Null arguments and the inverse problem

PAULI BRATTICO

Author affiliations can be found in the back matter of this article

]u[ubiquity press

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 the unvalued phi-features of a lexical element cannot be valued from the information available in the sensory input. The analysis unifies the theory of finite null subject pronouns and control and furthermore interprets null argument sentences without generating phrasal 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 was formalized and tested by means of a computer simulation.

CORRESPONDING AUTHOR:
Pauli Brattico

IUSS University School for Advanced Studies, Piazza della Vittoria n. 15, 27100 Pavia, Italy

pauli.j.brattico@gmail.com

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, as shown by (2) these relations depend on the nature of the main verb. Yet, the surface strings do not seem to contain any direct cues for any of these properties. How speakers infer them from the sensory input 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 in the sensory input. They are valued at the syntax-semantics interface, which results in the phenomenon known as control.

The argument 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 without technical details, while Section 4 addresses the details of formalization and subjects the analysis to a rigorous test by means of a 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 null arguments in the context of language comprehension. We consider one null argument, the phonologically null subject pronoun, usually referred to as "pro" in the literature, as an example. As a first approximation, pro replaces finite clause subject pronouns 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   hyvin   italiaa.
               (I)   speak-1SG   well   Italian
               'I speak Italian well.'

        c.   *English*
               *(I) speak Italian.
               Intended: 'I speak Italian.'

We might therefore consider that the native speaker processing the null subject variants infers the existence of the first-person subject/participant 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 intuitively compelling, turns out to be insufficient. Consider (4a–b).

(4)   a.   *Italian* (Cardinaletti 2018: 81)
               È   arrivato Gianni.
               is arrived   Gianni
               'Gianni arrived.'

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 do contain an overt grammatical subject, but it occurs in a noncanonical position at the end of the sentence. The canonical subject position is either empty or contains nonsubject material. For example, while the canonical word order in Finnish is SVO, the preverbal subject 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 in order to judge if a subject is missing and 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 the direct object, not the subject, 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 the first pass parse must perform real parsing with no artificial limit 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 whole comprehension problem, presents a nontrivial challenge.

I will call the problem of inferring properties of null arguments and the corresponding implicit participants from the bare sensory input as the "inverse problem" in this article. Since native speakers can interpret these properties effortlessly from all-new, out of the blue sentences, the model must likewise solve the problem by assuming nothing but an unannotated and contextless sensory input. In addition, the output must contain a list of participants and their thematic roles, correctly matched with native speaker intuition, and no such interpretation is allowed to arise if the input sentence is (judged) ungrammatical (e.g., **admires Mary* cannot be interpreted as 'he admires Mary', **John tries Mary to win* cannot be interpreted as 'John tries to make Mary to win'). Finally, it is required that the solution be presented in the form of a fully formal algorithm – a generative grammar – that can be tested rigorously.

2.2 INTRODUCTION TO NULL SUBJECTS

Before looking at the inverse problem specifically it is useful to gather the basic linguistic facts any solution to the inverse problem must minimally account for. The several decades of work that began with the discovery of null arguments (Rosenbaum 1967; Postal 1970; Rosenbaum 1970; Perlmutter 1971; Brame 1976) and continued during the GB-era and beyond (e.g., Chomsky 1980; Jaeggli 1980; Chomsky 1981; 1982; Rizzi 1982; Hyams 1989) have provided an overall empirical 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 subject-verb agreement (Perlmutter 1971; Taraldsen 1980; Rizzi 1982; Chomsky 1982; Huang 1984; Rizzi 1986; 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, subject pronouns 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. These conditions are illustrated in (5a–c). The second condition distinguishes Finnish from a consistent pro-drop language such as Italian.

(5)   *Finnish*

    a.   *Halua-a   ostaa   kirja-n.
        want-3SG   to-buy   book-ACC
        Intended: '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 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 the null subject requires both formal licensing, a set of grammatical properties determining when it is available, and some type of antecedent recovery mechanism using further linguistic context to infer the properties of the missing subject. This complicates the inverse problem: the acceptability of the third person null subject in Finnish 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 the subject positions of infinitival verbs and gerundive nouns that exhibit neither rich nor impoverished morphosyntax (see, e.g., Rosenbaum 1967; 1970; Postal 1970; Chomsky 1980; 1981; Martin 1996; Hornstein 1999; Landau 2000; Manzini and Roussou 2000; Culicover and Jackendoff 2001; Hornstein 2001; Landau 2003; Boeckx and Hornstein 2004; Landau 2004; 2013). 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 Landau 2013: Chapter 4.1.2).[1] Moreover, the existence of antecedent dependencies in partial pro-drop languages 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 the 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 when the object is null, as shown by (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.NOM   forces-3SG   to.practise   more
        'A failure forces people/one to practice more.'

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

---

[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 algorithm proposed in the present work will not be sensitive to the finite clause boundary, accepting the general line of thought in Landau (2013) that dispenses with the idea that control is confined to infinitival domains.

(8)   *Finnish*
        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 is grammatical (9a–b). This creates obligatory control (OC), in which the antecedent mechanism remains the only strategy pairing the predicate with its argument(s).

(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), glossed as VA/INF in this article.[2]

(10)   *Finnish*
        a.   *Jari   halusi   lähte-vän.
              Jari.NOM   wanted   leave-VA/INF
              Intended: 'Jari wanted to leave.'

        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 thus they are subject to considerable debate in the literature. The solution to the inverse problem must nevertheless capture them.

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 phrasal 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 problem is, instead, how to derive the attested 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.

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

---

2   Finnish A-infinitivals, such as *halu-ta* 'to 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).

be formulated by referring exclusively to overt sensory objects or to a structural interpretation or some higher-order property generated directly from such inputs. The justification for this assumption is that native speakers can provide correct structural and semantic interpretations for sentences in their own language without consulting anything else except 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 a formal framework implementing a *parser* (i.e. a function from sensory objects into sets of semantic interpretations) and an embodiment of an analysis of null arguments within the parser framework. I will begin by formulating the analysis of null arguments and then turn to the implementation of the parser component. Once both 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 the phonological words occurring in the input are matched with lexical items in the surface lexicon. Each lexical item is a set of features. If the phonological word contains several primitive grammatical items, such as tense (T), causativization (v) and a verbal stem (V), these elements are first decomposed and then matched with primitive lexical items. The process is illustrated in *Figure 1*. Merge, discussed later in this article, refers to a computational operation that builds syntactics structures from the incoming lexical items.
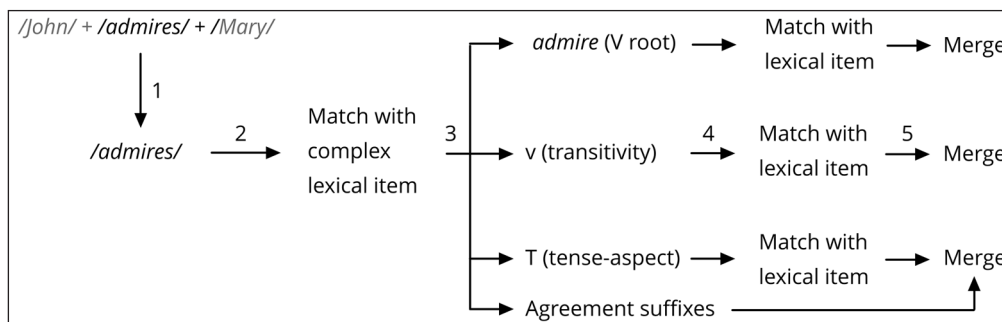

Figure 1 An overview of the computational operations involved in lexical decomposition. The pipeline contains the following steps: (1) consume a phonological word from the input; (2) match the word with a lexical entry or with a list of lexical items if it is ambiguous; if the matched entry is complex, break it into primitive parts (3) and match each with a primitive lexical item (4); Merge (i.e. attach) the lexical item into the syntactic structure. If a component of a word maps to an inflectional affix, it will be interpreted as a feature, not a lexical item. This means that a syntactic 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 sensory input).

To establish a connection between pro-drop and agreement, let us assume that agreement suffixes are extracted from phonological words and ultimately stored as features of lexical items, as shown by the arrow "agreement suffixes" in *Figure 1*. There are two types of phi-features that must be distinguished from each other. Proper names or pronouns retain their phi-features no matter where they occur in the sentence, whereas the corresponding 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 (at least potentially) via the input route. The received generative view, which I will adopt here as a starting point, maintains that finite T has an *unvalued* phi-set (denoted by $\varphi\_$ in this article) that comes to reflect the intrinsic phi-features of a local argument. Whether a lexical item has an unvalued phi-set is determined in the lexicon. 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→ $\varphi\_$
         goal   probe
         Lexical φ Lexical $\varphi\_$
         Argument   Predicate

The mechanism presented 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 extracted from the input (***Figure 1***). 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. The second route is illustrated by the Finnish pro-drop sentence (12).

(12)   *Finnish*
        Ihaile-   n   kovasti Merja-a.
        admire-   1SG   really   Merja-PAR
        φ_   ←   1SG
        'I really 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. Thus, they accomplish the same thing: identify the subject. Notice that no null pronoun is projected to the subject position. This solution agrees with the style of analysis developed by Borer (1986; 1989), Barbosa (1995), Alexiadou and Anagnostopoulou (1998), Manzini and Savoia (2002), Platzack (2004) and Barbosa (2009), all which differ from each other in details but are united by the assumption that agreement suffixes are pronominal enough to assume (at least part of) the role of a full subject pronominal. The analysis disagrees with Holmberg (2005) and most of the literature on Finnish (Vainikka 1989; Vilkuna 1989, 1995; Vainikka and Levy 1999) and other analyses (Cardinaletti 2004; Sheehan 2006) that claim that the clause contains a covert phrasal subject at SpecTP. I will return to this issue later in this article.

The overt phi-cluster at T, such as the first-person suffix *-n* in example (12), 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 in Finnish and is not present in Italian. Holmberg (2005) and Holmberg and Sheehan (2010) propose a solution 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 into the comprehension perspective, we must posit a condition to the effect that the overt Finnish third person agreement suffix values only the person and number features, leaving D_ without value. This would make the third person suffix to behave like a variable, in the sense that it lacks 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 sketched in (13).

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

In essence, then, the recovery mechanism deals with a situation in which an uninterpretable phi-feature passes through the morphological and syntactic component but remains unvalued at the syntax-semantic interface.

What constrains recovery? One condition is c-command: if an argument with a matching phi-set is found that c-commands the deficient feature, that element will be selected as the antecedent (Vainikka and Levy 1999; Holmberg 2005; Holmberg, Nayudu, and Sheehan 2009; Holmberg 2010). Thus, Jari *must* constitute the antecedent in (13). On the other hand, an element that is c-commanded by the variable is never selected. Very few 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) and

---

3   I am assuming that D_ hosts values such as definite, indefinite and generic. This list is obviously not exhaustive.

Brattico (2017). 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 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 null argument reconstruction fails. Conflicting or incoherent phi-sets block pro-reconstruction; otherwise the mechanism is available. If a predicate with an unvalued phi-set contains no valued phi-features, no conflict arises and null arguments are automatically licensed. I will exploit this loophole to capture the properties of agreementless radical pro-drop languages such as Chinese, Japanese and Korean (Huang 1984; Rizzi 1986).

The analysis presupposes a theory of Agree and its inverse version, call it Agree-1, operating with syntactic objects generated from the sensory input. I assume that an unvalued phi-set of a head H can agree with (i) the sister DP of H, (ii) a DP inside its sister, and with (iii) the specifier DP, in this order and with successful match blocking further search. These options are illustrated in (14), with the possible goals for $\varphi\_$ being αP, βP and γP. The operation explores the sister first (βP, γP), then the specifier (αP).

(14)   [SPEC$_\alpha$ [T($\varphi\_$) [$_\beta$...DP$_\gamma$...]]]
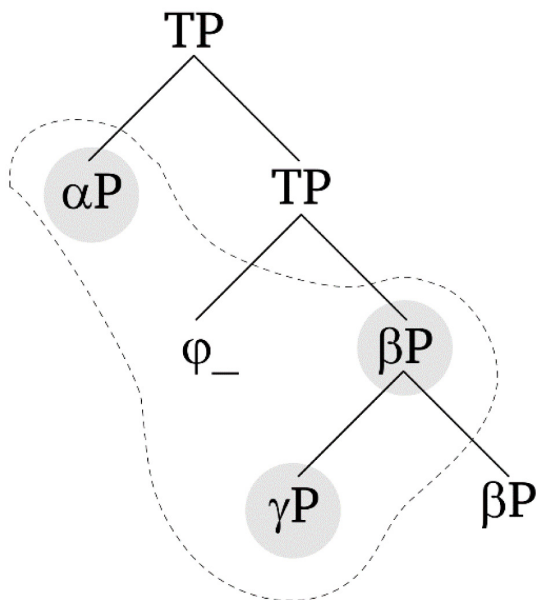
While the theory of Agree is usually formulated so that it only allows (i–ii), large-scale simulations, reported later, showed that condition (iii) might constitute a useful addition when working with linguistic input that will often have arguments at the preverbal specifier positions in the surface string. I will also assume the phase impenetrability condition (PIC) (Chomsky 2000: 108; Chomsky 2008), which prevents Agree-1 from searching below/above vP and CP. This condition 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, I assume that the final phi-set at any given head, resulting from Agree-1 and possible lexical phi-features, must not involve uninterpretable feature conflicts at the syntax-semantic interface.

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

---

[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).

property allowing it to host and express phi-features, triggering the mechanism elucidated above. I assume that this property is represented by a lexico-morphological 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. Furthermore, the feature 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 which will be captured by assuming that these infinitival heads are marked for +VAL. In English and Italian, infinitival heads are marked for –VAL. For present purposes we can think of ±VAL as a lexico-morphological feature triggering 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 determined 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 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 it by searching for an antecedent. *Table 1* summarizes the four types of lexical elements generated on the basis of the two lexical features, ±PHI and ±VAL.

UNVALUED PHI-FEATURES	Overt agreement possible	
	–VAL (no)	+VAL (yes)
–PHI (no)	Frozen particles that do not introduce or are not linked with arguments (e.g., *and*, *but*, *that*).	Words that agree by concord? Concord was not modelled in the present study and will be left for future research.
+PHI (yes)	Predicates linked with arguments that do not exhibit Agree-1. Antecedent recovery becomes mandatory. This category will play a key role in control.	Predicates linked with arguments by valuation or by phi-features, exhibiting Agree-1 (e.g., finite verbs in Finnish, Italian, English).

Table 1 Four types of predicates generated in this study.

The whole analysis has been set up in such a way that it works in principle by assuming nothing but the bare sensory input. There is no stage at which the existence of an invisible/inaudible pro is assumed and/or inferred. Features ±PHI and ±VAL can be retrieved from the lexicon on the basis of the sensory input (*Figure 1*), whereas Agree-1 is formulated in such a way that it applies to a syntactic structure reconstructed directly form the input.

3.3 CONTROL

We have assumed that an unvalued phi-feature occurring at the syntax-semantic interface triggers recovery. In the case of Finnish/Hebrew third person pro, the triggering feature was D_. Let us now generalize this idea 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 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.

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

The infinitival verb *(to) leave* does not exhibit overt agreement and there is no local argument that Agree-1 could target. Feature φ_ will arrive at the syntax-semantic interface unvalued. This will trigger recovery, which selects the closest antecedent, in this case *John* (16). I notate recovery from this point on by writing "φ_=John," with the left side containing the unvalued feature(s) and the right the antecedent selected by recovery.

(16)   John wants$_{\varphi\_=John}$ to leave$_{\varphi\_=John}$.
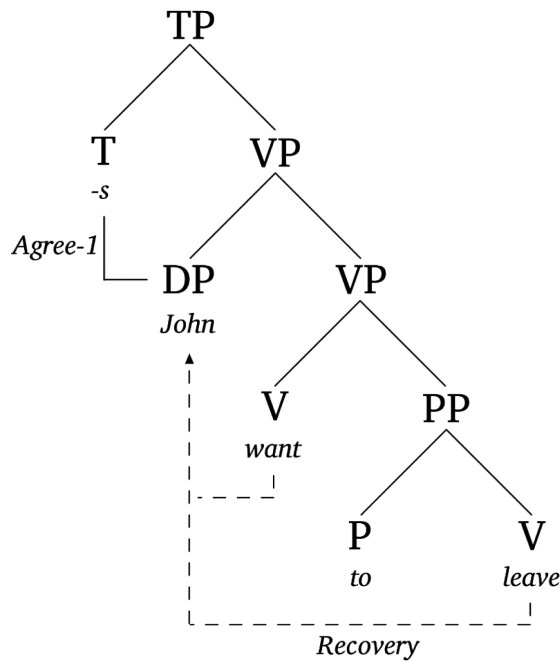
If an argument occurs between the agent of the main verb and the infinitival, then, all else being equal, it will be selected as an antecedent, deriving *John wants Mary to leave$_{\varphi\_=Mary}$*.

No null argument (PRO) is projected; everything is reconstructed form the 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 residing inside functional heads (here D_, φ_). 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 in the present work can be thought of an inverse of Borer's linking principle. Another precursor that has influenced the present approach is that of Janke (2008), who eliminated PRO by unifying control with anaphora resolution and proposed that the thematic roles of predicates are saturated by an upward looking percolation mechanism resembling the recovery algorithm proposed in the present study.[5] Both approaches were, in short, particularly useful in solving the inverse problem.

The analysis presupposes that there is a distinction between antecedent recovery for D_ and φ_: the latter will create standard control, the former the more liberal antecedent recovery signature observed in connection with Finnish partial pro-drop. The fact that Finnish third person null subject antecedent differs from obligatory and non-obligatory control 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 they are 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 after recovery, a generic interpretation, referring to people in general, is created as a last resort strategy.

An obligatory control structure (OC) arises under the present 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 or generate

---

5   I ultimately ended up rejecting the idea of unifying antecedent recovery with anaphora resolution. That hypothesis constitutes an interesting idea worth exploring in future work, however.

agreement suffices but still requires an antecedent. This phenomenon is illustrated by the examples in (17a–b).

(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 must 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 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 (Wilkinson 1971; Landau 2000).[6]

We still have to formalize these assumptions, again to anticipate computational work. The following formalization was assumed. Let us assume that a verb such as *begin* has a lexical feature SEM:INTERNAL, and a verb with the opposite profile (e.g. *persuade*) has 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 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 complement 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 will be 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. These assumptions are illustrated in (18a–c).

(18) a.   John   tries   –   to   leave.
                  SEM:INTERNAL   –ARG   V

      b.   John   persuades   Mary$_1$ to   —$_1$ leave.
                  SEM:EXTERNAL   +ARG   V

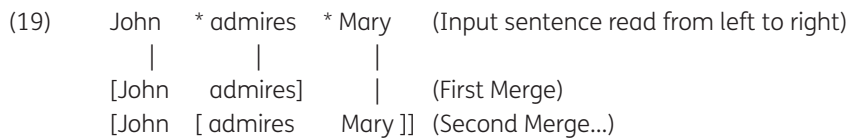   c.   John   wants   (Mary)   to   leave.

It follows that the unvalued phi-features at the embedded infinitival *leave* in the example (18a) must be valued by the main clause subject by recovery. This derives properties of obligatory control.

To summarize, the SEM-feature creates three classes of control predicates: SEM:INTERNAL (*John tried (*Mary) to leave*)(18a), SEM:EXTERNAL (*John persuaded *(Mary) to leave*)(18b) and neither (*John wanted (Mary) to leave*)(18c). Because SEM:EXTERNAL requires that the argument it projects and the argument below are not connected conceptually, the presence of this feature must also limit upward antecedent search at the syntax-semantic interface. This restriction was added to recovery: the operation is blocked by the presence of SEM:EXTERNAL at a head. Symbol v* is used in this study for a transitivizer with this feature (e.g. *order* = v* vs. *want* = v).

---

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.

3.4 MAPPING SENSORY INPUTS INTO PHRASE STRUCTURE OBJECTS: THE LEFT-TO-RIGHT ARCHITECTURE

An analysis was proposed in the previous sections that could, at least in principle, work within the narrow parameters defined by the inverse framework. Everything that the analysis requires can be put together from the material available in the sensory input. To show, however, that the proposed system *does* solve the inverse problem we must demonstrate 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 met. First, we need a function that maps the incoming phonological words into phrase structure objects that contain lexical items with features and structural configurations presupposed by lexical decomposition (*Figure 1*), Agree-1 and recovery. To this end, a left-to-right architecture of Phillips (1996; 2003) was adopted as a starting point. In this system Merge (syntactic structure building) operates 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…)
```

A system of this type, when supplied with the surface vocabulary mechanism sketched in *Figure 1*, generates phrase structure objects incrementally from the sensory input. This architecture was enriched with a Python based linear phase parsing recursion 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 partial phrase structure, explores that search space in a well-defined order while consuming words from the input until they have all been processed. The parser backtracks if the output is not well-formed, resulting in a reanalysis 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 crucial point is that the system will explore all phrase structure interpretations compatible with any linearly organized input sentence and thus maps any input sentence into a set of bare phrase structure objects that we can then use to test the proposed mechanisms of Agree-1 and recovery.

It is important to point out, however, that the proposed analysis of null arguments does not require one to use any specific parsing architecture. The analysis could be added to anything that maps sensory inputs into phrase structure objects or other kinds of syntactic objects that are sufficiently rich to sustain Agree-1 and recovery.

4 FORMALIZATION AND TESTING

4.1 INTRODUCTION

The analysis elucidated in the previous sections was formalized in order to verify that it can solve the inverse problem. The analysis was added to an existing linear phase parsing toolkit written in Python, which I used to automatize all calculations.[7] Formalization of a linguistic theory by means of a machine-readable language such as Python is not common, but it does not differ in any principled way from regular linguistic formalization; instead, one could say that it provides several advantages over purely symbolic formalization. Running the required calculations becomes a matter of starting one script. Thus, it is possible to test the logical consequences of a linguistic analysis automatically over a potentially huge number of test sentences. The second advantage is that a machine-readable formalization requires the researcher to be 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, in this study with the average speed of 70 ms per sentence. Fourth, the whole formalization can be deposited into public domain where it can be shared, examined, criticized, and developed. Consequently, the formalization proposed in this

---

[7]   The existing algorithm did not process agreement and phi-features and had no mechanism for antecedent recovery, so it failed to understand all inputs that contained null arguments.

study together with all the required files are available online.[8] Finally, the use of computational derivations allows the researcher to analyze the internal operation of the model at any level of detail desired. In the present study, the algorithm wrote a detailed log file containing all linguistically relevant computational steps it executed in connection with processing each input sentence. This will allow the researcher to evaluate the model also against performance data obtained from psycholinguistic experimentation. I will include several screenshots of these log files below. Psycholinguistic concerns did not play any role in the present study, however.

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 earlier in this article. In order to avoid bias in the selection of the test materials, the relevant properties (pro-drop, agreement, word order, embedding and control) were crossed to result in $2 \times 2 \times 2 \times 2 \times 2 = 32$ experimental categories. These categories are summarized in *Table 2* and discussed in more detail in the supplementary document.

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

SENTENCE	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
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
2152–2178	Yes	Grammatical	Canonical	Yes	Yes	Pro-drop with embedding
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 containing the analytical solutions together with null arguments and their antecedents, and another the detailed derivational logs. Correctness of the output was verified by the author by examining these files. Detailed comments on the verification procedure and the results are available in the supplementary document. Each sentence processed by the model was numbered by the algorithm during the execution of the script. Thus, it is possible to find a step-by-step derivation for any sentence discussed in the main text or included into the test corpus by searching for its identifier from any of the output files. These numbers and sometimes even the line numbers in the log files are also referred to in the main text, so that a reader can find the corresponding entries in the case further details are required. 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 file and can thus be found more conveniently. These are sentences with identifiers #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 and *Figure 1*. The overall methodological framework is illustrated in *Figure 2*.
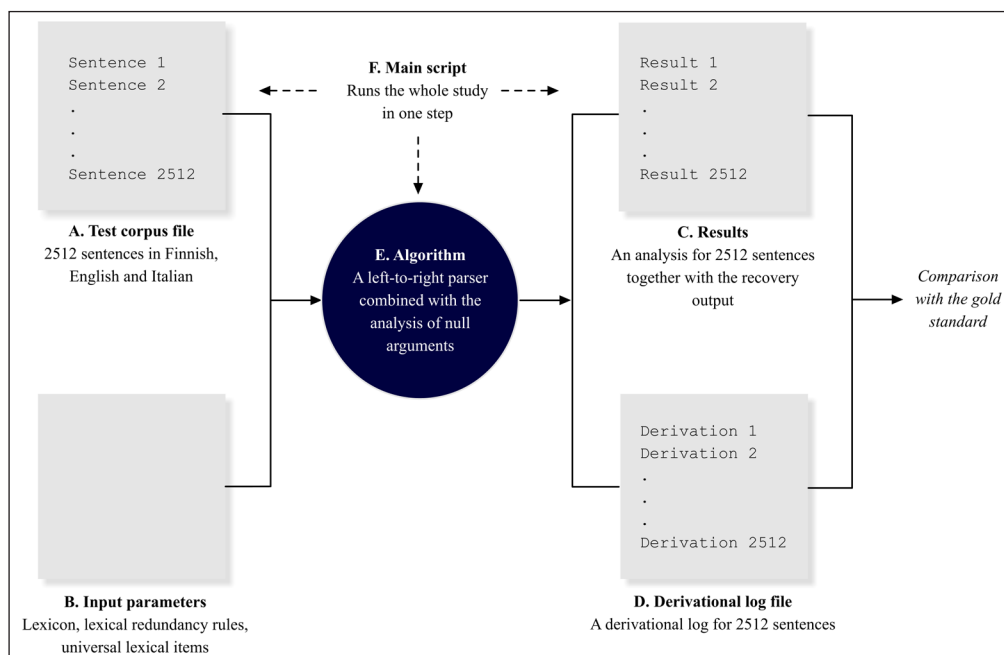

Figure 2 The general framework: an analysis (E), input parameters (A, B) and the raw output data (C, D). The output is verified by comparing it with a gold standard provided by a native speaker, in this case the author.

4.2 TESTING THE MODEL WITHOUT NULL SUBJECTS

The computational mechanisms were first tested without the presence of null arguments to make sure that the basic grammatical and parsing-related mechanisms were working correctly. Example (20) shows how the algorithm 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 reported analyses 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)
　　　　　　$[_{TP}$ noi$_{1,\varphi}$ $[_{TP}$ T$_{\varphi}$ $[_{vP}$ ——1 $[_{vP}$ v$_{\varphi\_=noi}$ $[_{VP}$ 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 indeed produces fairly standard bare phrase structure interpretations for the input sentences. After creating the subject, Steps 1–7, lines 45300–45326 in the derivational log file, the input word *adoriamo* 'admire-v-T-3pl' is decomposed into a complex

head V-v-T by the lexical-morphological component, with T containing features {φ_, φ:1pl} (Step 8, line 45323, see also **Figure 3** below). φ_ is provided in the lexicon, {φ:1pl} by the suffix *-iamo* extracted from the input and arriving through the morphosyntactic route, as assumed in the analysis, **Figure 1**.[9] These grammatical heads are then fed individually to the syntactic component (Steps 9, 10, 11 and **Figure 1**), where they are repackaged into a complex head TvV (line 45363). **Figure 3** provides a screenshot of the derivational log file containing these steps.
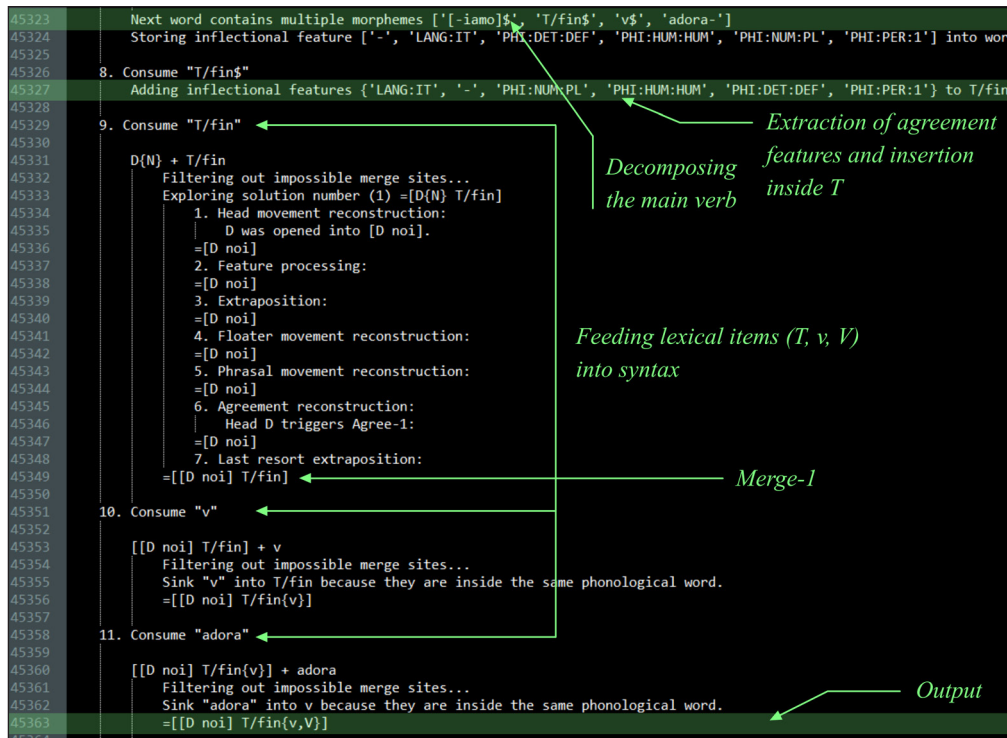
The direct object *Luisa* is treated in the same way (Steps 12–18). The result is (21), with "DN" and "φTvV" denoting complex polymorphemic words (D = determiner, N = noun head, T = tense, v = transitive verbal head, V = verb root).

(21)   We   admire   Luisa. (#101, Step 18, line 45423)
         [DN   [φTvV   DN]]

Complex heads are reverse-engineered by head reconstruction (lines 45426–45430).[10] DP arguments are spread into [D N] (=DP) structures, while the TvV complex generates a head chain. The result is (22).

(22) $[_{TP}$ We $[_{TP}$ T $[_{vP}$ v $[_{VP}$ admire(V) Luisa]]]] (line 45430)

The algorithm reconstructs the preverbal subject *noi* 'we' into the canonical thematic position at SpecvP, as indicated by the reconstructed position $\_\_\_{1}$ (lines 45438–9). Agreement reconstruction (valuation by Agree-1 and/or by means of input suffixes) will then be applied to the resulting structure (lines 45441–45447). Heads with unvalued phi-features and +VAL trigger morphosyntactic valuation. The result is (23)(line 45458). Details of the computational steps involved in these operations are provided in **Figure 4**.

(23) $[(Noi_{1})_{\varphi} [T_{\varphi\_=1pl} [\_\_\_{1} [v_{\varphi\_} [adora_{\varphi\_} Luisa]]]]]$

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 45453–4),


```
45423        >>> Trying candidate spell out structure [[D noi] [T/fin{v,V} D{N}]]
45424            Checking surface conditions...
45425            Reconstructing...                        Output of the parser
45426                1. Head movement reconstruction:
45427     Head         Target v{V} in T/fin
45428                  =[[D noi] [T/fin [v{V} D{N}]]]
45429   reconstruction Target adora in v
45430                  =[[D noi] [T/fin [v [adora D{N}]]]]
45431                =[[D noi] [T/fin [v [adora [D Luisa]]]]]
45432                2. Feature processing:
45433                =[[D noi] [T/fin [v [adora [D Luisa]]]]]
45434                3. Extraposition:
45435                =[[D noi] [T/fin [v [adora [D Luisa]]]]]
45436                4. Floater movement reconstruction:
45437                =[[D noi] [T/fin [v [adora [D Luisa]]]]]
45438                5. Phrasal movement reconstruction:
45439                   [D noi] will undergo A-reconstruction.
45440                =[[D noi]:31 [T/fin [__:31 [v [adora [D Luisa]]]]]]
45441                6. Agreement reconstruction:
45442                   Head T/fin triggers Agree-1:
45443   Agree-1          T/fin acquired PHI:GEN:N by phi-Agree from __:31.
45444                    T/fin acquired PHI:NUM:PL by phi-Agree from __:31.
45445                    T/fin acquired PHI:PER:1 by phi-Agree from __:31.
45446                    T/fin acquired PHI:DET:DEF from the edge of T/fin.
45447                   Head D triggers Agree-1:
45448                =[[D noi]:31 [T/fin [__:31 [v [adora [D Luisa]]]]]]
45449                7. Last resort extraposition:
45450              = [[D noi] [T/fin [[D noi] [v [adora [D Luisa]]]]]]
45451            Checking LF-interface conditions.
45452            Transferring [[D noi]:31 [T/fin [__:31 [v [adora [D Luisa]]]]]] into
45453                v with ['PHI:DET:_', 'PHI:NUM:_', 'PHI:PER:_'] was associated at
45454                   1. [D noi]    (alternatives: 2. T/fin )
45455   Recovery     adora with ['PHI:DET:_', 'PHI:NUM:_', 'PHI:PER:_'] was associated
45456                   1. [D Luisa]    (alternatives: 2. [D noi] 3. T/fin )
45457            Transfer to C-I successful.
45458            Semantic interpretation/predicates and arguments: [' ', 'Agent of
45459   -------------------------------------------------------------------------
45460                                              All tests passed
45461   -------------------------------------------------------------------------
45462        Solution:
45463        [[D noi] [T/fin [[D noi] [v [adora [D Luisa]]]]]]
45464        Grammar: [[D noi]:1 [T/fin [__:1 [v [adora [D Luisa]]]]]]
45465        Spellout TT/finP = [DP:1 [TT/fin [__:1 [v [V [D N]]]]]]
```

Figure 4 Screenshot of the derivational log file showing the application of head reconstruction, Agree-1, and recovery operations.

whereas the closest antecedent for the verb is *Luisa* (lines 45455–6). This provides (24)(line 45463). The antecedent relations can be found also from the results file (entry #101, lines 1019–1025), where they are expressed in a less technical way.

(24) $[_{TP}$ (Noi$_1$)$_{1pl}$ $[_{TP}$ T$_{\varphi=1pl}$ $[_{vP}$ —1 $[_{vP}$ v$_{\varphi\_=noi}$ $[_{VP}$ adora$_{\varphi\_=Luisa}$ Luisa]]]]]
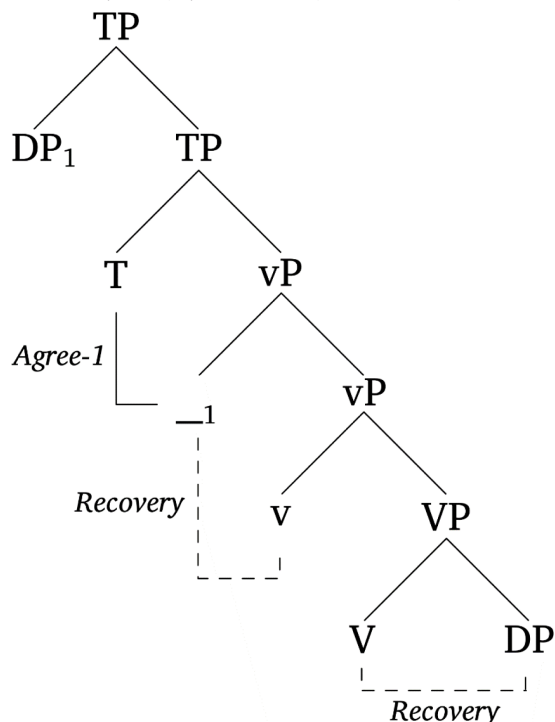
Recovery begins from the element triggering the operation (e.g., the main verb) and by examining if its sister can provide an antecedent; if not, the operation is applied iteratively to the mother node. The mechanism tries to establish an "upward path" from the triggering element into an antecedent. The agent argument for v is reconstructed by means of control: $\varphi\_$ 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 (redundant?) control relation between the verb and its complement in (24). If this solution were ignored, the verb would target a nonlocal antecedent and create a reflexive meaning 'we admire ourselves'.[11]

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 proposed agreement mechanisms did not break the mechanisms processing (*wh-*) operator movement. This was checked by feeding the algorithm with Finnish sentences (25–26), which it handled correctly.

(25)　　*Finnish*
　　　　Kuka　　ihaile-e　　Merja-a?　(#36)
　　　　who.NOM　admire-3SG　Merja-PAR
　　　　'who admires Merja?'

(26)　　*Finnish*
　　　　Ketä　　hän　　ihaile-e　　　___? (#37)
　　　　who.PAR　he.NOM　admire-3SG
　　　　'Who does he admire?'

These matters are discussed further in the supplementary document. Overall, however, these tests provided that the basic grammatical mechanisms were operating correctly. The algorithm is able to process input sentences in a linguistically meaningful way, and both Agree-1 and recovery were working as intended. These tests verified that the left-to-right parser component was doing what it was supposed to do, building plausible syntactic structures for the inputs it received.

4.3 PRO-DROP

The model was next tested with subjectless sentences in each of the three languages and for all person and number combinations (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 again be interpreted as representing logical consequences of the analysis, not independent and freely modifiable solutions generated by the author.

(27)　　a.　　*Finnish*
　　　　　　Ihaile-n　　Merja-a. (#2100)
　　　　　　admire-1SG　Merja-PAR
　　　　　　$[_{TP}\ T_{1sg}\ [_{vP}\ v_{\varphi\_=1sg}\ [_{VP}\ \text{admire}_{\varphi\_=Merja}\ \text{Merja}]]]$ (Simplified output)
　　　　　　'I admire Merja.'

　　　　b.　　*Finnish*
　　　　　　*Ihaile-e　　Merja-a. (#2102)
　　　　　　admire-3SG　Merja-PAR
　　　　　　(No parsing solution found.)

　　　　c.　　*Italian*
　　　　　　Ador-a　　Luisa. (#2112)
　　　　　　admire-3SG　Luisa
　　　　　　$[_{TP}\ T_{3sg}\ [_{vP}\ v_{\varphi\_=3sg}\ [_{VP}\ \text{adora}_{\varphi\_=Luisa}\ \text{Luisa}]]]$ (Simplified output)
　　　　　　'He admires Luisa.'

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

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 (27a) and Italian (27c) sentences generated by the algorithm is the correct one. It has been argued, convincingly, that most Finnish finite verbs have an EPP feature that requires their specifier positions to be filled in by a syntactic phrase (Vainikka 1989; Vilkuna 1989; Holmberg and Nikanne 2002; Huhmarniemi 2019). There is no such phrase at the preverbal subject position in the pro-drop sentences above. To understand how the analysis handles 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 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 the two 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 is different: 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).[12]

The Finnish third person pro-drop construction (27b) is correctly deduced as ungrammatical. The third person suffix present in the input values the number and person features at T (lines 1446511–4), but D_ remains unvalued and triggers recovery (line 1446521). No antecedent is found, and the input is judged ungrammatical (line 1446527). These steps are illustrated in **Figure 5**.

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 admire-3sg/D_ Merja-PAR
     'Pekka says that he (=Pekka) admires Merja.'

b.   *Minä sanon että ihaile-e   Merja-a. (#44)
     I   say-1sg that admire-3sg/D_ Merja-PAR
     Intended: 'I say that (he) admires Merja.'

Example (28b) is ruled out because the phi-features of the antecedent, first person singular, do not match with the third person phi-features of the unvalued phi-set. Notice that in a sentence in which there are two potential c-commanding antecedents, the model can only select the local one, as shown in (29).

---

12   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) argue that in Finnish the finite T checks a discourse-based topic feature NON-FOCUS. The linear phase algorithm uses a general specifier selection feature SPEC:F to derive these properties, forcing a phrase with some feature F (e.g., D, non-focus) to occur at the specifier position (Brattico 2019: 23–26, 63–65). This alone is now insufficient, however, because we have assumed that the agreement features at the head should be able to satisfy the same condition. Therefore, the notion of "specifier of head H" was generalized in this study 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, including, by definition, all specifiers and head-internal pro-elements.

```
1446494    >>> Trying candidate spell out structure [T/fin{v,V} D{N}]
1446495        Checking surface conditions...
1446496        Reconstructing...
1446497            1. Head movement reconstruction:
1446498                Target v{V} in T/fin
1446499                =[T/fin [v{V} D{N}]]
1446500                Target ihaile- in v
1446501                =[T/fin [v [ihaile- D{N}]]]
1446502            =[T/fin [v [ihaile- [D Merja-]]]]
1446503            2. Feature processing:
1446504            =[T/fin [v [ihaile- [D Merja-]]]]
1446505            3. Extraposition:
1446506            =[T/fin [v [ihaile- [D Merja-]]]]
1446507            4. Floater movement reconstruction:
1446508            =[T/fin [v [ihaile- [D Merja-]]]]
1446509            5. Phrasal movement reconstruction:
1446510            =[T/fin [v [ihaile- [D Merja-]]]]
1446511            6. Agreement reconstruction:
1446512                Head T/fin triggers Agree-1:
1446513                    T/fin acquired PHI:NUM:SG from the edge of T/fin.
1446514                    T/fin acquired PHI:PER:3 from the edge of T/fin.
1446515                Head D triggers Agree-1:
1446516            =[T/fin [v [ihaile- [D Merja-]]]]
1446517            7. Last resort extraposition:
1446518          = [T/fin [v [ihaile- [D Merja-]]]]
1446519        Checking LF-interface conditions.
1446520        Transferring [T/fin [v [ihaile- [D Merja-]]]] into the conceptual-intentional system...
1446521            T/fin with ['PHI:DET:_'] was associated at LF with:
1446522                No antecedent found, LF-object crashes.
1446523            v with ['PHI:DET:_', 'PHI:NUM:_', 'PHI:PER:_'] was associated at LF with:
1446524            ihaile- with ['PHI:DET:_', 'PHI:NUM:_', 'PHI:PER:_'] was associated at LF with:
1446525                1. [D Merja-]    (alternatives: )
1446526            Semantic interpretation failed, transfer to C-I crashed.
1446527        The sentence cannot be interpreted at LF
```

Figure 5 A screenshot from the derivational log file containing the failed recovery operation for the Finnish third person pro-drop construction. Only the person and number features are valued by Agree-1 (lines 1446511-514), hence D_ remains unvalued and crashes the derivation without an antecedent (lines 1446521-2). The term "LF-interface" occurring in the derivational log refers to the syntax-semantics interface, a level of representation that constitutes the input to the semantic system.

(29)   *Finnish*
       Pekka sanoo että Jukka sanoo että ihaile-e   Merja-a. (#45)
       Pekka says   that Jukka says   that admire-3SG$_{D\_=Jukka}$ Merja-PAR
       'Pekka says that Jukka says that he (=Jukka/*Pekka) admires Merja.'

To me the non-local argument is marginally possible when the semantic interpretation leans towards such interpretation.[13] Thus, nonlocal antecedents are registered and shown in the logs (see *Figure 4*, line 45456), but they are currently never selected. I do not know how to formulate and formalize selection conditions for nonlocal antecedents, as they seem to involve some type of extralinguistic plausibility considerations.

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 comfortably.'

      b.   *Istu-u tässä mukavasti. (Holmberg 2010, ex. 1a) (#47)
           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 finds *Pekka* from the main clause. This derives the correct properties. 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

---

13   ?*Murhaaja₁ uskoi yhä että poliisi₂ luulee että pro₁,₂ 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.

properties correctly. Virtually any phrase can (as long as it can constitute the topic) satisfy the EPP condition of Finnish (Holmberg & Nikanne 2002), hence the locative PP will do as well. This PP cannot, however, value D_ at T by Agree-1 because it is not a DP. The D-feature 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)
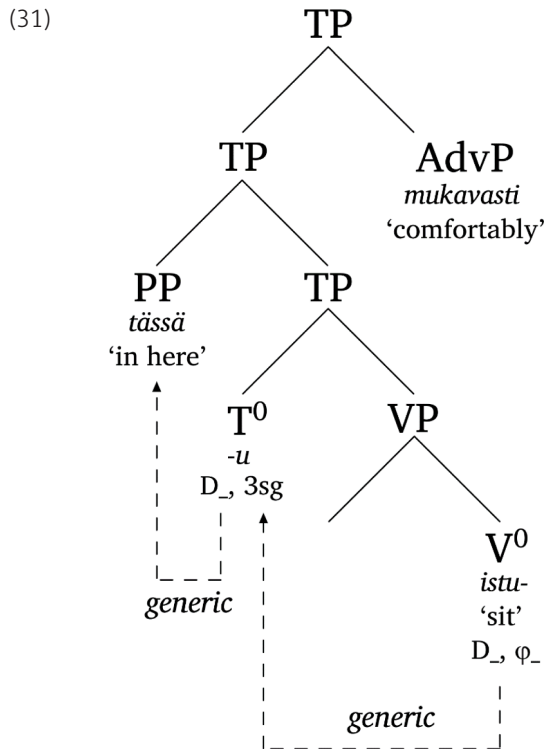
This deduction is not uncontroversial. No generic null pronoun fills in any argument position; the generic interpretation is created at the syntax-semantic interface as a last resort. Holmberg (2010) considers a number of arguments suggesting that a phrasal generic pronoun must be present in these constructions. His 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 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 automatically 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 complement selection feature of the main verb *want* is not satisfied.

(32)   *Finnish*
         *Minä halua-n.   (#50)
         I.NOM   want-1SG
         Intended: 'I want (myself/one/people in general).'

There is no derivational path for generating a pronominal complement from the phi-set of the verb, 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 overt agreement morphology. The present analysis licenses a null subject argument when a consistent pronominal element constructed from the 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 carries non-conflicting phi-feature(s) as a lexical property, then a null argument with those features will be generated. A crucial assumption we must make here is that the complete loss of agreement from a language at the surface must eliminate all conflicting phi-features from the lexicon (contrary to what is

the case in Swedish, English and French).[14] 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 with their arguments via a recovery mechanism that accesses discourse, perhaps in a way comparable to the behavior of the 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 in which lexical elements, therefore, have no conflicting phi-features. I tested both hypotheses, one in which the hypothetical finite verbs had only valued person features, corresponding to the first hypothesis, and another in which they had an unvalued person feature, corresponding to the second (the person feature was selected for testing purposes and is not supposed to reflect the situation in any real radical pro-drop language). 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).

To summarize, the four possible routes for licensing null arguments predicted to exist on the basis of the present analysis are summarized in *Table 3*.

Lexically valued phi-features	Overt agreement	
	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 mechanism (Swedish, English, French).
Yes	Null argument by lexically valued phi-feature(s). Interpretation depends on the nature of the phi-features (Chinese, Korean, Japanese?).	Null argument licensed by both overt agreement and lexical phi-features, which must match with each other (per Agree-1). This class corresponds to hypothetical, lexically specified frozen agreement forms that do not trigger recovery.

Table 3 Four routes for generating pro-drop sentences.

Notice that no radical pro-drop languages were included into the test corpus, and hence these remarks should be regarded as tentative.[15]

4.4 CONTROL

Let us next consider how the algorithm handles control. Consider the following examples of standard control first (33a–b). 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. The small verb v finds an antecedent at SpecvP, where the subject is after reconstruction. The lower verb *leave* is linked with *John* in (33a) and *Mary* in (33b). The algorithm, therefore, deduces the correct antecedent properties. Moving to more complex examples, (34a–k) represent the core cases of Finnish control and cover the 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 matches with the native speaker judgment. The results are explained below.

14　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 phi-features can be eliminated throughout the lexicon.

15　Construction of a systematic test corpus for a computational study such as the present one requires input from a native speaker with linguistic expertise.

(33)   *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 and disambiguates the A-infinitival into −ARG (34a–b). This prevents the underlying VP from projecting an external argument, as shown in (35).

(35)
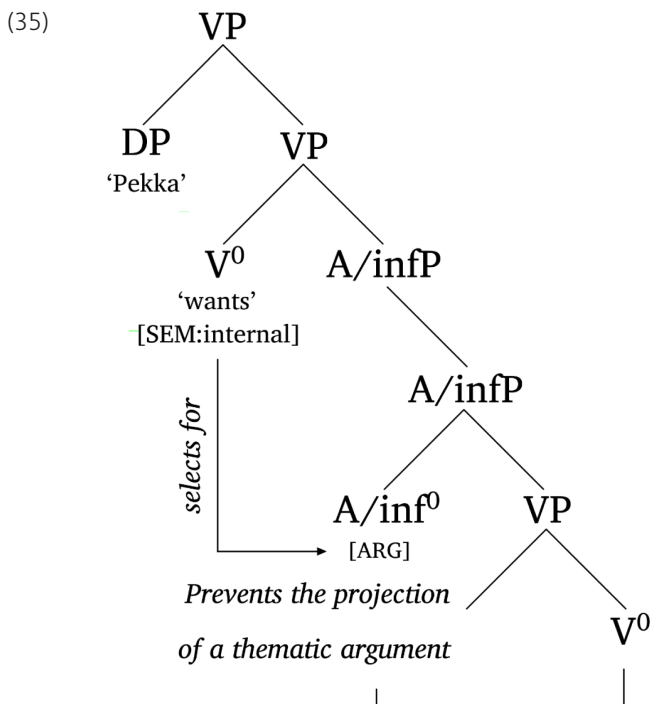
Sentence (34b) is judged ungrammatical, as the argument *Merja-n* 'Merja-gen' reconstructs into SpecVP but will be left without a thematic role due to the selecting −AGR head. The derivation fails at the syntax-semantics interface without interpretation. The verb 'order' (34c–d) is marked for SEM:EXTERNAL and disambiguates the selected infinitival into +ARG. The embedded verb projects a thematic role; hence a DP argument may appear 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 the antecedent. If there is no agreement, the VA-infinitival head cannot check its EPP feature. Notice that since the VA-infinitival does exhibit overt agreement (34j), it is marked for +VAL (it is worth recalling here that in Finnish many infinitivals exhibit overt phi-agreement, referred to as "possessive agreement" in the literature).

Example (36) shows noncanonical control clauses that the model also deduces correctly. The model calculates these data by applying phrasal reconstruction before Agree-1 and recovery (see **Figures 4** and **5**, #8, #2179–2194). These data show that the algorithm solves the problem of noncanonical subjects 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 tomorrow.'

      c.   Huomenna   käske-e   Pekka   Merja-n   lähte-ä.
           tomorrow   order-3SG   Pekka.NOM   Merja-GEN   leave-A/INF
           'Pekka orders Merja to leave tomorrow.'

      d.   Huomenna käske-e   Merja-n   Pekka   lähte-ä.
           tomorrow   order-3SG   Merja-GEN   Pekka.NOM   leave-A/INF
           'Pekka orders Merja to leave tomorrow.'

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 SEM:EXTERNAL (37a–ii). Subject control, illustrated by (37b–i), is often regarded as surprising because the 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 replace the Minimal Distance Principle 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 original linear phase algorithm, 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. These analyses are shown in (38).

(38)   a.   John persuades [Mary to leave$_{\varphi\_=Mary}$.] (=(34a–i), #67)
        b.   John promises [Mary] ⟨to leave$_{\varphi\_=John}$.⟩ (=(34b–i), #69)

This does provide the correct antecedent properties, though the solutions are not unproblematic and deserve a comment. Solution (38b) is reminiscent of Larson (1991), who proposed that at the level at which recovery applies the infinitival occurs at a higher position such that the object

antecedent becomes invisible. The algorithm, therefore, adopts a Larsonian solution when faced with an input sentence of this type. Analysis (38a) is, however, problematic. The linear phase parser is only allowed to create binary branching trees (Brattico 2019: Chapter 2.3) and cannot, therefore, create a structure in which the verb *persuade* takes two complements (e.g., *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 target 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 thematic dependency between the main verb and that subject. This resembles the situation with the Finnish examples in (35c), where we attest a similar mismatch between syntactic structure and semantic intuition.

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

(39)   *Finnish*
         Pekka   etsi-i   Merja-a   juost-en. (#73)
         Pekka   search-3SG   Merja-PAR   run-ADV
         'Pekka searches Merja by running.'

The algorithm finds 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)   *Finnish*
         Pekka   e-i   [$_{vP}$ nähnyt [Merja-a]   ⟨kävele-mässä$_{\varphi\_=Merja.}$⟩]
         Pekka   not-3SG   see   Merja-PAR   walk-MA/INF
         'Pekka did not see Merja walking.'

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

(41)   To leave would be a mistake. (#74)
         [to leave]$_1$ T$_{\varphi\_}$ ___$_1$ be [a mistake]
         'For people in general to leave (now) would be a mistake.'

T fails to value its 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 known to evoke similar 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 thematic subject of 'ask' (determined by v*) will be Pekka. 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   to.sing
               'Pekka asks (people, one) to sing.'

b. Pekka   haluaa laulamaan.
Pekka   wants   to.sing
'Pekka wants to sing.'

All these sentences are judged 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 sentences corresponding to 'John wants PRO'.

4.5 MIXED CASES

The model was tested with sentences containing both null subjects and control (#2117–2143, 2195–2260, 2267–2510). 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
Intended: '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 (=Pekka) wants to leave.'

In the example (44a), the algorithm values the phi-features of T by the agreement suffixes and then uses these features as an antecedent for the infinitival. In (44b) lack of an antecedent for D_ results in ungrammaticality. Finally, sentence (44c) is correctly judged as grammatical and interpreted so that Pekka constitutes the agent of saying, wanting and leaving.

4.6 FINITE CONTROL, TENSE, AND AGREEMENT

Finnish third person null subject can be said to exhibit 'nonlocal control'. The null subject is typically bound by a c-commanding argument from the next 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 seems to be that the antecedent in Finnish need not be local, perhaps it can circumvent even the c-command requirement. Furthermore, Landau argues that the prototypical finite control signature emerges if and only if the controlled finite clause is deficient either in terms of tense or agreement (or both), but this is not true of Finnish, in which the embedded finite clause is, at least on the surface, fully specified for both agreement and tense.

The fact that deficient agreement leads into control follows directly from the present analysis: an agreementless and argumentless finite verb will trigger recovery at the syntax-semantics interface. The finite clause boundary does not limit recovery. If T is marked lexically for −VAL, we further derive a 'finite OC signature'. The situation with deficient tense is less straightforward, however. If sentences that come with deficient tense specification indeed do possess full agreement, these data constitute a fundamental difficulty because full agreement should, all else being equal, block all control, leaving nothing unvalued at the syntax-semantics interface. In addition, there is nothing in the present analysis that links tense to control. What makes the matter nontrivial, however, 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 deficient. Perhaps Landau's deficient tense creates similar deficient agreement configurations? The matter remains to be shown, however, and the problem was left for future research.

4.7 SUMMARY

To summarize, the algorithm classifies control constructions in the manner depicted in *Figure 6*. If an antecedent cannot be found, the interpretation comes out as generic.
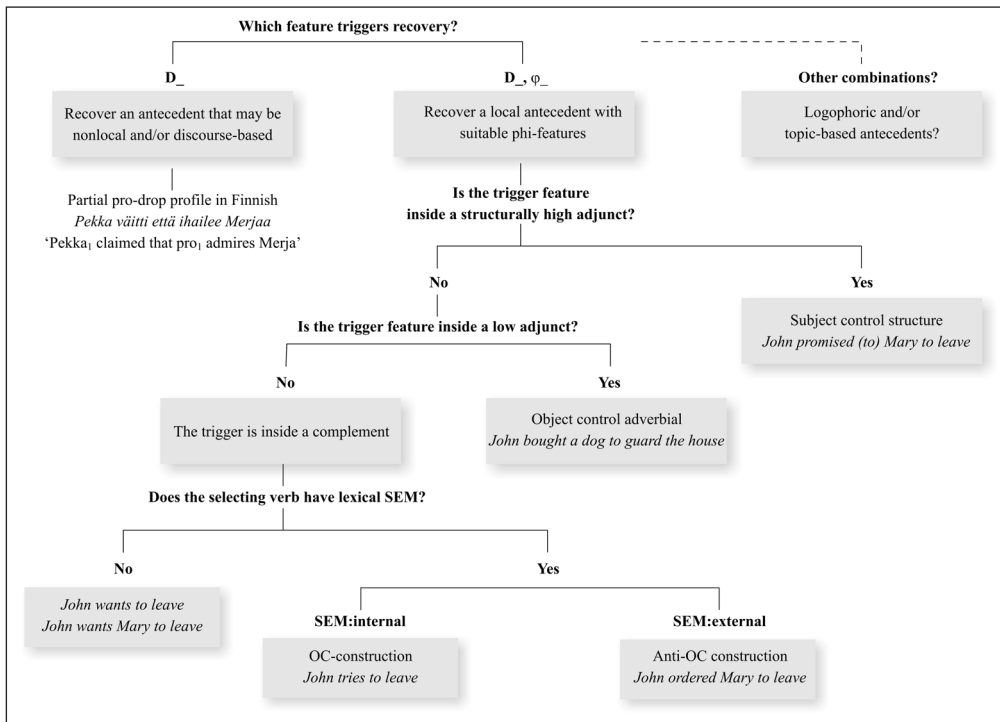
Figure 6 Classification of control constructions by the algorithm.

The classification is not meant as a complete or even correct classification of all control constructions; rather, it should be viewed as a logical consequence of the analysis that was designed as a possible solution to the inverse problem. Furthermore, the empirical literature indicates that the system might contain more branches, corresponding at least to logophoric and/or topic-based antecedents.

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 earlier work by Borer (1986; 1989), Phillips (1996, 2003), Cann et al. (2005), Janke (2008) and Brattico (2019). The algorithm was able to deduce correct phrase structures, null arguments, and their antecedents. Three languages were examined: Italian, English and Finnish, each with different behavior with respect to their null arguments (non-pro-drop, consistent pro-drop and partial pro-drop languages).

A finite null subject pronoun (pro) occurs if and only if 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 must not contain conflicting features. A controlled null argument is generated during the derivation if and only if an unvalued phi-feature cannot be valued from the resources available in the input. Unvalued phi-features trigger a recovery mechanism 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 technical solution to the inverse problem, it raises the larger question of the possible role of parsing in grammatical theorizing. It would be a mistake, in my view, to reconstruct this question as concerning only the division of labor between competence and performance, as no performance properties, such as efficiency, errors, garden paths, irrationality or 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 more traditional enumerative approaches. 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: 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, i.e. the requirement that a lexical item occurs together with a specifier; 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).

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Supplementary file 1.** Algorithm and methodology. DOI: *https://doi.org/10.5334/gjgl.1189.s1*

AUTHOR AFFILIATIONS

Pauli Brattico *orcid.org/0000-0002-4963-5301*
IUSS University School for Advanced Studies, Piazza della Vittoria n. 15, 27100 Pavia, Italy

REFERENCES

Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parametrizing AGR: Word Order, V-Movement and EPP-Checking. *Natural Language & Linguistic Theory* 16. 491–539. DOI: *https://doi.org/10.1023/A:1006090432389*

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

Barbosa, Pilar P. 2009. Two kinds of subject Pro. *Studia Linguistica* 63. 2–58. DOI: *https://doi.org/10.1111/j.1467-9582.2008.01153.x*

Boeckx, Cedric & Norbert Hornstein. 2004. Movement under control. *Linguistic Inquiry* 35. 431–52. DOI: *https://doi.org/10.1162/0024389041402625*

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

Borer, Hagit. 1989. Anaphoric AGR. In Osvaldo Jaeggli & Kenneth J. Safir (eds.), *The Null Subject Parameter*, 69–109. Amsterdam: Kluwer. DOI: *https://doi.org/10.1007/978-94-009-2540-3_3*

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)*. IUSS, Pavia: Technical software documentation.

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* 2. 115–65. Oxford: Oxford University Press.

Cardinaletti, Anna. 2018. On different types of postverbal subjects in Italian. *Italian Journal of Linguistics*, 30: 79–106.

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. DOI: *https://doi.org/10.7551/mitpress/9780262062787.003.0007*

Culicover, Peter & Ray Jackendoff. 2001. Control is not movement. *Linguistic Inquiry* 32. 493–512. DOI: *https://doi.org/10.1162/002438901750372531*

Farkas, Donka. 1988. On obligatory control. *Linguistic Inquiry* 11. 27–58. DOI: *https://doi.org/10.1007/BF00635756*

Holmberg, Anders. 2005. Is there a little pro? Evidence from Finnish. *Linguistic Inquiry* 36. 533–64. DOI: *https://doi.org/10.1162/002438905774464322*

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. DOI: *https://doi.org/10.1111/j.1467-9582.2008.01154.x*

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.

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.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30. 69–96. DOI: *https://doi.org/10.1162/002438999553968*

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

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

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. DOI: *https://doi.org/10.1007/978-94-009-2540-3_6*

Huhmarniemi, Saara. 2019. The movement to SpecFinP in Finnish. *Acta Linguistica Academica* 66. 85–113. DOI: *https://doi.org/10.1556/2062.2019.66.1.4*

Hyams, N. 1989. The null subject parameter in language acquisition. In Osvaldo Jaeggli & Kenneth Safir (eds.), *The null subject parameter*, 215–38. Dordrecht: Kluwer Academic Publishers. DOI: *https://doi.org/10.1007/978-94-009-2540-3_7*

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

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. DOI: *https://doi.org/10.1007/978-94-009-2540-3_1*

Janke, Vikki. 2008. Control without a subject. *Lingua* 118. 82–118. DOI: *https://doi.org/10.1016/j.lingua.2007.05.002*

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. DOI: *https://doi.org/10.1162/002438903322247560*

Landau, Idan. 2004. The scale of finiteness and the calculus of control. *Natural Language & Linguistic Theory* 22. 811–77. DOI: *https://doi.org/10.1007/s11049-004-4265-5*

Landau, Idan. 2013. *Control in Generative Grammar. A research companion*. Cambridge: Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9781139061858*

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. DOI: *https://doi.org/10.1007/978-94-015-7959-9_7*

Manzini, M. Rita & Anna Roussou. 2000. A minimalist theory of A-movement and control. *Lingua* 110. 409–47. DOI: *https://doi.org/10.1016/S0024-3841(00)00006-1*

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.

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

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. DOI: *https://doi.org/10.1162/002438903763255922*

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. DOI: *https://doi.org/10.1515/9783110883718*

Rizzi, Luigi. 1986. Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17. 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. 613–71. DOI: *https://doi.org/10.1023/A:1006225032592*

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.