanation for (34j) parallels Finnish finite partial control: overt infinitival agreement leaves D_ unvalued and recovery targets the main clause subject as an antecedent. If there is no agreement, the VA-infinitival head cannot check its EPP feature against anything. Notice that since the VA-infinitival does exhibit overt agreement, it is marked for +VAL. In Finnish, many infinitivals exhibit overt phi-agreement (referred to as “possessive agreement” in much of the relevant literature).

Example (36) shows noncanonical but grammatical word orders in connection with control that the model also deduces correctly. The model solves this problem by applying phrasal reconstruction before Agree-1 and recovery. These data show that the algorithm solves the problem mentioned in Section 2.1.

(36) *Finnish*

| | | | | | |
|---|---|---|---|---|---|
| a. | Huomenna | halua-a | lähte-ä | Pekka. | |
| | tomorrow | want-3SG | leave-A/INF | Pekka.NOM | |
| | 'Tomorrow Pekka wants to leave.' | | | | |
| b. | Huomenna | käske-e | Merja-n | lähte-ä | Pekka. |
| | tomorrow | order-3SG | Merja-GEN | leave-A/INF | Pekka.NOM |
| | 'Pekka orders Merja to leave.' | | | | |
| c. | Huomenna | käske-e | Pekka | Merja-n | lähte-ä. |
| | tomorrow | order-3SG | Pekka.NOM | Merja-GEN | leave-A/INF |
| | 'Pekka orders Merja to leave.' | | | | |
| d. | Huomenna | käske-e | Merja-n | Pekka | lähte-ä. |
| | tomorrow | order-3SG | Merja-GEN | Pekka.NOM | leave-A/INF |
| | 'Pekka orders Merja to leave.' | | | | |

Examples (37a-b) below illustrate subject and object control in English. The object antecedent (37a-i) works as expected: the closest antecedent is selected. The reason *persuade* is not compatible with a control clause without an argument is due to the feature SEM:external (37a-ii). Subject control, illustrated by (37b-i), is often regarded as surprising because recovery seems to skip a potential antecedent.

(37)

a. i. John persuades Mary to leave. (#67)
   ii. *John persuades to leave. (#68)

b. i. John promises Mary to leave. (#69)
   ii. John promises to leave. (#70)

There are at least two in principle ways to handle subject control (37b-i) under the present framework. One is to substitute MDP with a principle that distinguishes subject and object control from each other. This could be done by adding further formal criteria to the antecedent selection, allowing recovery to target objects and subjects selectively and specifically. This selection must then be made sensitive to a lexical feature, categorizing any given head as subject- or object oriented (or neutral). An alternative is that the infinitival is merged into a higher structural position in the clause so that the direct object becomes invisible to recovery. The model, when provided with no additional assumptions or mechanisms, adopts the second alternative: it right-adjoins *to leave* into a position in which recovery no longer sees the direct object (for the notion of adjunction assumed in the present algorithm, see Brattico in press). The analyses are shown in (38).

(38)

a. John persuades [Mary to leave$_{\varphi_=\text{Mary}}$] (=(34)a-i, #67)

b. John promises [Mary] ⟨to leave$_{\varphi_=\text{John}}$⟩ (=(34)b-i, #69)

This does provide correct antecedent properties, although the solutions are not unproblematic. Solution (38b) is reminiscent of Larson (1991), who proposed that at the level at which recovery applies the infinitival occurs at high enough adjoined position such that the object antecedent is invisible. The algorithm adopts a Larsonian solution when faced with an input sentence of this type. Analysis (38a) is problematic. The parser is only allowed to create binary branching trees (Brattico 2019: 2.2) and cannot, therefore, create a structure in which the verb *persuade* takes two complements (*John persuades* [*Mary*][*to leave*]). It handles an input of this type in one of two ways. If the verb is forced to select for a DP-complement, then the infinitival will be right-adjoined (*John persuades* [*Mary*] ⟨*to leave*$_{\varphi\_}$⟩) and the antecedent of the infinitival is determined by its structural position so that a higher attachment site will find the main clause subject (38b), lower attachment site the object (discussed below). If, on the other hand, the verb selects for an infinitival (either obligatorily or optionally), *Mary* will be generated inside the infinitival and, although the control properties are derived correctly, there is no selection dependency between the main verb and that subject. This resembles the situation with the Finnish examples in (35c), where we notice a similar mismatch between (unambiguous, unproblematic) syntactic structure and semantic intuition.

Unvalued phi-sets of controlled adverbials are valued by the same mechanism, with antecedent selection being again sensitive to the adverbial attachment site. Consider an input such as (39).

(39) *Finnish*
Pekka$_i$ etsii Merjaa [ PRO$_i$ juost-en.] (#73)
Pekka search-3SG Merja-PAR run-ADV
‘Pekka searches Merja by running.’

The model deduces the correct solution. The adverbial has a feature forcing it to occur at a high position at which recovery can only see the subject. An adverbial phrase that has a feature forcing it to occur inside the verb phrase (e.g. manner adverbials) will put it into a lower position where it could take the direct object as an antecedent, as shown by (40).

(40) Pekka$_2$ ei nähnyt [Merja-a]$_1$ ⟨PRO$_{1,*2}$ kävele-mässä⟩
Pekka not see Merja-PAR walk-MA/INF
’Pekka did not see Merja walking (lit. ‘see Merja *in walk-ing*’)

The model correctly interprets arbitrary or generic control sentences such as (41), with the structure and interpretation provided as given.

(41) To leave is a mistake. (#74)
[to leave]$_1$ T$_{\varphi_}$ __$_1$ be [a mistake]
‘For people in general to leave is a mistake.’

Finite T fails to value the phi-set, being unable to find phi-features from the infinitival clause, and the sentence comes out as generic. The infinitival will be interpreted in the same way. Null objects are also known to evoke the generic interpretation. Null objects are not licensed in English, unlike in Italian and Finnish. The null object construction (42) comes from Finnish.

(42) *Finnish*

| | | | |
|---|---|---|---|
| Pekka | pyytää | [lähte-mään.] | (#75) |
| Pekka | asks | leave-MA/INF | |

'Pekka asks people to leave.'

The transitive verb projects a v*-V structure, which leaves the unvalued phi-features of the main clause verb and the embedded infinitival without value. It cannot recover an antecedent from above v*, which then results in an interpretation in which both the thematic object of 'ask' and the thematic agent of 'leave' become generic, referring to 'people in general'. They remain unvalued throughout the derivation. The sentence is interpreted, by the model and by native speakers, so that Pekka is asking 'people in general' to leave. This interpretation depends on the presence of v* (=v with SEM:external); a verb with v will correctly allow control, as shown in (43):

(43) *Finnish*

| | | | |
|---|---|---|---|
| a. | Pekka | pyytää | laulamaan. |
| | Pekka | asks(v*) | to.sing |
| | 'Pekka asks (people, one) to sing.' | | |
| b. | Pekka | haluaa | laulamaan. |
| | Pekka | wants(v) | to.sing |
| | 'Pekka wants to sing.' | | |

All these sentences are ungrammatical if there is nothing at the complement position of the main verb. Thus, the model does not have a derivational path for generating Finnish/Italian/English sentence such as 'John wants PRO'.

### *4.5 Mixed cases*

The model was tested with sentences containing both pro and PRO (#2117-2143). The algorithm correctly deduces the properties of sentences of the type (44), in which the infinitival control verb must be linked with a null subject antecedent. Third person null subjects are ungrammatical.

(44) *Finnish*

| | | |
|---|---|---|
| a. | Halua-n | lähte-ä. (#76) |
| | want-1SG | leave-A/INF |

‘I want to leave.’

b. *Halua-a lähte-ä. (#77)
want-3SG leave-A/INF
‘He wants to leave.’

c. Pekka$_{3sg}$ sano-o$_{\varphi:3sg}$ että halua-a$_{\varphi_=Pekka}$ lähte-ä$_{\varphi_=Pekka}$. (#78)
Pekka say-3SG that want-3SG leave-A/INF
‘Pekka says that he wants to leave.’

In the example (44a), the model values the phi-features at T by the agreement suffixes and then uses the features as an antecedent for the infinitival. In (44b), the lack of an antecedent for the D_ results in ungrammaticality. Finally, sentence (44c) is correctly judged as grammatical and interpreted so that Pekka constitutes an antecedent of every predicate ‘say’, ‘want’ and ‘leave’.

### *4.6 Finite control, tense and agreement*

Finnish third person null subject exhibits nonlocal control. The null subject is typically bound by a c-commanding argument from the next finite clause up. This could be argued to constitute a form of “finite control,” but Landau (2013) shows that the phenomenon does not exhibit the finite control signature he finds from other languages. The main point of divergence is the fact that the antecedent in Finnish need not be local or perhaps not even constrained by c-command. Furthermore, Landau argues that prototypical finite control signature emerges if and only if the controlled finite clause is deficient either in terms of tense specification (–T+Agr) or agreement specification (+T–Agr) (or both –T–Agr), but this is not true of Finnish, in which the embedded finite clause is, at least on the surface, fully specified both for agreement and tense.

The assumption that deficient agreement leads into control follows directly from the present analysis: an agreementless and argumentless finite verb will trigger recovery. The finite boundary does not limit recovery, so the data is captured. If T is marked lexically for –VAL, we derive finite OC signature. The situation with Landau’s –T+Agr alternative is not straightforward, however, and raises further questions that I cannot answer satisfactorily on the basis of the current analysis. If +Agr denotes full agreement in such constructions, these data constitute a fundamental difficulty because full agreement should, all else being equal, block control, leaving nothing unvalued at the syntax-semantics interface. In addition, there is nothing in the present analysis that connects tense with control. What makes the matter nontrivial is the fact that in Finnish what looks to be full agreement still triggers recovery, here assumed to be due to the fact that the underlying agreement is still deficient. Perhaps Landau’s –T+Agr condition creates similar deficient agreement. This problem is left for future research.

### *4.7 Summary*

To summarize, the algorithm classifies various control constructions in the manner depicted in Figure 2. If an antecedent cannot be found, interpretation comes out as generic.

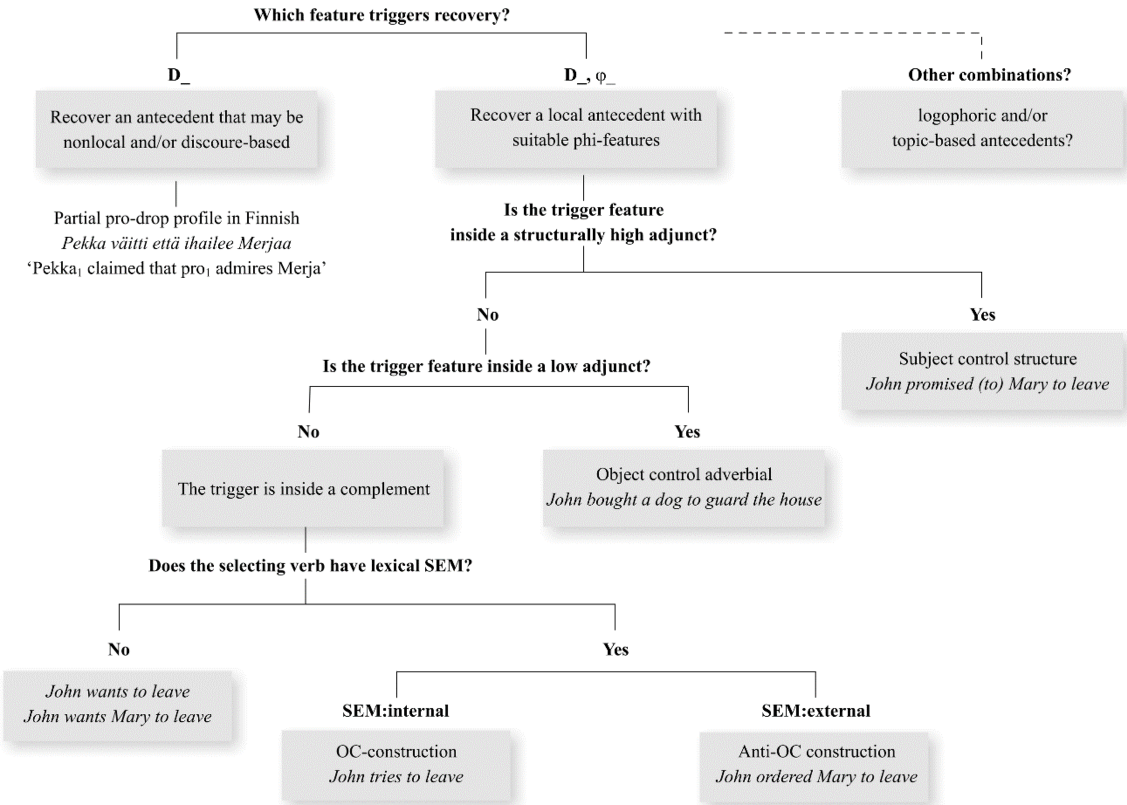



**Figure 2. Classification of control constructions by the algorithm.**

The classification is not meant to be interpreted as a proposal for a complete or even correct classification of all control constructions; rather, it should be viewed as a logical implication of the analysis that was designed as one possible solution to the inverse problem. Any change, addition or deletion to the above schema is acceptable within the parameters of the present study as long as the revised algorithm still solves the inverse problem.

## 5 Conclusions

The issue of licensing and recovering null arguments was analyzed from the point of view of sensory input. An analysis was proposed based on the earlier proposals by Borer (1986, 1989), Brattico (2019), Cann et al. (2005), Janke (2008) and Phillips (1996, 2003). The algorithm was able to deduce correct phrase structures, null arguments and their antecedents. Three languages were examined in detail: Italian, English and Finnish, each with different behavior with respect to their null argument behavior (non-pro-drop, consistent pro-drop and partial pro-drop languages).

A finite null subject pronoun occurs when a grammatical condition normally satisfied by an overt pronoun is satisfied by a consistent phi-set, a pro-element, residing inside a head. The phi-set may emerge from the input (as an agreement suffix) and/or from the lexicon, and it shall not contain conflicting features in order to function as a pronominal element. A controlled null argument is generated during the derivation when an unvalued phi-feature cannot be valued from the resources available in the input. Unvalued phi-features trigger recovery at the syntax-semantics interface, resulting in finite and non-finite control. Finnish partial pro-drop profile together with generic pro-constructions in the same language suggest that the two share a core set of features. Both are involved in the computation of anaphoric dependencies and the creation of the generic interpretation as a last resort. Finally, the analysis does not project phrasal null arguments, hence it reduces the number of syntactic objects stipulated by the analysis.

Although this study was limited to proposing a solution to the inverse problem, it raises the larger question of the possible role of parsing approaches in grammatical theorizing. It would be a mistake reconstruct this question as concerning only the division of labor between competence and performance, as no performance properties (e.g., efficiency, errors, garden-paths, irrationality, suboptimal sensory conditions) were addressed in this work; rather, the issue is whether we can learn something useful about the human language faculty by shifting the perspective from an enumerative approach to that of the inverse recognition problem. My view is that this is an empirical issue. The answer depends on what type of system the language faculty ultimately is. If the principles of the UG are principles of efficient parsing and/or processing of the sensory input – that is, if language is primarily a perceptual system – then we could learn something novel by shifting the perspective; if they are not, then the principles emerging from the inverse framework, if correct in the first place, are likely to receive more elegant formulations within the enumerative approach. It is also possible that the truth falls somewhere between these two extreme approaches. Then a useful approach would be to pursue both approaches while trying to converge them towards one unified theory.

**Abbreviations**

The following abbreviations are used in this article: φ = phi-features such as number and person; φ_ = unvalued phi-features; 1/2/3 = first, second and third person; ADV = adverbial suffix/head (used here in connection with Finnish adverbial suffixes); A/INF = Finnish *A*-infinitival (roughly a ‘desirative’ *to*-infinitival); ACC = accusative case; EPP = extended projection principle; GEN = genitive case; MA/INF = Finnish *MA*-infinitival; NOM = nominative case; PAR = partitive case; PL = plural; SG = singular; VA/INF = Finnish *VA*-infinitival (a ‘propositional’ infinitival complement).

**Supplementary files**

Supplementary file 1. Algorithm and methodology.

**Funding Information**

This research was conducted under the research project "ProGraM-PC: A Processing-friendly Grammatical Model for Parsing and Predicting Online Complexity," funded internally by IUSS (Pavia).

**Acknowledgements**

The author would like to thank three anonymous *Glossa* reviewers for their valuable and constructive criticism that led to a number of significant improvements. Anders Holmberg and Cristiano Chesi provided feedback on an earlier version of this work. I would also like to acknowledge the role of the hosting institution (IUSS) that provided me two years of uninterrupted academic freedom to pursue the research reported in this article. This provided me with the exceptional working conditions required to write and test the algorithm.

**Competing interests**

The author has no competing interests to declare.

**References**

Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parametrizing AGR: Word Order, V-Movement and EPP-Checking. *Natural Language & Linguistic Theory* 16. 491–539.

Barbosa, Pilar P. 1995. *Null subjects*. Cambridge, MA: MIT dissertation.

Barbosa, Pilar P. 2009. Two kinds of subject Pro. *Studia Linguistica* 63. 2–58.

Boeckx, Cedric & Norbert Hornstein. 2004. Movement under control. *Linguistic Inquiry* 35. 431–52.

Borer, Hagit. 1986. I-Subjects. *Linguistic Inquiry* 17(3). 375–416.

Borer, Hagit. 1989. Anaphoric AGR. In O. Jaeggli & K. Safir (eds.), *The Null Subject Parameter*, 69–109. Amsterdam: Kluwer.

Brame, Michael K. 1976. *Conjectures and refutations in syntax and semantics*. Amsterdam: North-Holland Publishing Company.

Brattico, Pauli. 2017. Null subjects and control are governed by morphosyntax in Finnish. *Finno-Ugric Languages and Linguistics* 6. 2–37.

Brattico, Pauli. 2019. *Computational implementation of a linear phase parser. Framework and technical documentation (version 6.x).* Technical software documentation, IUSS, Pavia.

Brattico, Pauli. In press. Finnish word order: Does comprehension matter? *Nordic Journal of Linguistics*

Cann, Ronnie, Ruth Kempson & Lutz Marten. 2005. *The dynamics of language: An introduction (Syntax and Semantics, Volume 35)*. Amsterdam: Elsevier Academic Press.

Cardinaletti, Anna. 2004. Toward a Cartography of Subject Positions. In Luigi Rizzi (ed.), *The structure of CP and IP. The cartography of syntactic structures*, 115–65. Oxford: Oxford University Press.

Chomsky, Noam. 1980. On binding. *Linguistic Inquiry* 11. 1–46.

Chomsky, Noam. 1981. *Lectures in Government and Binding: The Pisa lectures*. Dordrecht: Foris.

Chomsky, Noam. 1982. *Some concepts and consequences of the theory of Government and Binding*. Cambridge, MA.: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The Framework. In Roger Martin, Davic Michaels & Juan Uriagereka (eds.), *Step by Step: Essays on Minimalist Syntax in honor of Howard Lasnik*, 89–156. Cambridge, MA.: MIT Press.

Chomsky, Noam. 2008. On phases. In Carlos Otero, Robert Freidin & Maria-Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, 133–66. Cambridge, MA.: MIT Press.

Culicover, Peter & Ray Jackendoff. 2001. Control is not movement. *Linguistic Inquiry* 32. 493–512.

Farkas, Donka. 1988. On obligatory control. *Linguistic Inquiry* 11. 27–58.

Holmberg, Anders. 2005. Is there a little pro? Evidence from Finnish. *Linguistic Inquiry* 36 (4). 533–64.

Holmberg, Anders. 2010. The null generic subject pronoun in Finnish: A case of incorporation. In Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.), *Parametric variation: Null subjects in Minimalist Theory*, 200–230. Cambridge: Cambridge University Press.

Holmberg, Anders, Aarti Nayudu & Michelle Sheehan. 2009. Three partial null-subject languages: A comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica* 63. 59–97.

Holmberg, Anders & Urpo Nikanne. 2002. Expletives, subjects and topics in Finnish. In Peter Svenonius (ed.) *Subjects, expletives, and the EPP*, 71–106. Oxford: Oxford University Press.

Holmberg, Anders & Michelle Sheehan. 2010. Control into finite clauses in partial null-subject languages. In Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.), *Parametric variation: Null subjects in Minimalist Theory*, 125–52. Cambridge: Cambridge University Press.

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

Hornstein, Norbert. 2001. *Move! A minimalist theory of construal*. Malden, USA: Wiley-Blackwell.

Huang, C. T. James. 1989. Pro-drop in Chinese: A generalized control theory. In Osvaldo Jaeggli & Kenneth J. Safir (eds.), *The Null Subject Parameter*, 185–214. Dordrecht: Kluwer Academic Publishers.

Huang, C T James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15. 531–74.

Huhmarniemi, Saara. 2019. The movement to SpecFinP in Finnish. *Acta Linguistica Academica* 66 (1). 85–113.

Huhmarniemi, Saara & Pauli Brattico. 2015. The Finnish possessive suffix. *Finno-Ugric Languages and Linguistics* 4 (1–2). 2–41.

Hyams, N. 1989. The null subject parameter in language acquisition. In Osvaldo Jaeggli & Kenneth Safir (eds.), *The null subject parameter*, 215–38. Kluwer.

Jaeggli, Osvaldo. 1980. *On some phonologically null elements in syntax*. MIT, Cambridge, MA.

Jaeggli, Osvaldo & Kenneth Safir. 1989. The null subject parameter and parametric theory. In Osvaldo Jaeggli & Kenneth Safir (eds.), *The null subject parameter*, 1–44. Dordrecht: Kluwer Academic Publishers.

Janke, Vikki. 2008. Control without a subject. *Lingua* 118(1). 82–118.

Kempson, R., W. Meyer-Viol & D. M. Gabbay. 2001. *Dynamic Syntax: The flow of language understanding*. Malden, USA: Wiley-Blackwell.

Koskinen, Päivi. 1998. Features and categories: Non-finite constructions in Finnish. Toronto: University of Toronto dissertation.

Landau, Idan. 2000. *Elements of control: Structure and meaning in infinitival constructions*. Dordrecht: Kluwer.

Landau, Idan. 2003. Movement out of control. *Linguistic Inquiry* 34. 471–98.

Landau, Idan. 2004. The scale of finiteness and the calculus of control. *Natural Language & Linguistic Theory* 22. 811–77.

Landau, Idan. 2013. *Control in Generative Grammar A research companion*. Cambridge: Cambridge University Press.

Larson, Richard. 1991. Promise and the theory of control. *Linguistic Inquiry* 22. 103–39.

Lasnik, Howard. 1991. Two notes on control and binding. In Richard K. Larson, Sabine Iatridu, Utpal Lahiri & James Higginbothman (eds.), *Control and grammatical theory*, 235–51, Cambridge, MA: MIT Press.

Manzini, M. Rita & Leonardo Savoia. 2002. Parameters of subject inflection in Italian dialects. In Peter Svenonious (ed.), *Subjects, expletives, and the EPP*, 157–200. Oxford: Oxford University Press.

Manzini, M. Rita & Anna Roussou. 2000. A minimalist theory of A-movement and control. *Lingua* 110. 409–47.

Martin, Roger. 1996. A minimalist theory of PRO and control. University of Connecticut.

Nelson, Diane C. 1998. *Grammatical case assignment in Finnish*. London: Routledge.

Perlmutter, David. 1971. *Deep and surface constraints in syntax*. New York: Holt, Rhinehart and Winston.

Phillips, Colin. 1996. *Order and structure*. Cambridge, MA.: MIT Press.

Phillips, Colin. 2003. Linear order and constituency. *Linguistic Inquiry* 34. 37–90.

Platzack, Christer. 2004. Agreement and the person phrase hypothesis. *Working Papers in Scandinavian Syntax* 73. 83–112.

Postal, Paul. 1970. On coreferential complement subject deletion. *Linguistic Inquiry* 1. 439–500.

Rizzi, Luigi. 1982. *Issues in Italian syntax*. Amsterdam: Foris.

Rizzi, Luigi. 1986. Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17(3). 501–57.

Rosenbaum, Peter. 1967. *The grammar of English predicate complement constructions*. Cambridge, MA.: MIT Press.

Rosenbaum, Peter. 1970. A principle governing deletion in English sentential complementation. In Roderick Jacobs & Peter Rosenbaum (eds.), *Readings in English transformational grammar*, 220–29. Waltham, MA.: Ginn-Blaisdell.

Sheehan, Michelle. 2006. *The EPP and null subjects in Romance*. Newcastle: Newcastle University dissertation.

Taraldsen, Knut. 1980. On the NIC, vacuous application and the *that*-trace filter. *Unpublished Manuscript, MIT*.

Vainikka, Anne. 1989. *Deriving syntactic representations in Finnish*. Massachusetts: University of Massachusetts Amherst dissertation.

Vainikka, Anne & Yonata Levy. 1999. Empty subjects in Finnish and Hebrew. *Natural Language & Linguistic Theory* 17(3). 613–71.

Vilkuna, Maria. 1989. *Free word order in Finnish: Its syntax and discourse functions*. Helsinki: Finnish Literature Society.

Vilkuna, Maria. 1995. Discourse configurationality in Finnish. In Katalin É. Kiss (ed.), *Discourse configurational languages*, 244–68. Oxford: Oxford University Press.

Wilkinson, Robert. 1971. Complement subject deletion and subset relation. *Linguistic Inquiry* 2. 575–84.