Supplementary document to the article "Finnish word order: Does comprehension matter?" published in *Nordic Journal of Linguistics* (2020).

Analysis and derivation of Finnish finite clauses by the algorithm

Pauli Brattico

Research Center for Neurocognition, Epistemology and Theoretical Syntax (NETS)

School for Advanced Studies IUSS Pavia

2020

Contents

1 Introduction

This document describes in some detail how the algorithm derives various Finnish finite clause constructions and handles the free word-order property in this language. The description is written for theoretical linguists, not computer scientists. The algorithm is described elsewhere (Brattico 2019a) and will not be elucidated in detail. Derivations presented in this document are taken from the raw log file output of the study, thus the same information is available there but requires a certain level of understanding of the

internal workings of the algorithm that is not a prerequisite here. Relevant citations to the existing literature will be used sparsely and can be found from the main article.

2 The model

The model defines three information processing functions assumed to underlie human language comprehension. Specifically, it

(i) maps sensory PF-interface inputs (PHON) into a linear stream of lexical items (sets of features), which it acquires from the surface vocabulary (lexicon) by lexical retrieval and morphological decomposition;

(ii) maps the input stream into a set of morphology-syntax interface objects, called spellout structures, by ranked application of a Phillips-style Merge Right operation; and then

(iii) maps spellout structures into LF-interface phrase structures (SEM) that are checked for LF-legibility and then send off to the conceptual-intentional system (C-I) for semantic interpretation (Figure 1).

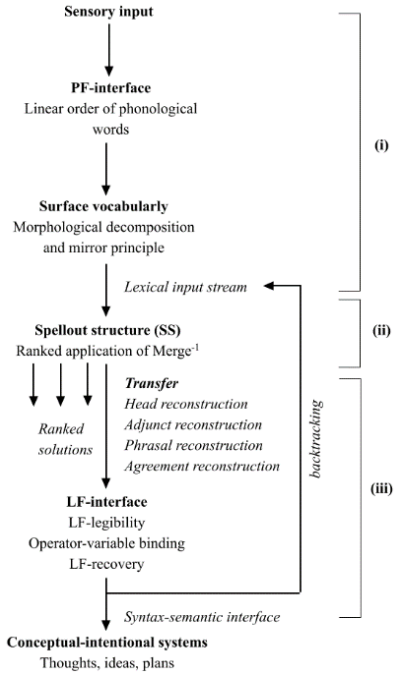


Figure 1. The general information processing flow of the linear phase comprehension algorithm.

Mapping (i) makes use of overt elements, therefore no morphemes or features are hallucinated without some concrete PF-condition being satisfied. The analyses created by the algorithm are therefore relatively austere when compared with what is found from the primary literature. Mapping (iii) is called *transfer*. Adjunct promotion and floating occurs during transfer. Spellout structures (SSs) are surface phrase structures in which the structure mirrors surface linear order, presupposing a depth-first left-to-right linearization. It may therefore be considered as a linguistic representation that contains instructions or an abstract plan for the sensorimotoric system(s). The LF-interface objects are phrase structures corresponding to the LF/d-structure in the standard theory and articulate linguistically relevant (compositional) semantic interpretation. The idea is that while (PHON, SS) can be, and are, subject to language-specific variation, LF-interface objects are normalized and universal, assuming, therefore, that the conceptual-intentional system is universal with little individual or language-specific variation in it. Perhaps (PHON, SS) captures to some type of cognitive fluidity corresponding to sensorimotoric repackaging or chunking of linguistic information. Spellout structure representations are often "compressed" when compared to those occurring at LF.

3 Derivation of canonical constructions (seeds)

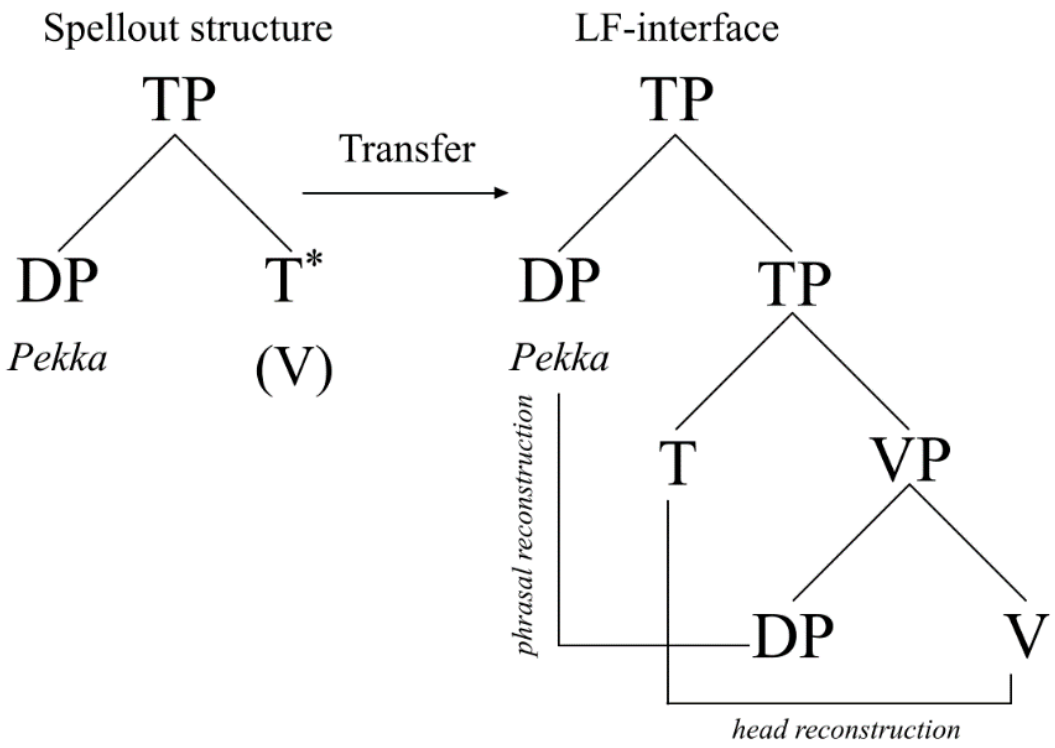
3.1 Introduction

This section shows how the model derives canonical finite clause constructions in Finnish. These examples are taken from Study 1 that derived the baseline seed sentences. The results and derivational logs consulted here can be found from Study 1. All analyses considered here are generated by the algorithm, not by the author, and should be understood as logical consequences of the abstract hypotheses put forward in the main article.

3.2 Declarative finite clauses

A basic intransitive clause is derived as shown in (1).

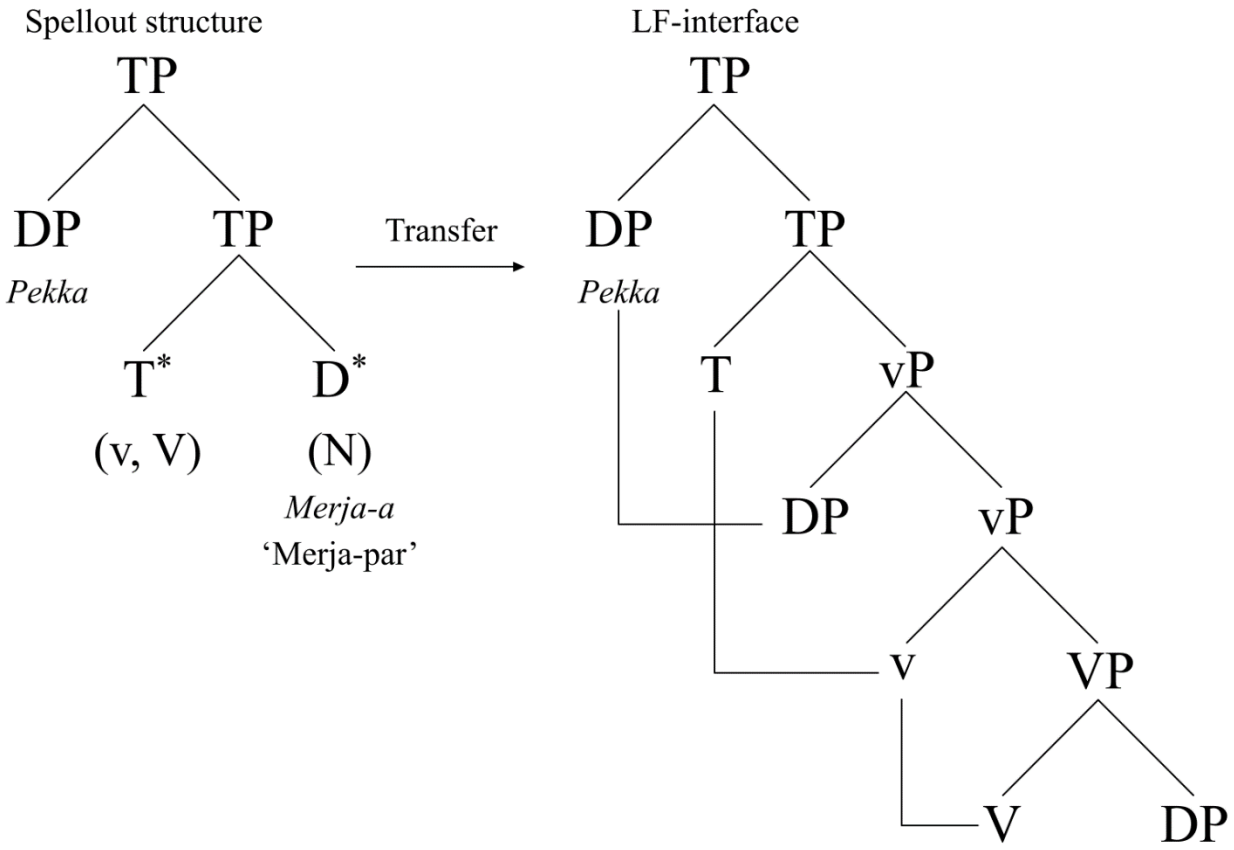
(1)



The algorithm uses bare phrase structure, with labels determined by label algorithm; labels are not part of the representations and are included for readability. All complex heads, such as the finite intransitive verb (T, V), are represented as non-phrasal complex elements at SS and will be denoted by using the asterisk (e.g., T*). The intransitive verb V is contained inside finite T head, which is denoted as T*(V). The verb is extracted from the complex head during transfer: an inverse of head movement. The complex argument *Pekka* is originally represented as D*(N) but is reconstructed during the derivation of the spellout structure, because left branches are assumed to be *phases* (in these sense of Chomsky's Phase Theory) and are transferred independently (Brattico and Chesi 2020). The D*(N) creates a trivial head reconstruction chain [D(N_1) $\__1$] = DP (label D by the labeling algorithm). The model uses the phase approach of (Chomsky 2000) but assumes that all left branches and only them are phases (cyclic domains). The grammatical subject is interpreted as a left DP-adjunct and is reconstructed to SpecVP by an inverse Agree, Agree-1, which positions it below the finite T, as explained in the main article. Both copies are present at LF-interface. The higher copy satisfies the Finnish finite clause EPP condition, while the latter provides a thematic role. Adjunct promotion is performed during transfer when the

algorithm notices that the grammatical subject is not locally c-commanding by a head with the required feature FIN; it is floated to SpecVP. There is no CP layer. There is no material in the PF-input corresponding to C-head(s). Transitive verbs generate the usual T-v-V spread shown in (2).

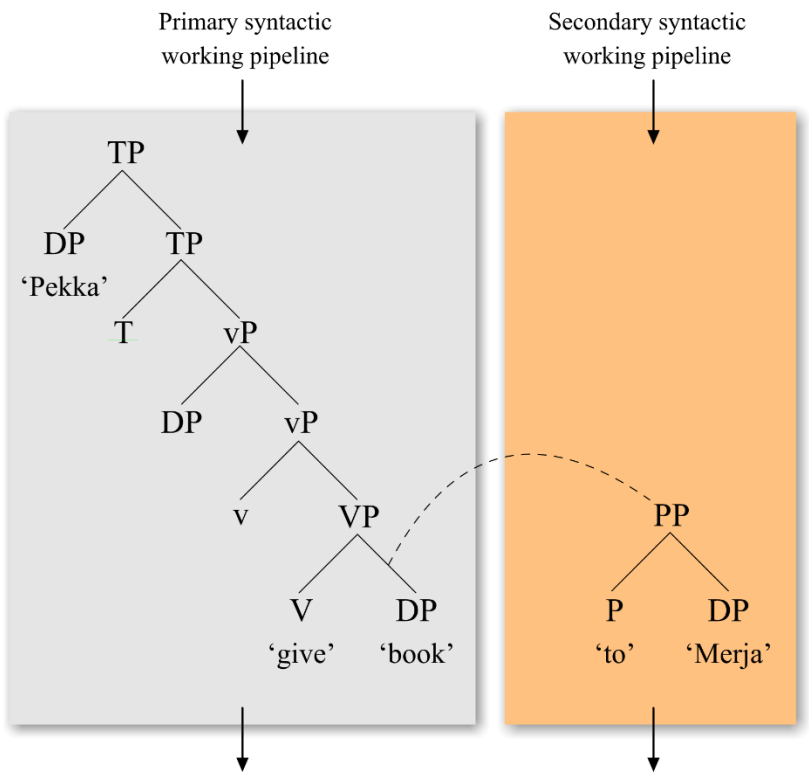
(2)



The direct object appears in the correct canonical position at SS-structure and requires no reconstruction. PAR is matched against a local non-agreeing functional head, here small v. The direct object is not promoted to an adjunct and is not floated. Transitive verbs contain two components v and V, with the latter having some feature requiring an object. Notice how head reconstruction "pushes" argument downwards. The subject is reconstructed to SpecvP.

Indirect arguments are PPs, in agreement with Nikanne (1993), and were promoted into adjuncts and attached under VP. In the example (3), the PP, which occurs after the direct object in the surface string, is interpreted as a right VP-adjunct.

(3)



For the properties of adjuncts, see the main article. Adjunct promotion turned out to be very nontrivial. The PP is promoted into an adjunct by an extraposition rule (lines 347-8 in the derivational log file *Study1_seeds_log_FINAL.txt*). Because selection and labeling ignore adjuncts, V selects for DP ‘the book’. The label of [DP <PP>] is DP, per labeling algorithm that ignores right adjuncts, as elucidated in the main article. The PP is licensed by a feature of the preposition (corresponding to the inherent case suffix per Nikanne’s thesis) that requires that it tails the V. No floating is required. The PP occurs as P*(D, N) at the MS-structure. Extraposition is triggered as a last resort by the fact that [$_{PP}$DP PP] leaves the internal DP without a thematic role.

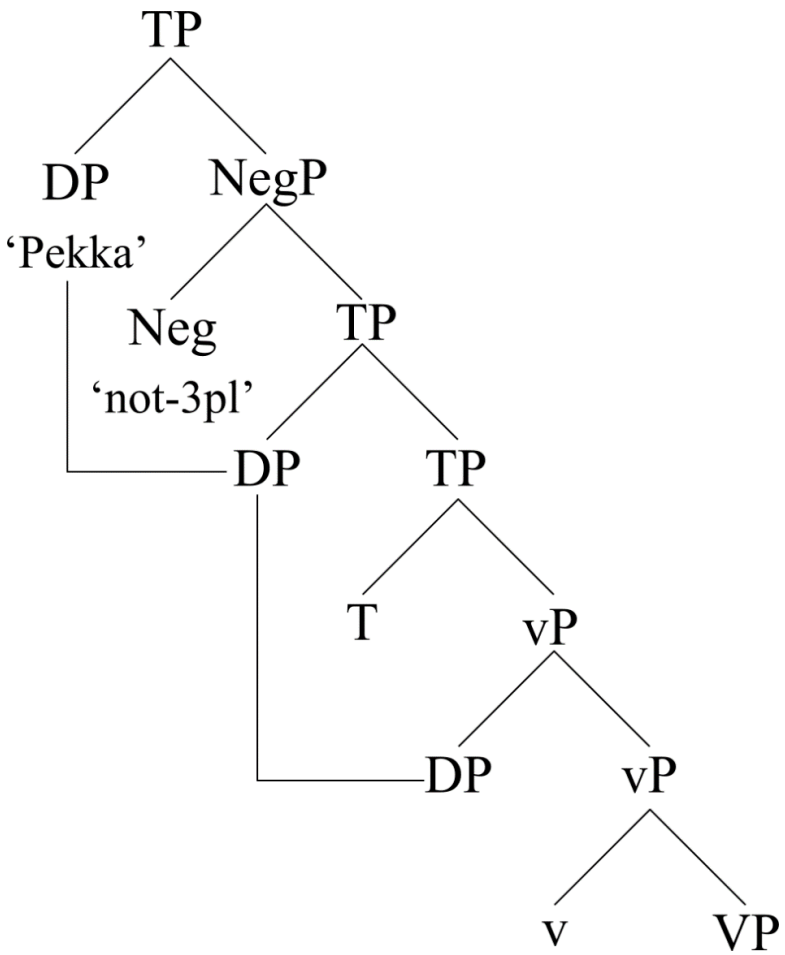
The notion of “secondary syntactic working pipeline” requires a further comment. Technically the system separates the two pipelines in two ways. First, it transfers both structures independently to the LF-interface in agreement with a cyclic architecture. Second, some computational operations targeting these structure regard them as independent, as mentioned earlier and elucidated further in the main article. This concerns labeling, sisterhood and other operations. The functions defining these notions therefore refer specifically to adjunct constituents in order to ignore them in such computations. However,

both constituents are still part of the same phrase structure object with some information flow, such as (adverbial) case marking and adverb licensing, occurring between. We can perhaps imagine that the adverbial is "pulled out" from the primary pipeline without breaking constituency. Finally, the notation should not be understood as implying the existence of two separate working spaces; rather, the secondary working pipeline is "secondary" only in relation to the primary working space with which it is linked. These assumptions are all motivated by the fact that they compute the properties of the relevant constructions correctly.

3.3 Negation and infinitival complements

Negation occupies the highest finite clause position and is followed by the tensed verb, which creates the usual T-v-V spread (4). The grammatical subject moves successive-cyclically from SpecNegP into SpecvP.

(4)

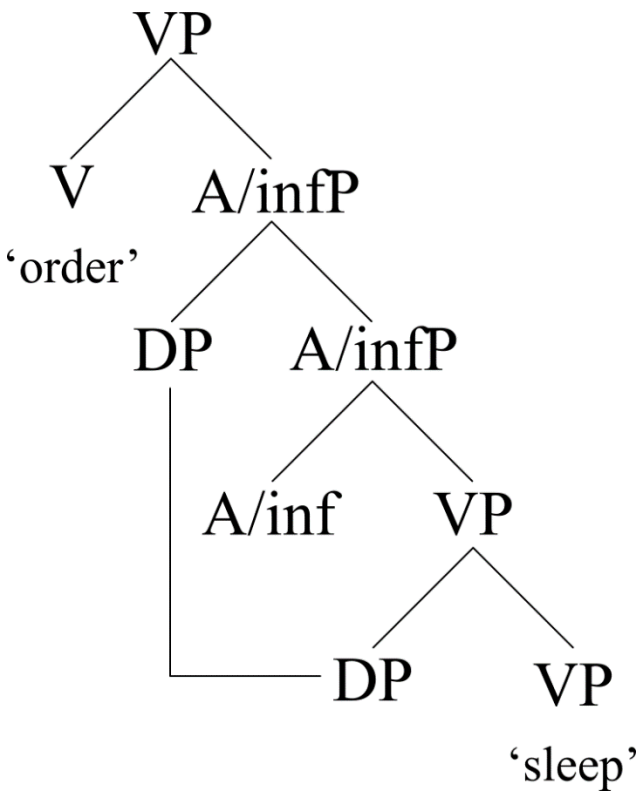


The explanation for successive-cyclic movement is nontrivial due to adjunct promotion and floating mechanisms. Here the subject is promoted into an adjunct due to a failure of Agree-1 at SpecNegP and is floated into SpecTP where it can Agree-1 with the FIN-feature at Neg. Notice that Finnish Neg is a finite element, it exhibits full finite phi-agreement with the subject. SpecTP is not, however, a thematic position. Phrasal A-reconstruction targets it subsequently (adjunct floating is done before A-movement reconstruction) and reconstructs it further down into a thematic position SpecvP. Therefore, the first reconstruction operation differs from the second.[1] In the current implementation, A-reconstruction takes place whenever a thematic argument without a criterial A-bar feature occurs in a non-thematic (EPP) position after possible adjunct reconstruction(s), and it reconstructs the element locally to the specifier position of the next head below (iteratively). The model has not been tested specifically with a comprehensive test corpus probing A-movement, however, so these assumptions must be regarded as tentative. The so produce correct results over a substantial amount of test sentences, both those examined in this study but elsewhere as well.

The A-infinitival was used to examine word order possibilities in infinitivals. Finnish infinitivals have the following structure, correctly captured by the model.

[1] If T has no FIN, Agree-1 cannot be checked at LF (unless we assume T is completely invisible) and reconstruction must rely on locality and selection, as it does here. An alternative is to assume that both Neg and T can check NOM which, under the current system, would require the "participle" T to have FIN, perhaps not an unreasonable assumption since the distribution of this word form (if we include the whole paradigm) is limited to the finite clause. I have experimented with both solutions.

(5)



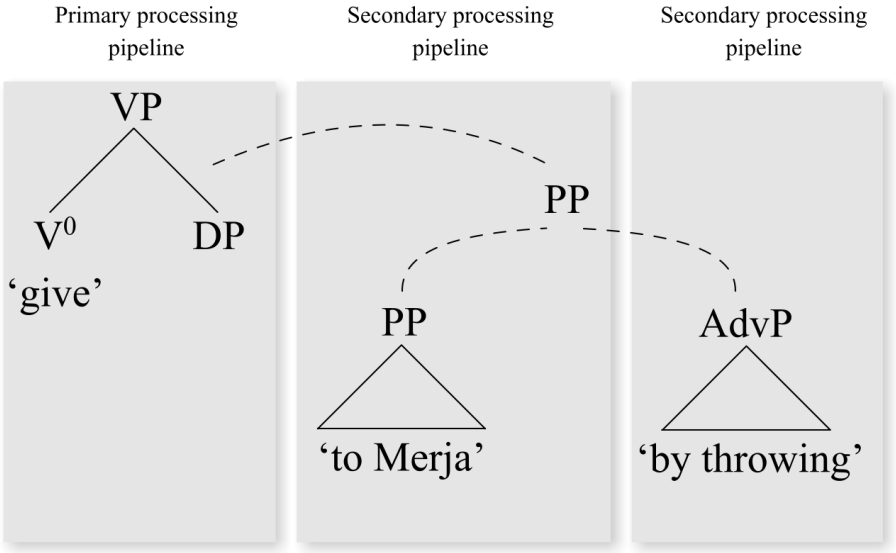
The intransitive A-infinitival verb consists of two parts, the infinitival morpheme and the verb stem (*nukku-a* 'sleep-A/inf'). The bimorphemic verb is decomposed into A/inf + V structure, perhaps analogously to the English *to* + *sleep*, and is reconstructed as shown in (5). The genitive case of the thematic agent of the infinitival is checked locally by -FIN at the infinitival head but notice that genitive DPs were not permitted to float. They are always reconstructed by means of local A-bar/A-reconstruction that uses selection and locality. Adjunct float was prevented by stipulation.[2] This assumption gets the data right but has no underlying motivation. The English pronominal accusative case must be marked similarly, so the mechanism is not specific to Finnish. I do not know what the ultimate explanation is.

[2] The problem is that all case features are checked by Agree-1, which by default licenses adjunct promotion (i.e. case features *are* an adjunct licensing mechanism), thus a separate stipulation is needed to stop the mechanism from operating. In the case of English all case features were ignored in the derivations and no conditions were checked: all arguments were reconstructed on the basis of selection and locality (this is not fully correct but works here).

3.4 Adverbials

Adverbs and adverbials that occur in their canonical positions at the end of the sentence are right-adjoined to the structure. The adverbials used in this study were manner adverbials that were adjoined to a low position in which they could tail the verb (hence they were VP-adverbs). A TP adverbial will be potentially merged higher, although this is not necessary and such structures can become ambiguous depending on the high or low attachment site of the adverbial. When the main clause contains an PP, the result was (6).

(6)



I went through several iterations in which various other assumptions were tested yet none worked as well as this.[3] As suggested in the main article, adjuncts are interpreted as existing in a "secondary syntactic working space." They are geometrical constituents, thus attached to the main structure in the primary space but are invisible for selection and

[3] One question is if the parser should be allowed to generate the symmetric left/right adjunction structure, in which both constituents are adjuncts (PP); and if so, what its label should be. The constituent is anomalous in the sense that it constitutes an adjunct PP that contains another adjunct within, but this representation requires nonstandard assumptions concerning the semantic interpretation of adjuncts, so that the right adjunct 'by throwing' is not interpreted as modifying the PP argument 'to Merja'. The semantic system did not have this problem, as the AdvP was tailing and thus semantically linked with the V. If this solution is blocked by some condition, then another will likely emerge in which both adjuncts are right-adjoined to the DP. The adverbial was promoted into adjunct during the adjunct reconstruction phase, not before; however, here it was at its canonical position and no reconstruction was required.

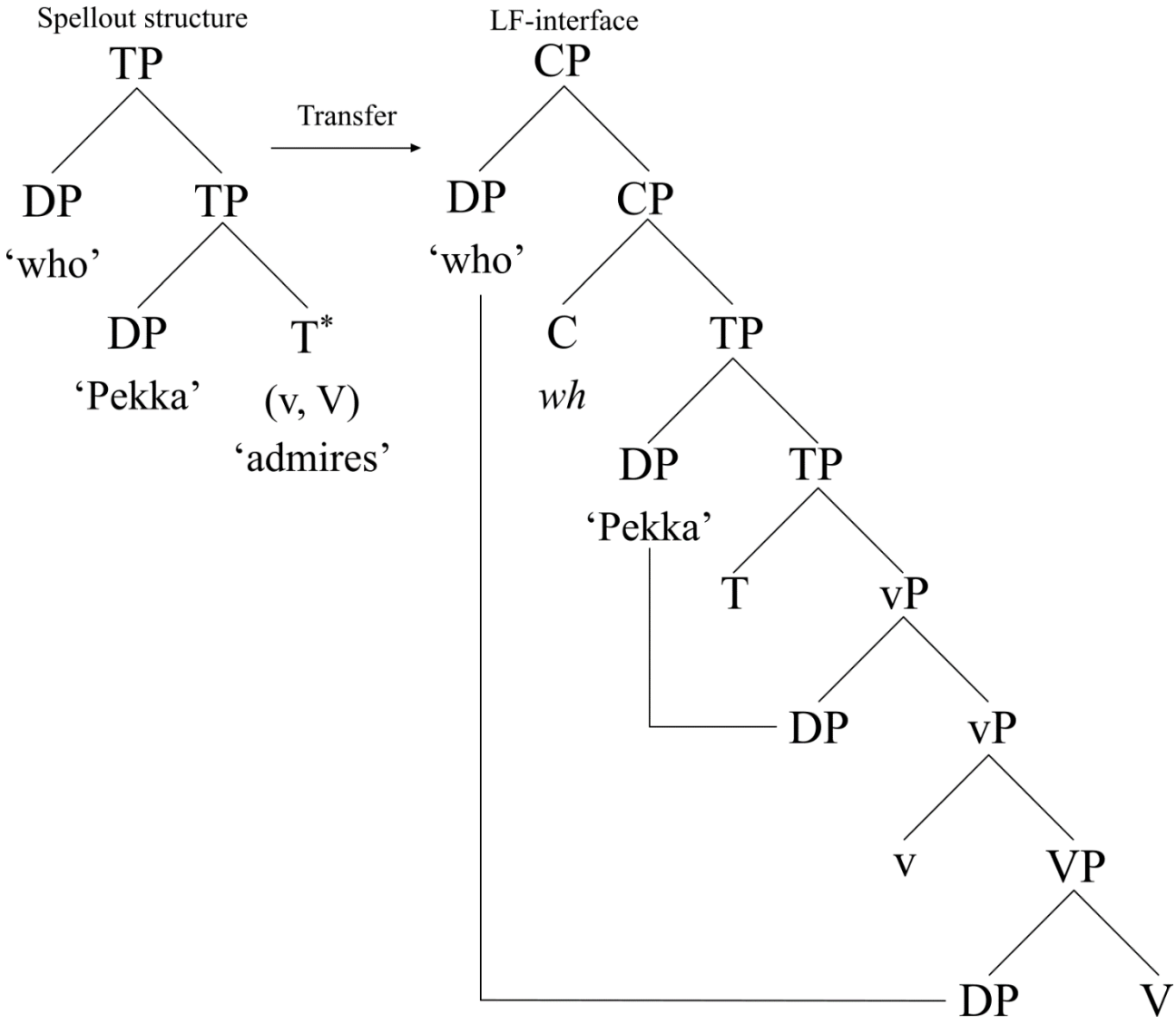
labeling. They are linked and licensed with the main structure by means of tail-head features and can be viewed as introducing an extra dimension to the phrase structure, meaning that any node in the main structure may not only be selected and select, but could also get shadowed by a constituent in the secondary working space. They are interpreted by using the shadowing mechanism as a form of predication. Thus, the manner adverbial is interpreted as denoting a property of the event denoted by the verb.

3.5 C-cartography and interrogatives

The structure of the Finnish C-cartography was relevant in this study because it affects left peripheral word order. A word order is possible in which either a head or a phrase occurs in the C-domain (but not both). The fronted element must come with an overt C-feature (e.g., wh, relative, foc, yes/no question). Interrogatives and focus constructions were included into this study.

Derivation of interrogatives and other constructions with some material in the C-domain implies several special and nontrivial properties under the comprehension perspective adopted here. In Finnish, standard interrogative clause has two phrases in the preverbal position, one which contains a criterial C-feature and another which is the subject (either one may also be missing). The C-head itself is covert and cannot be read off directly from the surface string. Pied-piping means that we cannot generate C-heads directly upon encountering an overt C-feature (as in “towards which river”). This problem was addressed in (Brattico and Chesi 2020). It was assumed that the function of A-bar movement is to represent the phonologically null C-head and its features at the PF-interface. Consequently, the presence of criterial features inside the moved phrase is used to generate the phonologically null C-head (i.e. criterial feature F → generate C with F). If the moved element itself represents the C-head, i.e. if the sentence is verb-initial with some C-feature marked at the verb, then it is interpreted as C*(T...). This is how LF-objects get populated with the C-head and corresponding criterial features. Movement and the nullness of phonological C-head are related to each other; perhaps they are ultimately the same thing. A standard transitive interrogative is derived as (7). Notice in particular the austere status of the spellout structure.

(7)

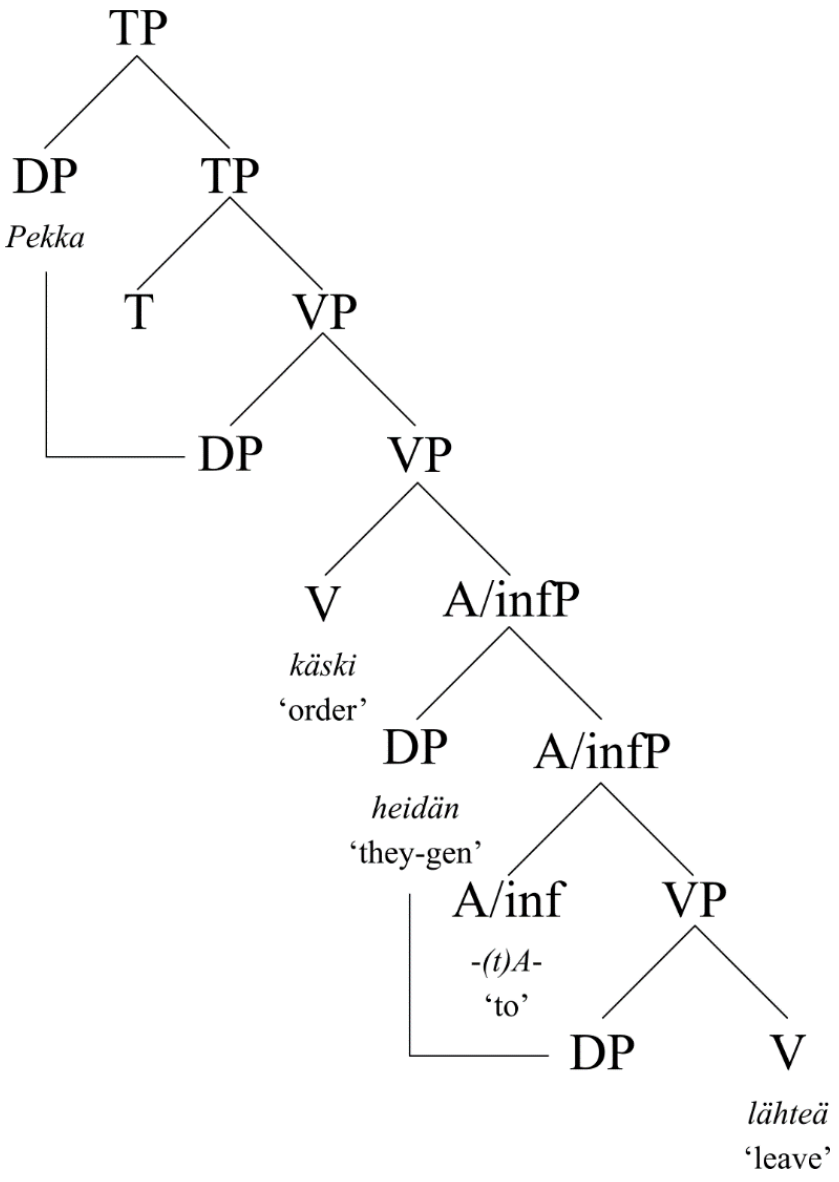


C is generated during transfer and does not exist at SS. The algorithm handles pied-piping, although it was not included into the present study. Wh-interrogatives were tested in connection with negated clauses, infinitival complements and with and without adverbials (file *Study1_seeds_wh_results_FINAL.txt*). Subject interrogative pronouns, direct object interrogative pronouns and genitive interrogatives were all tested and correctly derived. The model reconstructs interrogative pronouns on the basis of selection, locality and case information.

3.6 Infinitival complements

Finnish verbal infinitival complements (non-finite complements) have a basic structure that is represented in this study by the A-infinitival. The infinitival is made up of a verbal stem (hence, a VP shell vP or VP) plus one or two infinitival morphemes that correspond to infinitival heads. These infinitival heads check the genitive case from the external subject and possess a variety of tense-aspect properties but lack mood and other properties of finiteness. Derivation of the canonical infinitival is illustrated by example #365 in this corpus and is illustrated in (8).

(8)



The genitive argument *heidän* ‘they-gen’ cannot be promoted into an adjunct and floated, it can only be A-reconstructed. It is reconstructed into SpecVP position inside the A-infinitival where it receives the agent thematic role. The genitive case is checked against the INF (-FIN) feature at the A-infinitival. The MA-infinitival and the VA-infinitival have the same structure, but they were not included into this study.

4 Deriving standard noncanonical variations

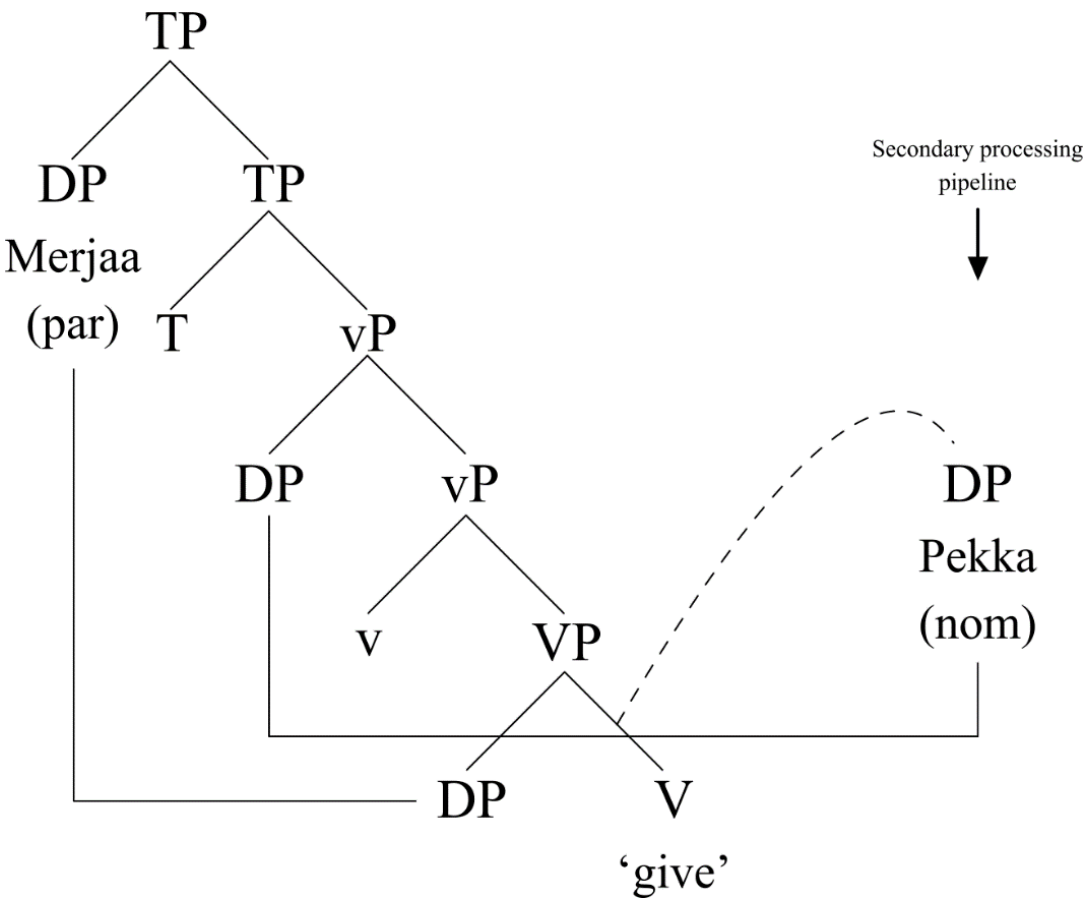
4.1 Introduction

In this section I consider the derivation of standard noncanonical clauses that have been discussed in previous literature. The derivations are taken from Study 0 which contains every word order; the numbers refer to the numbering in that study.

4.2 OVS, topicalization and the EPP

One prominent feature of Finnish is that the label of the first preverbal constituent at SpecTP can be anything. Often the constituent has topic properties, such a definiteness, specificity and/or discourse anaphoricity. We consider first a simple transitive OVS sentence such as *Merjaa ihailee Pekka* ‘Merja-par admires Pekka.nom’ (sentence number #15 in Study 0). The basic idea is that neither argument has a case feature that is satisfied at positions derived from the surface order, hence they are both promoted into adjuncts and floated. The spellout structure is [DP.par T*(v, V) DP.nom] which generates (9) by transfer. This is the analysis shown in the results file (#15).

(9)



The grammatical subject is first positioned into the CompVP position, where it is on the basis of surface order. The nominative case is not licensed in this position, because the DP is not locally c-commanded by a functional head with feature FIN. It is c-commanded by v at this position and flagged as illicit. It is promoted into an adjunct, as shown above, and floated into SpecvP, where its case is licensed. The same applies to the partitive DP in the preverbal topic position. Floating is implemented by searching the clause in a top-down order (ignoring left branches that have been transferred independently as phases) for the first position in which the case can be licensed, the positions shown above. Because the grammatical subject is promoted into an adjunct, it does not affect labeling and selection. It thus behaves like any DP-adverbial, such as *the whole day*.

Neither the conceptual-intentional system nor the LF induce any kind of topic reading for the surface subject *Merja-a* ‘Merja-par’. This would be trivial to add to the interpretation; there is a separate module taking care of semantic interpretation of the LF output (module “semantics.py”). The problem is that the constituent need not be the topic. This issue,

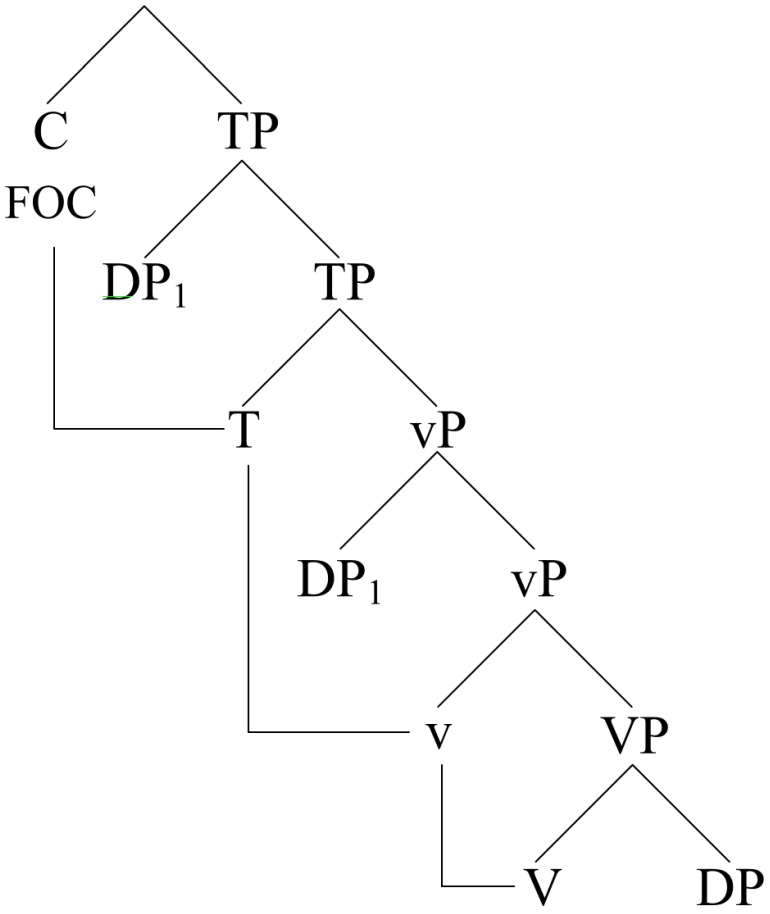
which is controversial at present, was ignored in this study. What is not controversial, though, is that Finnish finite clause enforces a strict EPP regimen. In the first published formulation of the present model the EPP was implemented by stipulation: some heads had an EPP feature that required that a phrase fills in their specifier position (Brattico and Chesi 2020). The feature is checked at LF, because a copy of the reconstructed element remains at the EPP position. The EPP was further reduced into a non-selective !SPEC:* feature which requires that a specifier with any label (*) is present. Thus, a sentence in which the position is empty is judged as ungrammatical (see, e.g., #11). This means that the EPP system adopted in this study was purely formal, having no semantic function.[4] Notice that the adjunct float operation will not position anything into SpecTP; it tries to satisfy case features.

4.3 Verb-initial clauses

The derivation of a transitive verb-initial clause is presented in #9-10. An unmarked transitive verb is interpreted as T*(v,V). When it occurs as the first element in the clause, spellout structure [T*(v, V) [DP.nom DP.par]] will be generated in which the grammatical subject cannot be licensed due to the lack of FIN at v. It is floated into the correct thematic position SpecvP. The sentence fails because the EPP feature of T is not checked (#9). When the verb comes with a C-feature, here FOC, the spellout structure will be [C*(T,v,V) [DP DP]] which reconstructs into grammatical (10)(#10).

[4] The assumption is most likely too simplistic. Following the insight of (Holmberg and Nikanne 2002) that the EPP position must be filled in by a "broadly referential" constituent, I argued in a recent paper that the EPP is involved in checking the nominal D-feature (Brattico 2019b). If so, then the checking mechanism must be criterial, as the D-feature need not occur at the edge of the phrase at SpecTP. This system, which would unify D-checking (EPP) with *wh*-checking, remains to be formalized, implemented and tested.

(10)



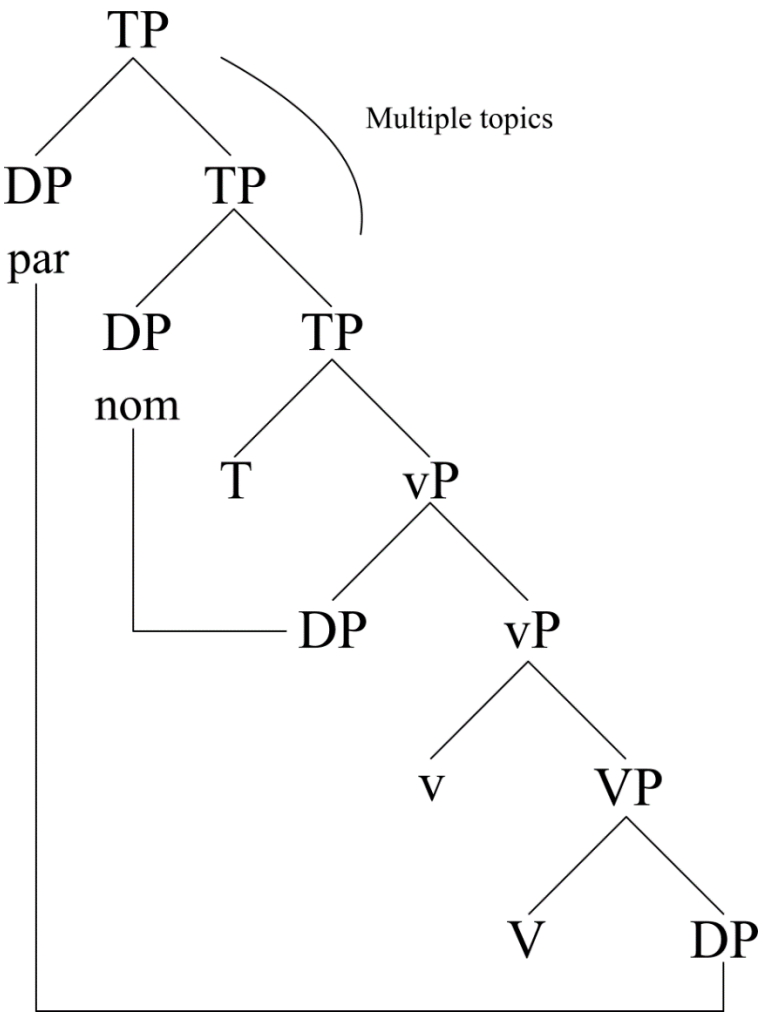
C*(T, v, V) is reconstructed by head reconstruction into the standard C-T-v-V spread. The nontrivial part is the positioning of the grammatical subject. Head reconstruction must "skip" it in order to have something into the EPP position. This is implemented by a condition in head reconstruction algorithm which tries to satisfy EPP. T*(v,V) is merged below the grammatical subject for this reason. If the subject and object appear in reverse order (*IHAILEE Merja-a Pekka* 'admire.foc Merja-par Pekka.nom'), the derivation as it as below, but the adjunct float operation additionally shifts the position of the S and O exactly as shown in Section 4.2 (#12).

4.4 Multisubject/topic constructions

One empirical puzzle this study aimed at solving was the presence of multi-subject/topic constructions such a *Merja-a Pekka ihailee* 'Merja-par Pekka.nom admires' (#13, see also #7). They are accepted as grammatical because the model promotes both preverbal subjects

into adjuncts and floats them into canonical positions. Because they are interpreted as adjuncts, the structure need not be analyzed as containing two non-adjunct specifiers.

(11)



Same analysis is applied to sentences in which the topics are followed by a negation, for example. Thus, sentence (12) is analyzed as (13), with Pekka and Merja being two DP adjuncts inside NegP.

(12) Pekka Merja-a e-i ihaile. (#85)
Pekka.nom Merja-par not-3sg admire
'Pekka (=topic) does not admire Merja (=topic).'

(13) $[_{NegP}$ (Pekka$_1$) $[_{NegP}$ (Merja$_2$) [Neg^0 [__$_1$ T [__$_1$ [v __$_2$ admire]]]]]]

It applies in the case of three pre-negation topics (#169), but the model judges these sentences are marginal due to the increased parsing load. Infinitival multi-topic constructions are also possible, as shown in (14)(#659).

(14)

a. Pekka käski [heidän kirjan Merjalle [anta-a]]. (#659)
   Pekkka orederd they.gen book to.Merja give-A/inf

b. Pekka order [they$_1$ [⟨book⟩$_2$ [⟨PP⟩$_3$ [A/inf [__1 V __2 __3]]]]

The semantic relevance of stacking several arguments into the edge of the A-infinitival (or any other infinitival) remains a puzzle. To me these sentences are possible by marginal, but with no clear semantic attribute associated with the pre-infinitival arguments.

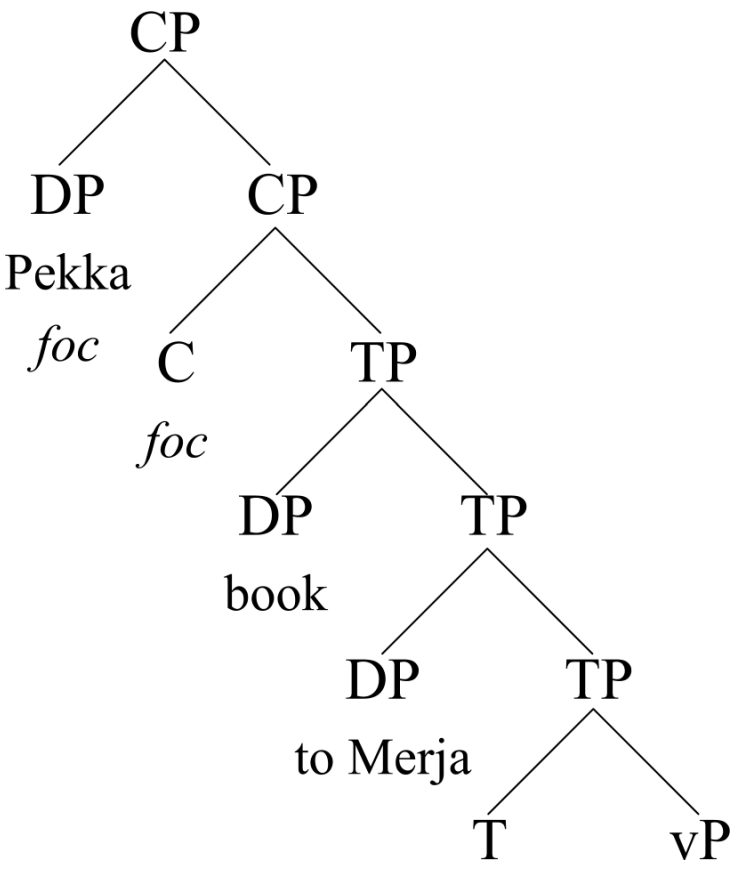
The presence of adjunct specifiers presents also challenges. One is that all functions that refer to the notion of "specifier" must be written in such a way that they take the possible occurrence of multi-spec/left adjunction structures into account. In the current model, this is done by using the notion of "edge" instead of specifier, with edge containing all phrases to the left until the label changes (plus a possible pro-element at the head, ignored in this study). But the general idea is clear from this example: while it is not possible to stack non-adjunct specifiers, adjunct stacking is accepted (e.g. stacking of adjectives, adverbials). Multitopic and multisubject sentences contain 'stacked argument adjuncts', in the same way as both adverbials and adjectives can be stacked.

Another theoretical issue concerns the status of left adjuncts. While right adjuncts are invisible for selection and labeling, left phrases have some of these properties even if they were not adjuncts. For example, they do not determine labels of the nodes that contain them. While complex left phrases can be selected, should complex left adjuncts be selected as well? Because one of the adjuncts present in the above analysis must satisfy the EPP/SPEC* feature, we must assume that left adjuncts can satisfy selection. In this they differ from right adjuncts.

A sentence with three subjects/topics is presented at #23, with the third and lowest subject/topic being a PP *Merjalle* 'to Merja'. The PP is floated and reconstructed into a right-adjunct position to the vP. Otherwise the derivation is as above. Notice that all

three argument constitute left adjuncts, as none of the case features are licensed in this position. If the highest phrase is marked for a C-feature, such as FOC, it will be analyzed as a phase at SpecCP with the C-head generated as explained earlier (#8, 14, 16 and others). A sentence *PEKKA kirjan Merjalle antoi* 'Pekka.foc book-acc to.Merja gave' (#27) is thus analyzed so that 'Pekka' is at SpecCP, with 'book.acc' and 'to Merja' as the two subject topics at the preverbal field, thus at the edge of TP (15)(#24).

(15)



Despite being at SpecCP, the subject is still interpreted as an adjunct and is reconstructed by adjunct float. A-bar reconstruction is not applied, even though the argument is in an A-bar/operator position. This is because adjunct float is applied before A-bar reconstruction in comprehension in this model. The chain between SpecCP and SpecvP is still available because the reconstruction leaves copies. It is possible to think that an A-bar chain is created during production, followed by a "trivial" adjunct chain (SpecCP, SpecCP) than functions to promote the element into an adjunct.

Reconstruction of C*(T, v, V) never creates multitopic constructions. The head is reconstructed so that only the highest phrase is captured to SpecT*P (16).

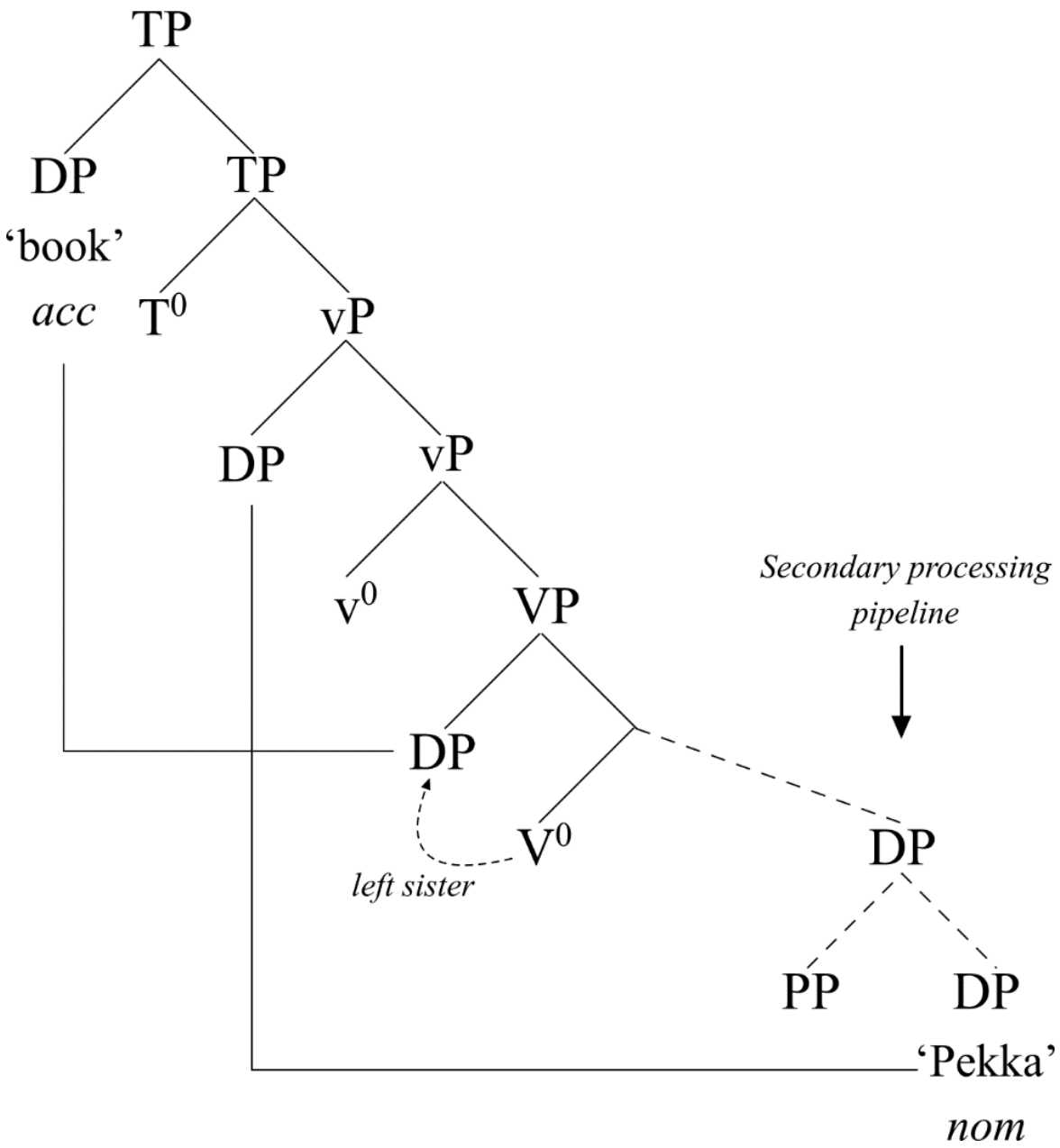
(16) ANTOI$_1$ kirja-n T_1 v V Pekka Merjalle.
GIVE book-acc T v V Pekka.nom to.Merja
'Pekka GAVE a book to Merja.'

This represents the configuration after head reconstruction; the postverbal subject is floated into SpecvP. The algorithm interprets multitopic constructions as marginal due to discourse complexity. These sentences are mildly marginal, but grammatical under a special context in which the preverbal elements all received a topic interpretation. The topic interpretation is generated by the conceptual-intentional system and corresponds to no grammatical feature.

4.5 Postverbal orders

A ditransitive V-O-IO structure is interpreted so that the IO argument is attached to the VP as a right adjunct, as explained before. Generation of V-IO-S order is nontrivial. The derivation proceeds by interpreting IO-S first as a slightly anomalous right adjunct DP, with PP at its Spec. Then, subject DP is floated to SpecvP, while the PP remains in situ. The remnant [$_{DP}$ PP __] constitutes an adjunct, so V does not see it and thus selects the reconstructed DP at its left as its complement (and specifier, accidentally).

(17)



This example shows why left complements are important. If we do not assume them, then this derivation will never converge: the verb cannot satisfy its complement selection feature and semantic interpretation fails to determine the patient of the transitive verb.

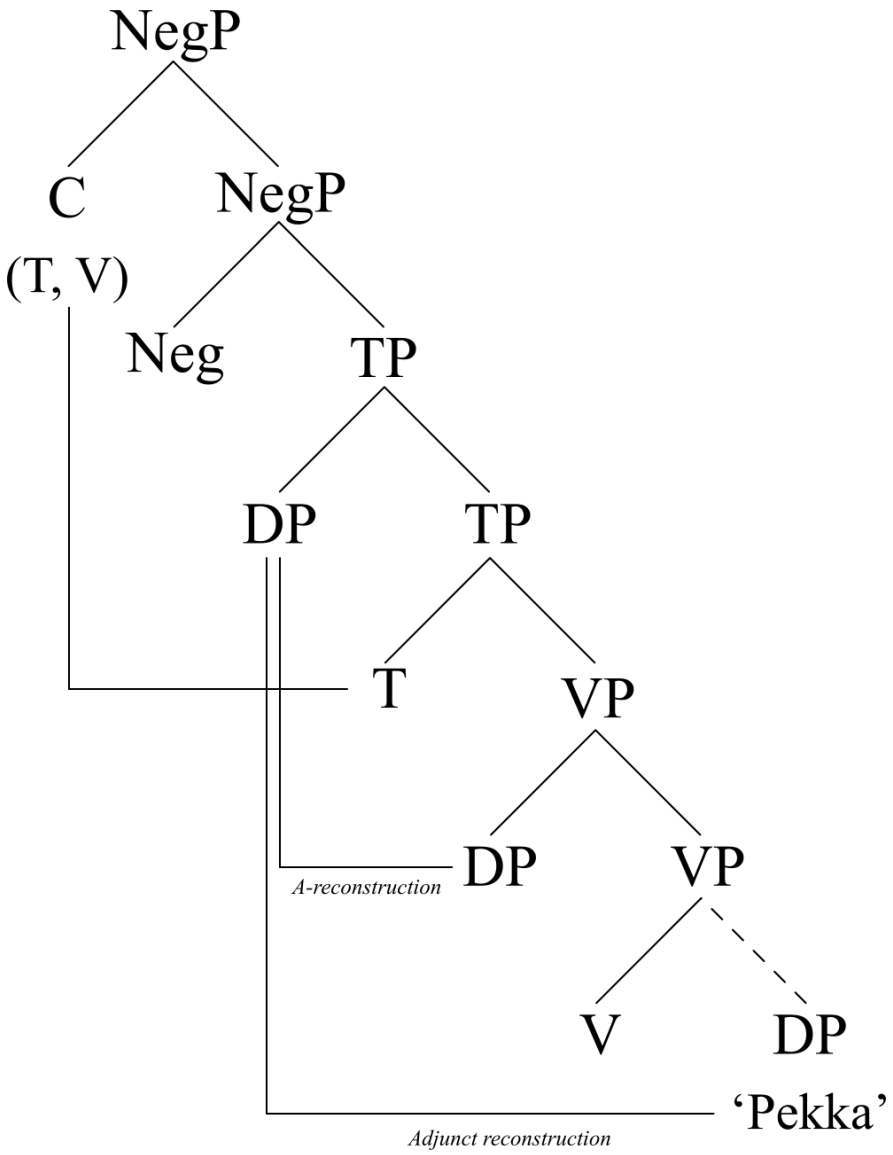
4.6 Nonlocal head movement

Finnish exhibits nonlocal head movement whose properties resemble phrasal A-bar movement. Based on a large catalog of facts concerning nonlocal head movement in Finnish, I proposed earlier (in a currently unpublished manuscript) an analysis in which a head that is marked

by an overt C-feature (and only such heads) can be reconstructed downwards into a first position in which it can be selected. This allows the system to derive nonlocal head movement constructions in the current corpus. Consider the derivation of (18), shown in (18).

(18) Nukkunut e-i Pekka. (#76)
sleep.past not-3sg Pekka.nom
'It was SLEEPING that Pekka (=information focus) did not do.'

(19)



The complex head C*(T,V) is reconstructed over the local head Neg, as shown by the bold arrow, because this is the first position in which T*(V) can be selected (namely, by Neg).

The postverbal grammatical subject is promoted into an adjunct and first floated to SpecTP, in which it satisfied the nominative case by Agree-1. From this position it is lowered into SpecVP by local A-reconstruction, which is a thematic position. Notice that since the postverbal DP is an adjunct, V can select this constituent both as its complement and specifier. Obviously, V is extracted from T*(V), which is not shown above. All nonlocal head movement constructions were derived similarly (e.g., #104, 110, 112, 594). Nonlocal head movement is available only for heads that were marked by an overt C-feature, here FOC.

(20)	IHAILLA$_1$	Merja-a	Pekka	käski	heidän	__$_1$ (#594)
	admire-A/inf.foc	Merja-par	Pekka.nom	ordered	they.gen	

One noteworthy feature of the above analysis is the fact that Neg does not satisfy the EPP feature. This will normally lead into ungrammaticality. Here the EPP was not in operation because my judging of the large corpus of Finnish word orders revealed, as pointed out in the main article, that T-to-C movement knocked out the EPP principle. This condition was added to the model (as an unexplained condition), hence we see in the above example that no phrase needs to appear at SpecNegP. The EPP condition, and the above exception, was checked at the LF-interface. A reviewer disagreed with some of my judgments, but the matter cannot be decided by looking only at the few examples reported in the main article. One has to look at the whole corpus. The matter was mentioned in a footnote in the main article. It is important to keep in mind, however, that the EPP principle was knocked off in cases in which a head had moved over T as this does capture my grammaticality intuitions, which constitute the gold standard used in this study. I do not doubt my own judgments and suspect that the reviewer was misinterpreting the examples.

5 Canonical ungrammatical sentences and how they are judged as such

5.1 Wrong order between heads

One category of ungrammatical word orders in Finnish is constituted by sentences in which the word order between two or more heads is wrong. This statement must be prefaced with the remark that the word order between two heads may change if one of them is moved into the C-domain and is reconstructed downwards. We have already examined these cases. The key feature in such cases is the presence of an overt feature or morpheme (e.g., foc, wh)

signaling the presence and content of the C-head. The C head must in such cases occur in a position in the surface string in which it can be reconstructed while satisfying selection. If no such feature is present and the (monomorphemic) head occurs in an unmarked form, no head reconstruction takes place and the head is positioned into the structure on the basis of its spellout structure position. Consider example (21) (#67), which the model rules as ungrammatical.

(21) *Pekka nukkunut e-i.
Pekka sleep.past not-3sg

The spellout structure for this sentence is [DP [T*(V) Neg]], which reconstructs into [DP [T [V Neg]]]. The verb is extracted from within the complex head but positioned below T, as T selects it. This representation is ruled ungrammatical at LF-interface because it does not satisfy selection features of Neg. All other possible derivations fail for similar reasons (see e.g., #81, 82, 83, 84, 87, 88, 107, 121, 122). The following two derivations are possible in theory: [[DP T] [V Neg]] and [DP [T V] Neg], neither which satisfies LF-legibility.

Interestingly, I judged some of these sentences as marginal. The fact remains that in general we must rule out wrong head orders, but in some cases, especially if the heads are adjacent or close to each other, and when the structure is otherwise relatively simple, the structure is grammatical but comes with a feeling of an idle “word play.” The word order above does not trigger discourse interpretations. I don’t know how to explain this phenomenon. One possibility is that they are produced by an extralinguistic stylistic component. The assumption is nontrivial, and I do not at present have any solution of how to implement such process formally (if they exist in the first place). One possibility is to allow the algorithm to permutate surface order within strict boundaries and try to parse them. Another possibility is that these cases are handled by general cognition allowed to interpret the structure under deviation from the norm. The same mechanism could interpret sentences with simple grammatical errors, such as S-V agreement mistakes, for example.

5.2 Genitive DP and rightward movement

Genitive DPs were prevented from floating, by a lexical feature associated with the genitive case itself in the surface vocabulary. This does not prevent them from undergoing A-

reconstruction or A-bar reconstruction. Therefore, all examples in which the genitive is located rightwards of its own predicate are judged correctly as ungrammatical, since adjunct float is the only displacement mechanism that reconstructs phrasal arguments into any direction, including leftwards (rightward/downward in the standard theory)(22)a (see #367, 373, 375, 379, 385, 437-448 for a few examples). In contrast, topicalization of the genitive argument is possible (b), because the argument can be A-reconstructed into its canonical position, as shown in the analysis provided by the algorithm (c). A-bar reconstruction of the genitive argument is illustrated by example (d)(#390).

(22)

a. *Pekka käski nukku-a heidän. (#367)
Pekka.nom ordered sleep-A/inf they-gen

b. PEKKA heidän$_1$ käski __$_1$ nukku-a. (#370)
Pekka.foc they-gen ordered sleep-A/inf
'It was Pekka, not Jukka, who ordered them (=topic) to sleep.'

c. PEKKA$_2$ C heidän$_1$ T __$_2$ V __$_1$ A/inf __$_1$ V (#370)
——————A——————→ ————A————→

d. HEIDÄN Pekka käski nukku-a. (#390)
they-gen.foc Pekka.nom ordered sleep-A/inf
It was them, not us, who Pekka ordered to sleep.'

6 Controversial and nontrivial issues

6.1 Verb-initial clauses by TP fronting

The model analyses sentence (23)(#225) as grammatical but extremely marginal.

(23) ##Antanut Pekka kirja-a e-i Merjalle.
give.past Pekka.nom book-par not-3sg to-Merja
'Pekka did not give a book to Merja.'

It finds the analysis (24).

(24) [$_{NegP}$ [$_{TP}$ T [give Pekka$_1$]]$_2$ [(book) [not [__$_1$ __$_2$]]]](to Merja)$_4$

The analysis has several exceptional features. First, it illustrates TP-fronting, in which the whole TP = *antanut Pekka* 'give.past Pekka.nom' is fronted to SpecNegP, while the DP 'the book' occurs as a left adjunct inside the same projection. This construction was accepted, because there was no selection feature at Neg in the lexicon that ruled out a TP specifier (that feature would have been -SPEC:T). This could be added trivially, but the sentence is acceptable to me. The assumption, then, is that both 'Pekka to give' and 'the book' are topics. The second interesting point is that the algorithm leaves room from movement-within-movement: once the TP is reconstructed to the complement position of Neg, the arguments are floated to their thematic positions, which means that Pekka, for example, is moved out of the TP, while the PP is moved inside it. Both operations therefore target a phrase that has already moved. The interesting thing about this is that the ordering between adjunct float and A-reconstruction (TP reconstruction here) is always 'adjunct float => A-reconstruction', which is seemingly violated here. The reason this analysis is possible is because the PP is merged to the right of the rest of the structure, which, because all left branches are derived independently, means that the TP is fronted *during that operation* and the arguments can be then floated *during the transfer of the whole structure*. The process is illustrated below:

(25)

a. Reconstruct NegP as a phase upon Merge(NegP, PP)

   $[[_{NegP}\ TP_1\ DP\ Neg\ \_\__1](PP)]$

b. Reconstruct arguments during final transfer

   $[TP_1\ [DP_2\ [Neg\ [_{TP}\ .\ .\ .\ DP_2\ .\ .\ .\ DP_3.\ .\ .\ PP_4.\ .\ .]_1]]](PP)_2$

What makes this movement-within-movement analysis possible is therefore the fact that under this parse the PP is first merged to the top right position, triggering and licensing operation (25)a.

6.2 Infinitival topicalization

The algorithm finds a solution for sentences in which an infinitival has been topicalized. It will also correctly reconstruct A-movement inside the topicalized A-infinitival (which implements a movement-inside-movement construction).

(26) Pekka [heidän nukku-a] käski. (#371)
Pekka.nom they.gen sleep-A/inf ordered
'Pekka asked them to sleep.'

According to my judgment, this result is correct; I accept these sentences. The algorithm provides (26) with the (correct) structure (27).

(27) $[_{TP}$ ⟨Pekka⟩$_3$ [$[_{A/infP}$ they.gen$_1$ A/inf __$_1$ sleep]$_2$ [T __$_3$ order __$_2$]

It reconstructs the genitive external argument inside the A-infinitival before moving the fronted phrase to its canonical CompVP position due to the assumption that all left branches are transferred to LF as a phase. The same assumption captures internal wh-movement and pied-piping in Finnish. By using the same logic, the model is able to analyze sentences that involve two movement steps inside the fronted infinitival. These come with a heavy parsing load and are judged as clearly marginal (??) by the model.

(28) ⟨Pekka⟩ [Merjaa$_1$ heidän$_2$ ihailla __$_2$ __$_1$]$_3$ käski __$_3$. (#455)
Pekka to.Merja they.gen admire-A/inf order

(29) ⟨Pekka⟩ [heidän kirjan antaa] käski __ Merjalle.(#72)
Pekka they.gen book give-A/inf order to.Merja

7 Discourse interpretation of noncanonical word orders

7.1 The C-domain and focus

Some word order permutations give rise to robust semantic effects. Head- or phrasal movement to the C-domain results in an interpretation that depends on the overt feature attached to the moved element. If the feature is missing, speakers will interpret the moved element as expressing correction or contrastive focus/topic (30). In spoken language this is typically associated with prosodic effects, such as stress on the moved element, and especially if the special interpretation is not emphasized in the conceptual content of the sentence.

(30)
a. Auton Pekka osti (ei mopoa)
Car.foc Pekka bought (not motorbike

’It was a car, and not a motorbike, that Pekka bought.’

b. MYI Pekka __ koko omaisuutensa!
sold.foc Pekka all possessions
‘Pekka DID sell all his possessions.’

This has led to an analysis, assumed in the present work as well, in which the C-domain is associated with special interpretation effects that consists of a bare operator-variable construction (perhaps best illustrated by the relative clause) combined with discourse interpretation arising from C-features that may be used optionally. Few examples are shown in (31).

(31)

a. Ketä Pekka ihailee __?
who Pekka admires
’Who does Pekka admire?’

b. Ketä-hän Pekka ihailee __?
who-hAn Pekka admires
’I wonder who Pekka admires?’

c. Ketä-kö-hän Pekka ihailee __?
who-Q-hAn Pekka admires
’I wonder who Pekka admires?’

d. Merja-a-ko-han Pekka ihailee __?
Merja-par-Q-hAn Pekka admires
‘Is it Merja who Pekka admires?’

The model is able to interpret and reconstruct these sentences correctly, as shown in Study 1, because it recognizes that the first element carries a criterial feature (e.g., wh, Q, hAn) that then triggers A-bar reconstruction or head reconstruction. The criterial feature is copied to C and can therefore be selected. It also affects compositional semantic interpretation at LF. But because relative pronouns move to the same SpecCP position and cannot be combined with discourse features, I do not believe that the C-domain and associated grammatical mechanism can be defined by relying on discourse interpretation and/or discourse features (such as focus); rather, the position itself is part of a

mechanism that implements an operator-variable construction, with the discourse features and interpretation as an additional semantic dimension.

7.2 TP-domain and the topic

The Finnish preverbal subject position is usually interpreted as a topic position. In this work I explored constructions with several preverbal subjects, which were analyzed as adjuncts. They all tend to receive a topic interpretation. For example, sentence (32) is felicitous in a context in which Pekka, Merja and the car are well-known from previous discourse.

(32) Miksi Pekka Merjalle auto-n lainasi?
why Pekka.nom to.Merja car-acc borrowed
'Why did Pekka borrow the car to Merja?'

On the other hand, the topic interpretation is not necessary. Non-topics occupy the same position(s).

(33) Joku rikkoi tuon auton.
somebody broke that car
'Somebody broke that car.'

The topic position hypothesis is therefore controversial and, in my view, incorrect. I have argued that these sentences exhibit a definiteness effect, and only in the case of nonsubject fronting. But because it is uncontroversial that the preverbal field is not restricted to topics, the currently analysis, contrary to (Holmberg and Nikanne 2002), has no grammatical topic features (e.g., topic, non-focus) that would drive operations. The preverbal field is filled in by an ordinary EPP mechanism. The interpretation associated with (32) is assumed to arise in the interpretative component, either the at LF-interface or in the postsyntactic conceptual-intentional system. The definiteness hypothesis was not part of the present analysis either.

7.3 Postverbal positions and information focus

Noncanonical postverbal position is associated with information focus position, thus it contains the "main news" of the sentence.

7.4 Infinitival fronting

Infinitival fronting, exhibited in (34), is not associated with a robust semantic effect (when the phrase does not move to SpecCP).

(34) ??Milloin [heidän osta-a lahjan] Pekka käski __?
when they-gen buy-A/inf present Pekka ordered
'Why did Pekka order them to buy a present?'

This could be related to the fact that only referential arguments and phrases can receive a topic interpretation at the preverbal SpecTP position. It is possible that the event denoted by the infinitival must still be known in the discourse, but the intuition is less clear than in the case of referential arguments. Perhaps "indefinite infinitivals" are less grammatical in this position (?* *Milloin heidän tehdä jotain Pekka käski* 'when they.gen do something Pekka ordered'). It still seems possible, though, that infinitival fronting constitutes an "idle word play" that is made possible by the grammar but has no use.

7.5 Multitopic infinitivals

Another word order manipulation that seems possible but not associated with clear semantic effects are multitopic infinitivals. Sentence (35) provides an example, with 'Merjalle' being an extra subject/topic inside the A-infinitival. These extra topics are adjoined within the infinitival projection, as shown in the analysis.

(35) Pekka käski [$_{A/infP}$ heidän [$_{A/infP}$ Merjalle$_1$ anta-a kirjan __$_1$]].
Pekka ordered they-gen to.Merja give-A/inf book-acc
'Pekka ordered them to give a book to Merja.'

I cannot associate these word order permutations, grammatical as they are, with any clear semantic interpretation.

8 Conclusions

A fully formal, algorithmic grammatical theory allows one to rigorously test the logical consequences of a scientific hypothesis or grammatical analysis by running the theory over a set of test sentences. There is often no other way to fully justify an analysis of this kind, as they tend to depend on complex interactions of several principles yielding many

"emergent" properties that are not obvious even to those who participated in the creation of the model in the first place. Thus, the analysis proposed in this work was the result of several trial-and-error runs in which various intuitively appealing but ultimately wrong technical ideas were first tried but then shown to lead into unexpected problems in simulation experiments. Consequently, the analysis of Finnish provided by the algorithm, as reported here, should not be taken as something that the author constructed specifically for the purposes of this study; rather, it should be interpreted as a logical consequence of the abstract hypotheses put forward in the main article. By the same token, the only justification for the model is its ability to calculate the dataset correctly and provide correct semantic interpretations and syntactic structures for the test set. Although many of the assumptions could be criticized on some external grounds – for example, the assumption that the semantic system operates with a binary branching bare phrase structure, that there exists an autonomous language systems, or innate grammatical rules, could all be viewed with skepticism – such criticism is irrelevant unless it is tested by using the same standard, hence unless it too is implemented as a formal algorithm and tested over the same or very similar set of sentences. It would indeed be interesting to see how a model of human language comprehension that does not operate with phrase structures, innate rules or language-specific computations could satisfy even the minimal requirement of observational adequacy.

REFERENCES

Brattico, Pauli. 2019a. *A Computational Implementation of a Linear Phase Parser. The Framework and Technical Documentation*. Pavia.

Brattico, Pauli. 2019b. "Subjects, Topics and Definiteness in Finnish." *Studia Linguistica* 73:1–38.

Brattico, Pauli and Cristiano Chesi. 2020. "A Top-down, Parser-Friendly Approach to Operator Movement and Pied-Piping." *Lingua* 233:102760.

Chomsky, Noam. 2000. "Minimalist Inquiries: The Framework." Pp. 89–156 in *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, edited by R. Martin, D. Michaels, and J. Uriagereka. Cambridge, MA.: MIT Press.

Holmberg, Anders and Urpo Nikanne. 2002. "Expletives, Subjects and Topics in Finnish." Pp. 71–106 in *Subjects, Expletives, and the EPP*, edited by P. Svenonius. Oxford: Oxford University Press.

Nikanne, Urpo. 1993. "On Assigning Semantic Cases in Finnish." Pp. 75–87 in *Case and other functional categories in Finnish syntax*, edited by A. Holmberg and U. Nikanne. Mouton de Gruyter.