Computational analysis of Finnish nonfinite clauses

Pauli Brattico

2023

Abstract. Finnish nonfinite clauses constitute a complex grammatical class with a seemingly chaotic mix of verbal and nominal properties. Thirteen nonfinite constructions, their selection, control, thematic role assignment, nonfinite agreement, embedded subjects and syntactic status were targeted for an analysis. An analysis is proposed which derives their syntactic and semantic properties by relying on a computational model of human information processing. The model analyses Finnish nonfinite constructions as truncated clauses with one functional layer above the verb phrase. Research methods from naturalistic cognitive science and computational linguistics are considered as potentially useful tools for linguistics.

Keywords: Finnish, nonfinite clause, participles, infinitives, computational modeling, information processing, computational linguistics

1 Introduction

Finnish infinitives and participles constitute a complex and large syntactic class with a mix of nominal and verbal properties (Ikola 1974, Hakulinen & Karlsson 1979:Chapter 14, Wiik 1981, Vainikka 1989, 1995; Toivonen 1995, Koskinen 1998, Vilkuna 2000:Chapters 8.1, 9; Ylikoski 2003, Visapää 2008, 2022; Ylinärä 2018, Kiparsky 2019, Jussila 2020). Application of computational modelling to this class suggests, however, that it constitutes a homogeneous natural kind based on a truncated clause with one functional layer above the verb phrase.

The argument is organized as follows. Section 2 introduces the grammatical class of Finnish nonfinite clauses as a whole, Section 3 then defines the research agenda by focusing on five syntactic properties, namely selection, grammatical role, control, agreement, and the presence of overt subjects. Section 4 develops the hypothesis. A Python-based model of human language apperception is proposed and employed to analyze and classify Finnish nonfinite clauses. The approach is multidisciplinary and combines ideas from computational linguistics, cognitive science and linguistics. In Section 5, the hypothesis is put into a rigorous test by means of a computational experiment. Section 6 discusses certain additional topics while Section 7 summarizes the main conclusions.

The source code and the raw input/output files associated with this study are available in the source code repository.[1] There is also a supplementary document which addresses issues that are technical in nature and provides further instructions on how to work with the underlying source code and the raw data.

2 FINNISH NONFINITE CLAUSES

Finnish is an agglutinative, suffixing language and forms several types of deverbal predicates by combining verbal stems with suffixes. DEVERBAL NOMINALS are derived by suffixing the stem with one of the many nominalizing suffixes (e.g., *osta-minen* ‘buy-ing’, *ost-o* ‘purchase’). The results behave like ordinary nouns and noun phrases. They are put aside in this study, together with most of nominal syntax. DEVERBAL ADJECTIVES are derived by suffixing the verbal stem with one of the adjectivizer suffixes (e.g., *tutki-maton* ‘explore-without’, i.e. ‘unexplored’). Some deverbal adjectives project nonfinite clauses called PARTICIPLES or PARTICIPLE ADJECTIVES (*tutki-va* ‘explore-VA/A’ meaning ‘x who explores y’ and *tutki-ma* ‘explore-MA/A’ meaning ‘x explored by y’). Participle adjectives project clause-like structures that can contain direct objects, thematic subjects, adverbial modifiers and even other nonfinite clauses. Because the participle adjectives have several properties that the rest of the nonfinite clauses do not have, they are treated as a separate matter in Section 5.2.5.

Once we put the deverbal nominals and adjectives aside, a residuum of deverbal predicates and nonfinite clause structures projected from them remain that are neither nouns nor adjectives; rather, they seem to exhibit a mixture of nominal and verbal properties. We focus on this group. This class contains, to begin with, a group of MA-INFINITIVES with five different deverbal predicates made of the *-mA* morph (bolded in the examples below) followed by a semantic case suffix (ILL = illative 'into'; ABE = abessive 'without'; INE = inessive 'inside'; ELA = elative 'from'; ADE = adessive 'on/at')(1).[2]

(1) a. Pekka pyys-i Merja-a [pese-**mä**än ikkuna-t].
Pekka.NOM ask-PST.3SG Merja-PAR wash-MA.ILL/INF window-PL.ACC
'Pekka asked Merja to wash the windows.'

b. Pekka läht-i [pese-**mä**ttä ikkuno-i-ta].
Pekka.NOM leave-PST.3SG wash-MA.ABE/INF window-PL-PAR
'Pekka left without washing the windows.'

c. Pekka näk-i Merja-n [pese-**mä**ssä ikkuno-i-ta].
Pekka.NOM see-PST.3SG Merja-ACC wash-MA.INE/INF window-PL-PAR
'Pekka saw Merja washing the windows.'

d. Pekka kiels-i Merja-a [pese-**mä**stä ikkuno-i-ta].
Pekka.NOM deny-PST.3SG Merja-PAR wash-MA.ELA/INF window-PL-PAR
'Pekka denied Merja from washing the windows.'

e. Pekka aloitt-i siivoukse-n [pese-**mä**llä ikkuna-t].
Pekka.NOM begin-PST.3SG cleaning-ACC wash-MA.ADE/INF window-PL.ACC
'Pekka began the cleaning by washing the windows.'

The class also contains constructions called the ‘E-infinitive’ and ‘A-infinitive’ in the traditional literature, both which have two forms. This 2 + 2 classification makes less sense syntactically, as we will see later, so it was expanded into the four infinitives listed in (2).

(2) a. *A-infinitive*

Pekka halus-i [lähte-ä].

Pekka.NOM want-PST.3SG leave-A/INF

‘Pekka wanted to leave.’

b. *KSE-infinitiv*e

Pekka pakkas-i [lähte-ä-kse-en].

Pekka.NOM pack-PST.3SG leave-A/INF-KSE/INF-PX/3P

‘Pekka packed in order to leave.’

c. *E-infinitive*

Pekka pakkas-i [laula-en].

Pekka.NOM pack-PST.3SG sing-E/INF

‘Pekka packed by/while singing.’

d. *ESSA-infinitive*

Pekka laulo-i [lähti-essä(-än)].

Pekka.NOM sing-PST.3SG leave-ESSA/INF(-PX/3P)

‘Pekka sang while leaving.’

Next we consider the VA-INFINITIVE which constitutes a nonfinite complement clause expressing propositional meaning that can be best translated into English by a regular finite clause.[3] It has both the present and past tense forms (3).

(3) a. Pekka usko-i Merja-n pese-vän ikkuna-t.

Pekka.NOM believe-PST.3SG Merja-GEN wash-PRS.VA/INF window-PL.ACC

‘Pekka believed that Merja will wash the windows.’

b. Pekka usko-i Merja-n pes-seen ikkuna-t.

Pekka.NOM believe-PST.3SG Merja-GEN wash-PST.VA/INF window-PL.ACC

‘Pekka believed that Merja washed the windows.’

Finally, the TUA-INFINITIVE describes past actions or underlying causes and rationalizations (4).

(4) *TUA-infinitive*

Pekka laulo-i [lähde-tty-än].

Pekka.NOM sing-PST.3SG leave-TUA/INF-PX/3P

‘Pekka sang after he left.’

The nonfinite clauses differ from the finite clause in a number of respects. They do not exhibit finite agreement,[4] nominative case assignment, mood or modality, and cannot host high complementizers such as *että* ‘that’. The sentential negation *e-* ‘not’, which is a tenseless auxiliary-type element in Finnish, cannot appear inside nonfinite clauses. Finally, the nonfinite clauses do not provide a domain for an operator, thus there are no such things as nonfinite relative clauses headed by a relative pronoun.[5] Still, they all describe an event with participants, assigns subject and object cases, incorporates additional clauses, and hosts adverbs.

The constructions enumerated above and targeted for a detailed analysis in this study are summarized in Table 1.

Table 1. Finnish nonfinite clauses (infinitives and participles) selected for an analysis.

Name	Example	Meaning (approx.)

Infinitives	A-infinitive	*syö-dä* 'eat-A/INF'	'to eat'
	VA-infinitive	*syö-vän* 'eat-VA/INF'	'to eat'
	MA-infinitives	*syö-mässä* 'eat-MA.INE/INF' *syö-mällä* 'eat-MA.ADE/INF' *suö-mättä* 'eat-MA.ABE/INF', *syö-mästä* 'eat-MA.ELA/INF' *syö-mään* 'eat-MA.ILL/INF'	'in eating' 'by eating' 'without eating' 'from eating' 'to eat'
	E-infinitive	*syö-den* 'eat-E/INF'	'by eating'
	ESSA-infinitive	*syö-dessä* 'eat-ESSA/INF'	'while eating'
	TUA-infinitive	*syö-tyä* 'eat-TUA/INF'	'after eating'
	KSE-infinitive	*syödä-kse-en* 'eat-KSE/INF-PX/3P'	'in order to eat'
Participles	VA-participle	*syö-vä* 'eat-VA/A'	'x who eats'
	MA-participle	*syö-mä* 'EAT-MA/A'	'x who was eaten'

Much of the previous literature on these constructions have focused on whether and how the overt morphological forms of the infinitival predicates in Table 1 match with their underlying syntactic structure. Notice that the *-mA* morph of the MA-participle *syö-**mä*** 'eat-MA' also occurs inside the MA-infinitives (*syö-**mä**-än* 'EAT-MA-ILL'), the latter containing an additional illative semantic case form. The MA-infinitives could therefore be classified as 'semantically case-marked MA-participles'. The KSE-infinitive could be analyzed as a translative-marked A-infinitival with structure 'V + A/inf + translative case', and the VA-infinitive as a VA-participle (*-vA*) case-marked by the genitive or accusative (*-n*). An alternative is that some or all of these morphological similarities reflect diachronic development or perhaps pure coincidences. Thus, Vilkuna (2000) suggests that the *-vAn* morph (see Table 1) is related neither to the VA-participle nor to the genitive/accusative case form *-n*, instead modern Finnish speakers perceive it as an "unanalysable whole" functioning as a sign for the VA-infinitive (p. 244)(see also Ylikoski 2003:203–205). We return to this controversy.

The first modern syntactic approach to the Finnish nonfinite clauses was presented by Vainikka (1989), who analyzed them as verb phrases, noun phrases and preposition phrases.[6] Specifically, Vainikka assumed that the nonfinite clauses are either bare verb phrases, where the infinitival suffix corresponds to a feature inside the verb phrase (A-infinitive), verb phrases wrapped inside preposition phrases (MA-infinitives, containing semantic case suffixes), or (iii) verb phrases wrapped inside noun phrases (VA-, KSE-, TUA-infinitives). We can regard this analysis as a null hypothesis of sorts in the sense that it tries to survive without positing anything beyond the standard lexical categories. The problem, though, is that Finnish nonfinite clauses have neither the distribution nor the properties of verb phrases, noun phrases or preposition phrases. As pointed out above, they are often considered to exhibit a mixture of nominal and verbal properties, resisting clear-cut classification. Vainikka considered the issue but did not offer a solution. In later work, she expanded the functional structure by assuming that the verb phrase was embedded inside a nominalizing infinitival head Y which functions to transform (Vainikka's term) the verb phrase into a nominal projection, allowing the resulting construction to be embedded inside nominal projections hosting case and agreement (Vainikka 1994). The syntactic structure mirrors morphological form according to the scheme 'verb + nominalizer Y + case form + nonfinite (possessive suffix) agreement', with the latter two optional. The problem of differentiating between infinitives and standard verb phrases, noun phrases and preposition phrases remained unsolved, however.

Koskinen (1998), to my knowledge the first large work devoted in its entirety to the syntactic analysis of Finnish nonfinite clauses, analyzed the infinitival predicates as hybrid categories instead of the major supercategories V, N or P. The model created new lexical categories by mixing features. For example, she proposed that the VA-infinitive clause (3) is a verb phrase embedded inside a hybrid tense/adjectival head, where adjectives were further analyzed as a combination of N and V. The tensed adjective was wrapped inside a further 'DP-like projection' (p. 169) giving the nonfinite clause its nominal properties. Crucially, the hybrid approach can differentiate the nonfinite predicates (and hence also the nonfinite clauses) from ordinary verbs, nouns and adpositions by modifying their feature content. Her analysis is similar to Vainikka's in that the morphological forms guided syntactic analysis.[7]

Ylinärä (2018) developed another hybrid analysis where the nonfinite clauses were analyzed as projecting both verbal and nominal categories, but the analysis was developed within the more recent cartographic framework. The analysis first combines category-neutral roots √ with an aspect head, creating an aspectual template $[_{\text{AspP}}\ \text{Asp}^0\ [_{\surd\text{P}}\ \surd^0]]$ that serves as the basic structure common to all infinitives targeted for an analysis in her study. Several additional projections were then required to derive full nonfinite clauses, such as projections hosting the (possibly null) subject and object arguments (AgrS^0, AgrO^0), projections contributing nominal properties such as case (K^0, D^0) and, finally, projections related to information structure (σ^0, γ^0) accounting for clause-internal topic/focus readings and construction-internal argument scrambling.

We approach the data first from a slightly different perspective and return to the analyses reviewed above later. Specifically, we begin from the constructions listed in Table 1 without attempting to decompose or reduce them on the basis of their morphological surface forms and instead create a computational model which calculates their syntactic and semantic properties thematized in the next section. Once we have a successful model we will treat it as an observationally adequate baseline with which the other approaches are compared. When compared with the previous approaches reviewed above, the resulting analysis is slightly more complex than Vainikka's null hypothesis but also less complex than the models proposed by Koskinen (1998) and Ylinärä (2018).

3 RESEARCH AGENDA

We focus on five syntactic traits of the Finnish nonfinite clauses: selection, control, grammatical role, agreement and the properties of subjects. To illustrate, let us consider more closely the properties of the Finnish VA-infinitive (3), shown again in (5).

(5) Pekka ties-i [osta-va-nsa uude-t kengä-t].
Pekka.NOM know-PST.3SG buy-VA/INF-PX/3P new-PL.ACC shoe-PL.ACC
'Pekka knew that he (Pekka/*third party) will buy new shoes.'

The sentence contains a main clause segment 'Pekka knew…' plus a nonfinite segment '…to buy new shoes', which together express a propositional attitude with the content in which Pekka knew what is stated in the nonfinite segment. The nonfinite segment in (5) can only be selected by certain kinds of verbs. While it is possible to know the proposition described by the VA-infinitive, it is not possible to order it (6a). Furthermore, the VA-infinitive cannot occur in connection with intransitive verbs (6b).

(6) a. *Pekka käsk-i [osta-va-nsa uude-t kengä-t].
Pekka.NOM order-PST.3SG buy-VA/INF-PX/3P new-PL.ACC shoe-PL.ACC
Intended: 'Pekka ordered him/third party to buy new shoes.'
b. *Pekka nukaht-i [osta-va-nsa uude-t kengä-t].
Pekka.NOM fall.asleep-PST.3SG buy-VA/INF-PX/3P new-PL.ACC shoe.PL-ACC

Something makes the transitive verb 'order' and all intransitives incompatible with the VA-infinitive. Furthermore, nonfinite clauses do not generally occur out of the blue and must occur in a selected position (7).[8]

(7) *Osta-va-nsa uude-t kengä-t!
buy-VA/INF-PX/3P new-PL.ACC shoe-PL.ACC

We include selection, illustrated by the example above, into this study as a phenomenon we want to model. How selection (and other notions introduced in this section) was operationalized into concrete stimulus materials will be discussed in Section 5.1.1.

The thematic agent of the infinitival predicate ‘to buy’ in (5) must be the same as the main clause subject. In (5), Pekka knows and buys something. Moreover, it is not possible to insert an embedded subject inside sentence (5), as shown by (8).

(8) *Pekka ties-i [hän-en osta-va-nsa uude-t kengä-t].

Pekka.NOM know-PST.3SG he-GEN buy-VA/INF-PX/3P new-ACC.PL shoes-ACC.PL

Intended: ‘Pekka believed that he will buy new shoes.’

Properties of this type are referred to as CONTROL. Some infinitival sentences, such (1), exhibit subject control, where the thematic agent of the embedded infinitival clause must be the same as the subject of the superordinate clause. Other infinitives exhibit object control, where the thematic agent of the embedded clause must be the same as the direct object of the superordinate clause (9)(symbol PRO stands for the embedded thematic agent when it is not expressed overtly).

(9) Pekka$_1$ komens-i hän-tä$_2$ [PRO$_{*1,2}$ lähte-mään].

Pekka.NOM order-PST.3SG he-PAR leave-MA.ILL/INF

‘Pekka ordered him to leave.’

The participant who was asked to leave was the patient of asking. If we replace the MA-infinitive *lähte-mään* ‘leave-MA.ILL/INF’ with *huuta-malla* ‘yell-MA.ADE/INF’ meaning ‘by yelling’, control shifts back to the subject (10).

(10) Pekka$_1$ komens-i hän-tä$_2$[PRO$_{1, *2}$ huuta-malla].

Pekka.NOM *order*-PST.3SG *he*-PAR *yell*-MA.ALL/INF

‘Pekka ordered him (to so something) by yelling.’

Control (subject control and object control) was included into this study as a phenomenon targeted for modeling.

Some nonfinite clauses are complements, some are nonselected adjuncts (i.e., adverbials), while others exhibit mixed behavior.[9] To show that the VA-infinitive is unable to appear in a nonselected adjunct position, we can try to combine it with a full transitive finite clause (11).

(11) *Pekka usko-i Merja-a [osta-va-nsa uude-t suka-t].

Pekka.NOM *believe*-PST.3SG *Merja*-PAR *buy*-VA/INF-PX/3P *new*-ACC.PL *sock*-ACC.PL

Intended: ‘Pekka believed Merja that s/he should buy new socks.’

The direct object (*Merja-a* ‘Merja-PAR’) reserves the direct object slot and excludes the VA-infinitive from the same grammatical role, which is the only syntactic role the VA-infinitive can have. Some infinitives do, however, appear in adjunct positions (12).

(12) Pekka usko-i Merja-a [tarkista-matta asia-a neti-stä].

Pekka.NOM *believe*-PST.3SG *Merja*-PAR *check*-MA.ABE/INF *thing*-PAR *Internet*-ELA

‘Pekka believed Merja without checking it/the matter from the Internet.’

This shows that the MA-infinitive, unlike the VA-infinitive, can behave like an adverbial. SYNTACTIC STATUS (complement, adjunct) was included into this study as a phenomenon targeted for modelling.

The infinitival form *osta-va-nsa* ‘buy-VA/INF-PX/3P’ contains three overt morphological elements: the verb stem *osta-* ‘buy’, the infinitival affix *-vA(n)-*, and the third person nonfinite agreement (possessive) suffix -*nsA*. Example (13) shows that the third suffix -*nsA* represents agreement.

(13) a. Minä usko-n osta-va-**ni** uude-t kengä-t.
I.NOM believe-PRS.1SG buy-VA/INF-PX/1SG new-PL.ACC sock-PL.ACC
'I believe that I will buy new shoes.'

b. Sinä usko-t osta-va-**si** uude-t kengä-t.
you.NOM believe-PRS.2SG buy-VA/INF-PX/2SG new-PL.ACC sock-PL.ACC
'You believe that you will buy new shoes.'

Nonfinite agreement was also targeted for modelling.

Only some infinitival predicates exhibit agreement, and there are cases where agreement is optional. In the case of the VA-infinitive, agreement is optional but affects other properties of the construction. Example (14) shows that when the nonfinite agreement disappears, a separate overt subject must appear inside the infinitive. This, furthermore, breaks subject control.

(14) Pekka usko-i [*(Merja-n) osta-van uude-t kengä-t].
Pekka.NOM believe-PST.3SG Merja-GEN buy-VA/INF new-ACC.PL shoe-ACC.PL
'Pekka believed that Merja would buy new shoes.'

Examples (8) and (14) show that the embedded subject can be both obligatory and impossible. Example (15) demonstrates the same effect for the A-infinitive.

(15) a. *Embedded subject is obligatory*
Pekka käsk-i *(Merja-n) osta-a uude-t kengä-t.
Pekka.NOM order-PST.3SG Merja.GEN buy-A/INF new-PL.ACC shoe-PL.ACC
'Pekka ordered Merja to buy new shoes.'

b. *Embedded subject is ungrammatical*

Pekka	halus-i	(*Merja-n)	osta-a	uude-t	kengä-t.
Pekka.NOM	*want-PST.3SG*	*Merja-GEN*	*buy-A/INF*	*new-PL.ACC*	*sock-PL.ACC*

'Pekka wanted (Merja) to buy new shoes.'

These data show that when the A-infinitive is selected by 'order/ask', the embedded subject is obligatory (15a), while selection by 'want' blocks it (15b). The presence/absence of the embedded subject was added to this study as a phenomenon to be modelled. We want to model also the interaction between subjects and agreement.

Selection (5)–(7), subject and object control (8)–(10), syntactic status (11)–(12), nonfinite agreement (13) and the syntax of overt subjects (14)–(15) define the properties we focus on in this study and target for a rigorous analysis. We pay less attention to morphology, adjective participles, binding, lexical semantics and word order. Some of the infinitival predicates have passive forms, but this matter was put aside since many nontrivial questions that have to do with derivational morphology were excluded. Full nominalizations belong to nominal syntax and were likewise excluded.

4 AN INFORMATION PROCESSING ANALYSIS

We develop an analysis of Finnish nonfinite clauses that is based on a computational, Python-based information processing model of the human brain. The approach combines ideas from linguistics, naturalistic cognitive science and computational linguistics. Section 4.1 introduces the hypothesis, Section 4.2 describes the implementation.[10]

4.1 Hypothesis

We begin by outlining some more general assumptions concerning linguistic processing that form the immediate theoretical context of the analysis. We assume that human linguistic information processing consumes phonological words from the sensory input and transforms them into lexical representations. These operations correspond to a process where the hearer recognizes and retrieves words arriving through the sensory systems. The LEXICO-MORPHOLOGICAL SYSTEM, which performs these computations, is also required to handle polymorphemic words such as the infinitival predicate *osta-va-nsa* 'buy-VA/INF-PX/3P'. The lexico-morphological system has access to the LEXICON, a storage of lexical information.

Let us assume that the lexico-morphological system delivers its output to the SYNTACTIC SYSTEM calculating hierarchical dependencies between the incoming lexical items. For example, in order to represent the differences between complements and adjuncts the model must have access to the corresponding syntactic notions. They, like other similar notions such as selection and agreement introduced in Section 3, are defined and computed inside the syntactic system.

Once the syntactic computations have been completed, the output is interpreted semantically. Since semantic interpretation in the broad sense includes phenomena that do not belong to linguistics or language processing specifically (e.g., emotions, holistic perception, music appreciation), it is useful to posit an INTERFACE between syntax and broad semantics as the last linguistic representation generated by the syntactic processing pathway before nonlinguistic processing takes over. We can imagine it as a circuit that connects the endpoint of the syntactic processing pathway to the semantic system(s). The overall architecture is depicted in Figure 1.

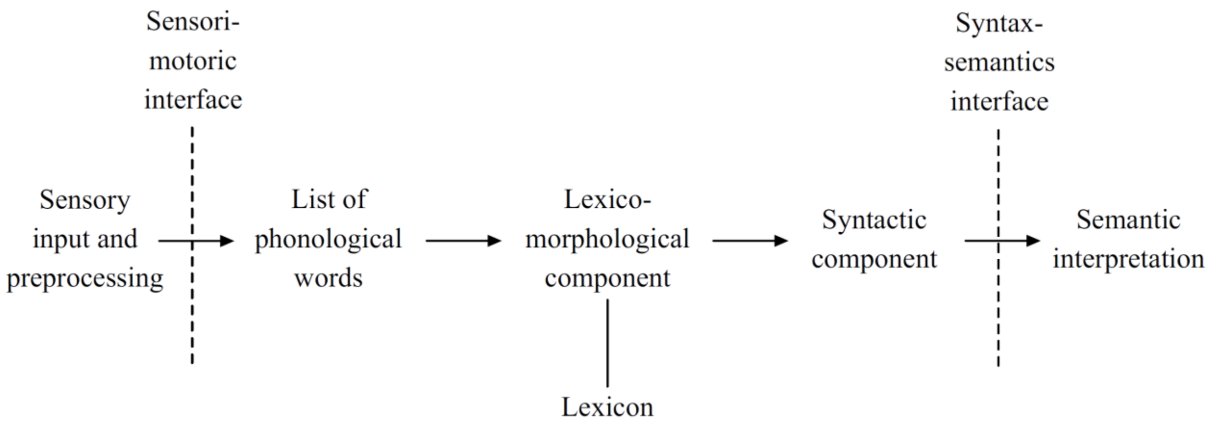


Figure 1. Overall architecture of the language processing system used as a background in this study. The model is a simplification, but sufficient in the light of the research agenda defined in Section 2.

Assuming the architecture in Figure 1 as a background, let us consider the Finnish nonfinite clauses. If the infinitival predicate consists of two overt morphemes, then those items will be used as seeds for generating the corresponding lexical items into the syntactic structure. The idea is illustrated in (16).

(16)	Pekka	uskoi	Merja-n	osta-van	uude-t	kengä-t.
	Pekka.NOM	*believe*-PST3SG	*Merja*-*GEN*	*buy*-*VA/INF*	*new*-*ACC.PL*	*shoe*-*ACC.PL*
				[VA/inf [buy	[new shoes]]]]]	

The first line represents the original sentence with the morphological boundaries marked as analysed by the lexico-morphological system (i.e. *osta-van* ‘buy-VA/INF’). The last line sketches the intended syntactic representation where the two components of the infinitival predicate, in this case the verbal stem and the VA-suffix, have been assembled into the syntactic structure as independent lexical items. More generally, an infinitival predicate composed out of a verb stem V and an infinitival suffix X (e.g., *osta-van* ‘buy-VA/INF’ = V#X) will be transformed by the lexico-morphological component into a sequence X^0 + V^0

where X^0 and V^0 are lexical items. Examples (17) and (18) show how an input sentence containing an A-infinitive is processed, according to these assumptions.

(17)	Pekka	halus-i	[osta-a	kengä-t].	(Input)
	Pekka.NOM	*want-PST.3SG*	*buy-A/INF*	*shoe-PL.ACC*	(Morphology)
	$D^0 + N^0$	$T^0_{pst} + V^0$	A/inf^0 + V^0	$D^0 + N^0$	(Lexical items)

(18) $[_{TP}$ DP $[_{TP}$ T^0 $[_{VP}$ V^0 $[_{A/infP}$ A/inf^0 $[_{VP}$ V^0 DP]]]] (Syntax)

The model therefore assumes a variation of the syntax-morphology mirror principle (Baker 1985, Julien 2002) which maps morphological decompositions transparently into syntactic structure. Once we have repackaged morphology into syntactic structure, selection can be modelled by relying on standard head-complement selection. Example (19) shows how selection is applied to (17). The main verb establishes a grammatical selection dependency with the head of the infinitival clause to generate the interpretation 'want + to V'.

(19) $[_{TP}$ DP $[_{TP}$ T^0 $[_{VP}$ $V_1{}^0$$[_{A/infP}$ A/inf^0 $[_{VP}$ $V_2{}^0$DP]]]]

Notice that the selection dependency will be established between the main verb and the infinitival component of the infinitival predicate; there is no relationship between the main verb V_1 and the lower verb V_2.

Control will be captured by assuming that every predicate must be paired with an argument. If no local argument can be found, a nonlocal argument is detected by scanning upwards. The model will create the control dependency shown in (20). The direct object and anything below the verb are ignored because only upward/leftward scanning is possible.

(20) Pekka uskoi osta-va-nsa uudet sukat.

[*Pekka*[T [*believe* [*VA/inf* [*buy* [*new* *socks*]]]]]

⟵ (‘Who is the buyer? Pekka’)

This accounts for the intuition that Pekka is both the believer and the buyer in the thought expressed by (20). If an embedded subject intervenes, it will be the target (21).

(21) Pekka uskoi Merja-n osta-van uudet sukat.

[*Pekka*[*believed* [*Merja-GEN* [*VA/inf* [*buy* [*new* *socks*]]]]]]

⟵ (‘Who is the buyer? Merja’)

We assume that this process takes place at the syntax-semantics interface (see Figure 1).

Some Finnish infinitival predicates exhibit nonfinite agreement. We assume that agreement is reconstructed as features inside lexical items. The third-person singular agreement suffix in (20) is transformed into a lexical feature cluster [3SG] inserted inside the lexical item corresponding to the VA-morpheme in the sensory input. Agreement features of finite verbs are inserted inside finite T. The result is shown in (22).

(22) Pekka usko-i osta-va-nsa uude-t kengä-t.

Pekka.NOM *believe-PST.3SG* *buy-VA/INF-PX/3P* *new-PL.ACC* *shoes-PL.ACC*

[Pekka [$T_{[3sg]}$ [believe [VA/$inf_{[3p]}$ [buy [new shoes]]]]]]

$T_{[3sg]}$ signifies that T (tense, here ‘past’) contains features corresponding to ‘third person singular’, VA/$inf_{[3p]}$ means that the VA-infinitival head contains the feature ‘third person’. Had we assumed that the third person agreement cluster corresponds to its own head in the lexicon and not to inflectional features, the model would have projected a separate Agr^0 head positioned above finite tense by the mirror principle (i.e., V#T#3sg ~ [...Agr^0...[...T^0...[...V^0...]]]). This alternative, which generates a further finite agreement head above TP, is not linguistically implausible (Holmberg *et al.* 1993, Holmberg & Nikanne 2002, Mitchell 1991, Pollock 1989) and will be experimented with in Section 6.

We will also have to capture the fact that nonfinite agreement can be absent, optional or obligatory (see, e.g., examples (13)–(14)). Finnish nonfinite agreement was explored recently by Brattico (2023) in a computational study. Accordingly, a head that never shows agreement has feature –ΦPF, signifying that overt agreement is not possible (Φ refers to phi-feature sets, PF to the PF-interface responsible for spellout, so –ΦPF means 'do not spell out phi-features'). Sentences such as **Pekka halusi osta-a-nsa sukkia* 'Pekka wanted buy-A/INF-PX/3P socks' can be ruled out by using this feature to block the agreement features from going inside the A-infinitival head. Feature +ΦPF makes overt agreement obligatory. Some Finnish nonfinite clauses can be described by a generalization which says that agreement (when possible in the first place) occurs if and only if an overt phrasal subject is absent (23).

(23) a. *Embedded subject is ungrammatical in the presence of agreement*

Pekka usko-i (*Merja-n) osta-va-nsa kengä-t.

Pekka.NOM believe-PST.3SG (Merja-GEN) buy-VA/INF-PX/3P shoe-PL.ACC

'Pekka believed that he/*Merja will buy shoes.'

b. *Embedded subject is obligatory in the absence of agreement*

Pekka usko-i *(Merja-n) osta-van kengä-t.

Pekka.NOM believe-PST.3SG (Merja-GEN) buy-VA/INF shoe-PL.ACC

'Pekka believed that *he/Merja Merja will buy socks.'

Feature Φ1, which was posited to handle this situation, requires that either an overt phrasal subject or overt agreement must occur but not both redundantly. Lexical elements which allow redundant co-occurrence have feature Φ2, a profile that characterizes several nonfinite predicates as well as finite verbs, nominals and adpositions in Finnish. These assumptions do not yet capture cases in which an overt phrasal subject is mandatory regardless of what happens to agreement. The former has commonly been captured by positing an EPP feature (Chomsky 1981, 1982) requiring that the head has an overt phrasal specifier. To this author's knowledge the idea of extending the EPP mechanism to the analysis of Finnish nonfinite clauses was first proposed by Vainikka (1989). Feature –EPP prohibits the head from projecting an overt phrasal specifier. The features are summarized in Table 2.

Table 2. Lexical features posited in this study (α^0 = grammatical head)

Lexical feature	**Explanation**
+EPP	α^0 requires an overt phrasal specifier
–EPP	α^0 cannot have an overt phrasal specifier
+ΦPF	Overt agreement at α^0 is mandatory
–ΦPF	Overt agreement at α^0 is illicit
Φ1	α^0 must exhibit overt agreement or phrasal subject but not both
Φ2	α^0 must exhibit overt agreement, a phrasal subject, or both
X/inf^0	Infinitival heads where X refers to the type of infinitival suffix (Table 1).

4.2 Algorithm

We will build the model on the existing Python-based linear phase algorithm proposed in (Brattico 2019, 2022), which is a linguistic information processing platform (essentially, a collection of Python functions) based on the architecture provided in Figure 1. Understanding the exact operation of the underlying implementation is not necessary for interpreting the linguistic results. Some of the technical material is in the supplementary; this section provides a nontechnical summary.

The model will be processing Finnish. Speakers of Finnish, English and Italian are not identical, however. The program creates an idealized SPEAKER MODEL at runtime for the speaker of any language, dialect or variation L present in the lexicon and uses the model to process all inputs in L that it determines on the basis of each sentence. The speaker model used in the present study is illustrated in Figure 2.

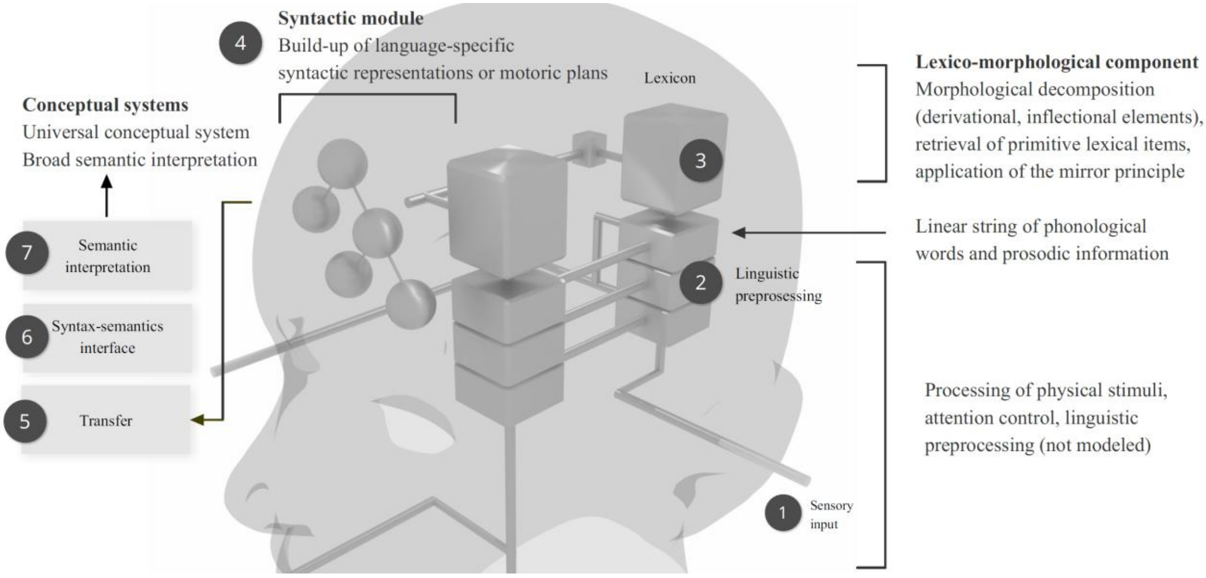


Figure 2. A speaker model for some language, dialect or variation. The model is selected automatically on the basis of the language of the input sentence.

It is assumed that all language-specific properties are encoded into the lexicon, while the syntactic processing pathway (components 4–7, Figure 2) is universal. Thus, Finnish sentences are processed by the same, ultimately neuronal syntactic pathway that processes also other languages, but the two languages use different lexicons.

The lexicon (component 3, Figure 2) is a list of lexical entries. Polymorphemic words are mapped into morphological decompositions, which are linear lists of pointers to further entries in the same lexicon; primitive lexical items are either heads, inflectional features or clitics, all which are sets of features wrapped inside primitive constituents. The features posited in Section 4.1 and summarized in Table 2 are among the features a primitive constituent can have. A constituent is an object that the syntactic component can access and process. Primitive constituents are inserted into the phrase structure as such (they become heads in the sense of the generative theory), inflectional features are inserted inside heads (clitics are not present in the current dataset). These assumptions are illustrated in Figure 3.

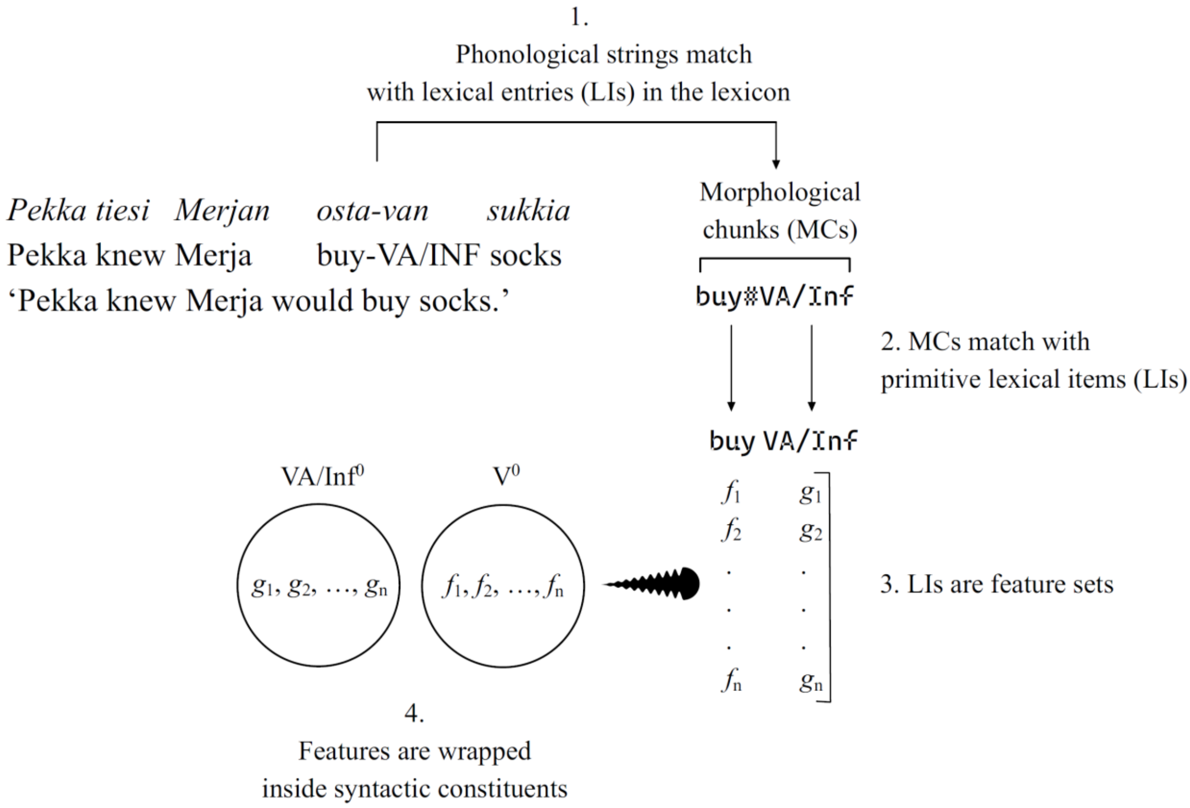


Figure 3. A phonological word enters the processing pipeline and activates lexical entries, here the two elements 'buy' and the VA-infinitival suffix VA/Inf (1). These items are matched with further entries in the same lexicon representing primitive lexical items (2). Primitive lexical items are feature sets (3). Features are wrapped inside constituents (4), objects that the syntactic component assembles into syntactic representations (4, Figure 2).

The syntactic module (component 4, Figure 2) receives primitive constituents in a left-to-right order from the sensory input, following the original idea by Phillips (1996), and attaches them into the phrase structure in the active syntactic working memory. The incoming heads are attached to the phrase structure incrementally. All incoming heads are attached to the right edge of the existing phrase structure, thus either to the top node or to some of its right daughters. Complex constituents have the form [A B] where A is the left constituent, B is the right constituent. A and B can be primitive or complex. Active syntactic working memory holds all linguistic objects under processing. The attachment process is illustrated in Figure 4.

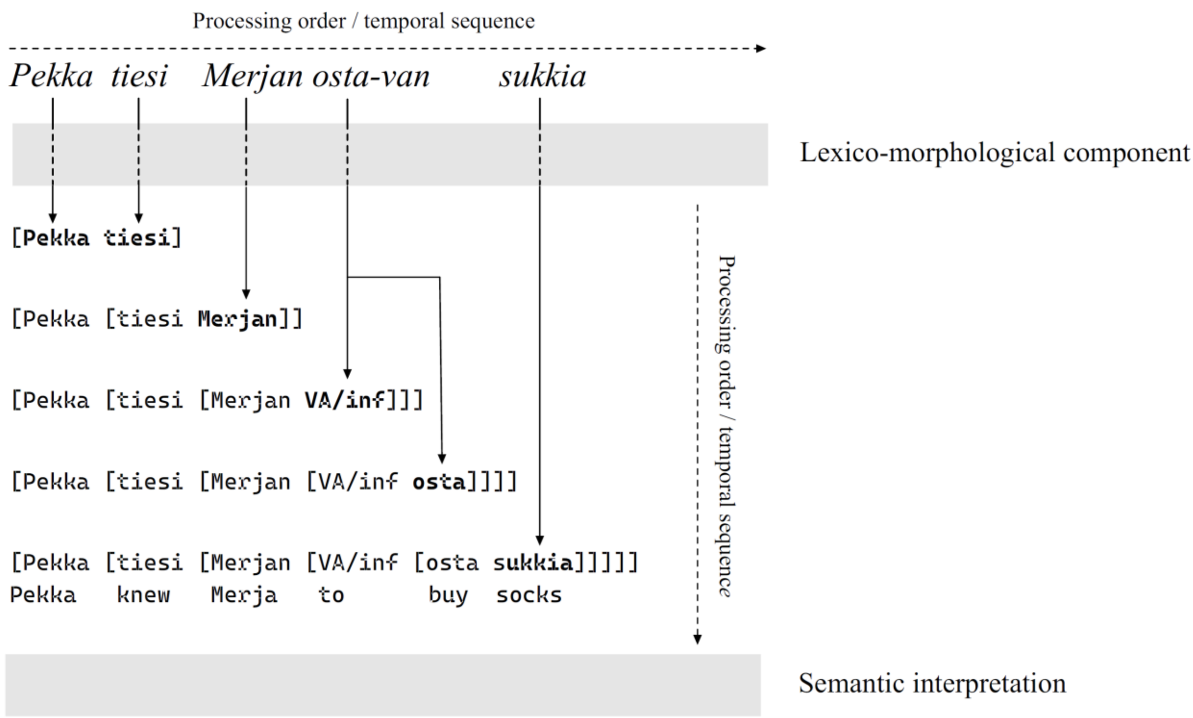


Figure 4. The input sentence is read from left to right. Each phonological word is decomposed into primitive lexical items, which are attached incrementally to the phrase structure in the active syntactic working memory. Some decompositions (e.g., *tiesi* ‘knew’ ~ $T^0 + V^0$) were ignored for readability and are examined in detail in Section 5.2.

This description presupposes that the system can determine the correct or intended right edge position for all incoming lexical items. Consider the parsing of *the horse raced past the barn.* The most reasonable way of attaching the incoming words incrementally into the phrase structure given this input would be [[*the horse*] [*raced* [*past the barn*]]], but this strategy fails if another word such as *fell* appears. To solve this issue, the algorithm considers all possible right edge nodes during each attachment operation, filters solutions which it deems too ill-formed and ranks the residuum by using cognitive parsing heuristics. The filtered and ranked nodes are explored recursively to create a parsing tree. Thus, after the algorithm arrives at the first possible solution (called the 'the first-pass parse') after consuming all words in the input, it will backtrack and search for alternative solutions by using the parsing tree created by filtering and ranking. In the case of [[*the horse*] [*raced* [*past the barn*]]]] + *fell*, the model backtracks until it finds [[*the horse* [*raced past the barn*]] *fell*] with the meaning 'the horse which was raced past the barn fell'. Backtracking causes an increase in the use of cognitive resources that we should ideally detect in psycholinguistic experiments. All solutions generated by this method are evaluated at the syntax-semantics interface (component 6, Figure 2). Those which pass are forwarded to semantic interpretation for further processing (components 7, 8, Figure 2).

Once a candidate solution is generated and considered worthy of testing at the syntax-semantics interface, it is TRANSFERRED to it (component 5, Figure 2). Transfer applies a limited amount of error correction or normalization to the parsed solution. It has a limited role in the present study because chain creation was not specifically selected for an analysis in this study. It does play a role when the input sentence has uncanonical and/or unexpected properties that need adjusting before universal semantic interpretation and conceptual processing can apply. From the point of view of the generative theory, transfer corresponds to a reverse-engineered transformational component creating head-, Ā-, A- and scrambling chains. See the supplementary document §2.4 for details.

Once the whole system has been set up, the algorithm will read all input sentences and provide them with syntactic analyses and semantic interpretations, all according to the design principles provided in this and the previous subsection. The model is therefore evaluated against a whole dataset of expressions in a computational experiment, reported in the next section.

5 COMPUTATIONAL EXPERIMENT

We tested the model by letting the algorithm implementing the analysis process Finnish nonfinite clauses. We say that the model is justified to the extent that the behavior of the model matches with the behavior of native speakers. Section 5.1 describes the experiment and contains subsections elucidating the design together with the construction of the dataset (Section 5.1.1) and the simulation procedure (Section 5.1.2); Section 5.2 reports the results in separate subsections discussing the A- and VA-infinitives (Section 5.2.2), MA- and E-infinitives (Section 5.2.3), ESSA-, KSE- and TUA-infinitives (Section 5.2.4) and finally the two participles (Section 5.2.5).

5.1 Methods

5.1.1 Design and stimuli

A set of input sentences containing Finnish nonfinite clauses was created. The test corpus was created by crossing the syntactic variables defined in Section 3. The first variable was NONFINITE CLAUSE TYPE, which contains the five MA-infinitives plus six other types (Table 1), repeated in (24). Participles were not included in this study as an independent variable, but see the discussion in Section 6.2.5.

(24) a. *A-infinitive*

Pekka halus-i osta-a sukk-i-a.

Pekka.NOM want-PST.3SG buy-A/INF sock-PL-PAR

‘Pekka wanted to buy socks.’

b. *VA-infinitive*

Pekka usko-i hän-en osta-van sukk-i-a.

Pekka.NOM believe-PST.3SG he-GEN buy-VA/INF sock-PL-PAR

‘Pekka believed that he will buy socks.’

c. *MA- infinitives (five forms, one shown here)*

Pekka näk-i hän-et osta-massa sukk-i-a.

Pekka.NOM see-PST.3SG he.ACC buy-MA.INE/INF sock-PL-PAR

‘Pekka saw him buying socks.’

d. *E-infinitive*

Pekka kulutt-i tunnin luki-en.

Pekka.NOM spend-PST.3SG hour read-E/INF

‘Pekka spend one hour by reading.’

e. *ESSA-infinitive*

Pekka väsäht-i juost-essa.

Pekka.NOM get.tired-PST.3SG run-ESSA/INF

‘Pekka got tired while running.’

f. *TUA-infinitive*

Pekka väsäht-i juos-tua-an.

Pekka.NOM get.tired-PST.3SG run-TUA/INF-PX/3P

‘Pekka got tired after running.’

g. *KSE-infinitive*

Pekka nukku-i tunnin levätä-kse-en.

Pekka.NOM sleep-PST.3SG hour rest-KSE/INF-PX/3P

‘Pekka slept one hour in order to rest.’

The next variable was the SELECTING VERB. Four different verbs were used in order to model selection and the complement/adjunct distinction (OC = obligatory control)(25).

(25) a. *Anti-OC verbs, which require an embedded subject*

Pekka käsk-i *(hän-en) lähte-ä.

Pekka.NOM order-PST.3SG (he-GEN) leave-A/INF

‘Pekka ordered him to leave.’

b. *OC verbs, not compatible with an embedded subject*

Pekka halus-i (*hän-en) lähte-ä.

Pekka.NOM want-PST.3SG (he-GEN) leave-A/INF

‘Pekka wanted to leave.’

c. *Verbs which select propositions and proposition-like objects*

Pekka usko-i hän-en lähte-vän/ *lähte-ä.

Pekka.NOM believe-PST.3SG he-GEN leave-VA/INF leave-A/INF

‘Pekka believed that he will/would leave.’

d. *Intransitive verbs*

Pekka nukku-i.

Pekka.NOM sleep-PST.3SG

‘Pekka slept.’

The third variable was the absence/presence of the PHRASAL SUBJECT, while the fourth was AGREEMENT, which could also be absent or present. When nonfinite agreement was present, the agreement suffix was added to the infinitival stem as the outmost element and was represented in the lexicon as an inflectional affix. The syntactic position of the infinitival was the fifth variable. It had two options, complement and adjunct. Complement infinitival constructions were created by putting the infinitive right after the main verb, which licenses the main verb to select the infinitive whenever possible (26a). An adjunct infinitival construction was created by positioning a separate direct object between the transitive main verb and the infinitival to block the head-complement interpretation (26b–c).

(26) a. *Complement configuration*

Pekka halus-i [osta-a suka-t].

Pekka.NOM want-PST.3SG buy-A/INF sock-PL.ACC

‘Pekka wanted to buy socks.’

b. *Adjunct configuration, ungrammatical and grammatical*

*Pekka halus-i auto-n [osta-a suka-t].

Pekka.NOM want-PST.3SG car-ACC buy-A/INF sock-PL.ACC

c. Pekka halus-i auto-n [voida-kse-en matkustel-la].

Pekka.NOM want-PST.3SG car-ACC able-KSE/INF-PX/3P travel-A/INF

‘Pekka wanted a car in order to travel.’

Because the complement configurations (26a) are already covered by the main verb tests, and because the adjunct tests (26b–c) involve transitive verbs not used in any other condition, the adjunct configurations were added to the list of main verbs as a fifth level. These assumptions generated 11 (construction type) x 5 (selecting main verb + adjunct configuration) x 2 (embedded subject) x 2 (nonfinite agreement) = 220 core constructions which capture the notion of logically possible infinitival sentence given the research agenda and the independent syntactic variables it defines.

A few special tests were added to the dataset. First, before running the actual test sentences we want to make sure that the lexical elements and other presupposed grammatical mechanisms such as verb valency, case marking and word order work correctly. Ten (10) baseline test sentences were added for this purpose (Group 0 in the dataset). If any of these sentences were calculated wrongly, examination of further test results was deemed meaningless. Then, when the appearance of some infinitival type required the presence of a selecting lexical item from a special semantic class, the required test sentence was included as a single datapoint. For example, the Finnish MA.ELA-infinitive (roughly 'from doing') occurs with the main verb *estää* 'prevent' but not with *nähdä* 'see' (27).

(27) Pekka *näki/esti hän-tä lähte-mästä.
Pekka saw/prevented he-PAR leave-MA.ELA/INF
'Pekka saw/prevented him from leaving.'

Both sentences were added to make sure that the contrast works correctly, but only to one test group. Therefore, lexical semantic selection was tested in the experiment but was not explored systematically as an independent or dependent variable. In some cases where the selecting and selected verb formed pragmatically odd combinations, a pragmatically plausible alternative was added to clarify the intended interpretation and syntactic structure. We also added a few sentences at the end of the corpus to test basic cases of binding and noncanonical word orders to ensure that the modifications made to the model did not break these mechanisms. They are discussed in the technical supplementary §4.9, being outside of the original research agenda. The participles (four VA-participles and four MA-participles) were included inside their own group to see how the model processes them, but without attempting systematic analysis. Finally, three experimental sentences were included to test some of the alternative analyses that will be discussed in Section 6. In total, the dataset had 263 sentences/construction types (220 core examples + 43 further tests). Each input sentence was a linear list of tokenized and normalized phonological words. The model processed the input sentences incrementally from left to right as if it was comprehending them in a real language use context. The same model processed all sentences from the dataset.

The dependent variables were grammaticality judgments (grammatical, ungrammatical), control (i.e., antecedent selection, object and subject control), thematic roles of all arguments and the plausibility of the syntactic analyses calculated by the model. Grammaticality judgments were assessed by comparing the model output with native speaker judgments provided by the author and, in a few cases where there was uncertainty, by a group of native speakers.[11] Control and thematic role interpretations, which the model provided as output, were matched with native speaker semantic intuitions.

5.1.2 Procedure

The script processed all sentences from the input and paired them with an output. Each sentence was processed incrementally, one word at a time. The output contained, per each input sentence, a grammaticality judgment and a derivation (the whole process of calculating each sentence, word by word) and, for each grammatical sentence, a syntactic and semantic analysis. The latter contained control dependencies and thematic roles among other attributes. The output was provided by the algorithm in the form of text files. These files, together with the input dataset and the lexicons, constitute the raw data of this study.

5.2 Results

5.2.1 Observational adequacy

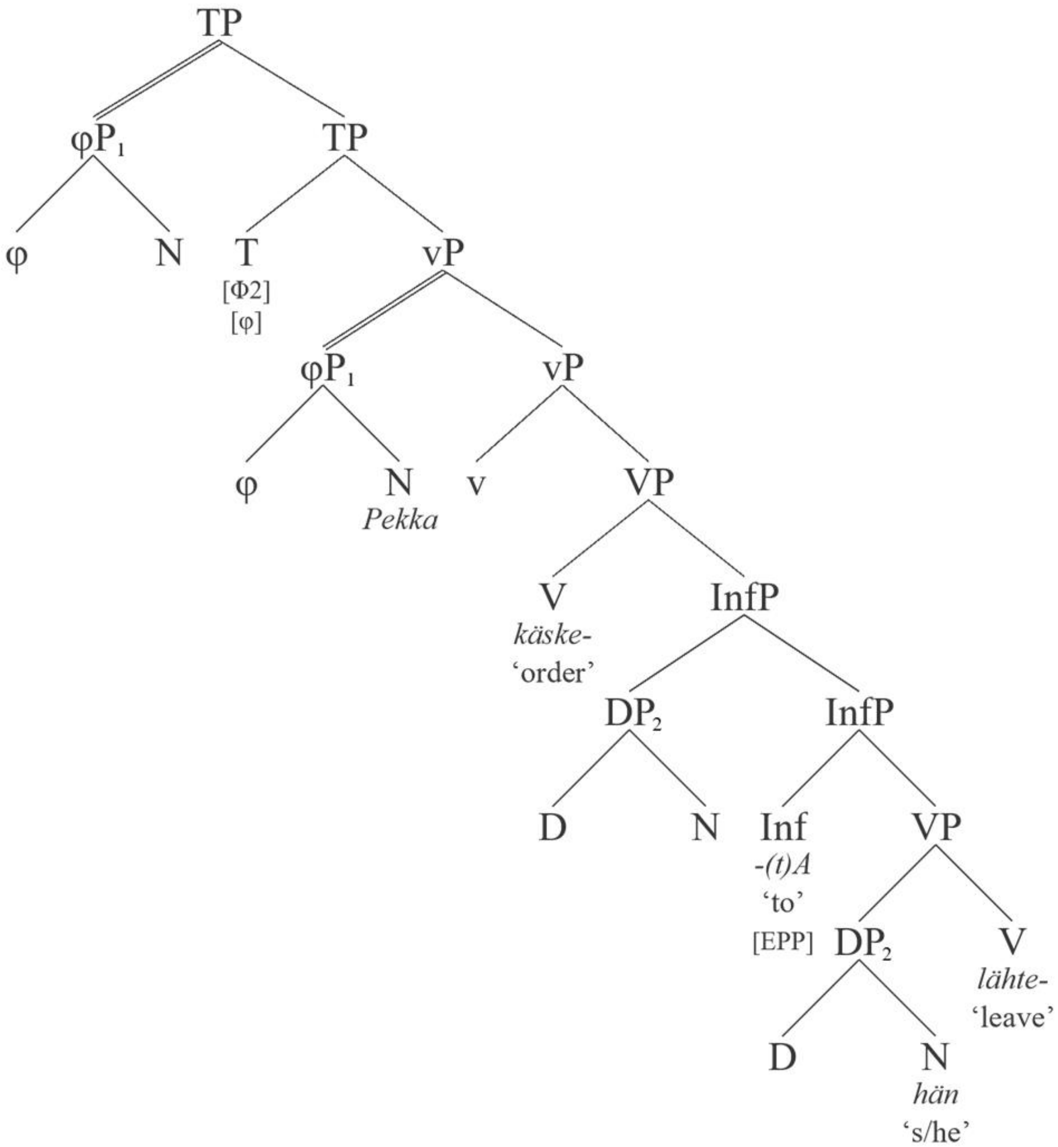
When it comes to the grammaticality judgments, the model reached 100 % accuracy. The hypothesized principles and lexical features predicted the native speaker intuitions concerning grammaticality. In what follows, we will examine the linguistic analyses, antecedent dependencies, derivations and the thematic roles predicted by the model. When a sentence number is prefaced with #, it refers to the sentence number in the dataset, not in this article.

5.2.2 A-infinitive, VA-infinitive

We begin by examining how the model analyses Finnish A-infinitives and VA-infinitives. A grammatical sentence with an embedded A-infinitival clause (28) is analysed by the model as (29). All phrase structures shown here are the syntax-semantics (LF-)interface objects generated, and transformed into images, by the algorithm.

(28)	Pekka	käsk-i	hän-en	lähte-ä. (#11)
	Pekka.NOM	order-PST.3SG	he.GEN	leave-A/INF
	‘Pekka ordered him to leave.’			

(29)



The symbol [φ] means that overt agreement was displayed at spellout and was realized inside the head as features; ±EPP refers to the requirement that the head must have an overt phrasal specifier (+) or that it cannot have one (−), while the lack of EPP means that the subject is optional; Φ1 requires that either agreement or an overt phrasal specifier is present but not both; Φ2 requires one but allows also both. All phrase structures produced by the algorithm are asymmetric and binary-branching. The phrase structure at the LF interface terminates to primitive constituents (e.g., D, φ, T, v, V and Inf in (29)) that contain feature sets retrieved from the lexicon (3, Figure 2). They are represented in the images by the major lexical categories. The labels for complex constituents, shown in the images, are not part of the constituents but were provided by a labelling algorithm. Subscripts indicate chains in the usual sense. Double line represents adjunction and scrambled constituents.

Both the finite and the infinitival predicate were decomposed based on the input words as specified in Section 4.1. The finite predicate *käsk-i* ‘order-PST.3SG’ contains the verbal stem *käske-* ‘order’, transitivizer/voice head v, tense (past) and agreement (third person singular). Agreement features were inserted inside T, as shown by the occurrence of [φ]. The infinitival *lähte-ä* ‘leave-A/INF’ consisted of the infinitival head A/inf^0 *–(t)A* and the verbal stem *lähte-* ‘leave’, both which appear as independent lexical heads in (29). The infinitival phrase (labelled as InfP in the image by the image generation algorithm) was merged to the complement position of the main verb *käske-* ‘order’ to represent the fact that the event of leaving was interpreted as being the object of ordering. *Pekka* was interpreted as the agent of ordering, while *hän-en* ‘s/he-GEN’ was interpreted as the agent of leaving. The whole sentence is interpreted so that Pekka asked or ordered somebody (not himself) to leave. This information is visible in the results file generated by the algorithm, part of which is shown in Figure 5. The predicted thematic roles are visible on the line 238.

```
233 11.  Pekka käski hänen lähteä
234
235     [<ϕ Pekka>:1 [T [<__>:1 [v [käske- [[D hän]:2 [-(t)A [__:2 lähte-]]]]]]]
236
237     Semantics:
238     Thematic roles: ['Causer/Agent of v(v)°: <ϕ Pekka>', 'Patient of V(order)°: InfP', 'Agent of V(leave)°: [D hän]']
239     Arguments: ['Argument for T° is <ϕ Pekka>', 'Argument for v° is [D hän]', 'Argument for käske-° is [D hän]', 'Argu
240     Speaker attitude: ['Declarative']
241     Assignments:
242     [ϕ Pekka] ~ 2, [D hän] ~ 4, Weight 1
243     [ϕ Pekka] ~ 4, [D hän] ~ 2, Weight 1
244     Information structure: {'Marked topics': ['<ϕ Pekka>'], 'Neutral gradient': ['[D hän]'], 'Marked focus': []}
```

Figure 5. A screenshot from the results file generated by the algorithm showing the syntactic analysis (at the syntax-semantics interface, line 235) and aspects of semantic

interpretation (lines 238–244) created by the semantic component (see Figure 1). The predicted thematic roles, specifically, are listed on line 238. Every sentence that was judged grammatical by the model is associated with a similar entry and much be checked for correctness.

Consider (30) where the A-infinitival was combined with the nonfinite agreement morpheme. The A-infinitival head contains –ΦPF as a lexical property, which rules out (30a) and (30b).

(30) a. *Pekka käsk-i [hän-en lähte-ä-nsä]. (#12–13)

*Pekka.*NOM *order-*PST.3SG *he-*GEN *leave-*A/INF-PX/3P

b. *Pekka käsk-i [lähte-ä-nsä]. (#15)

*Pekka.*NOM *order-*PST.3SG *leave-*A/INF-PX/3P

Consider the variation (31) next.

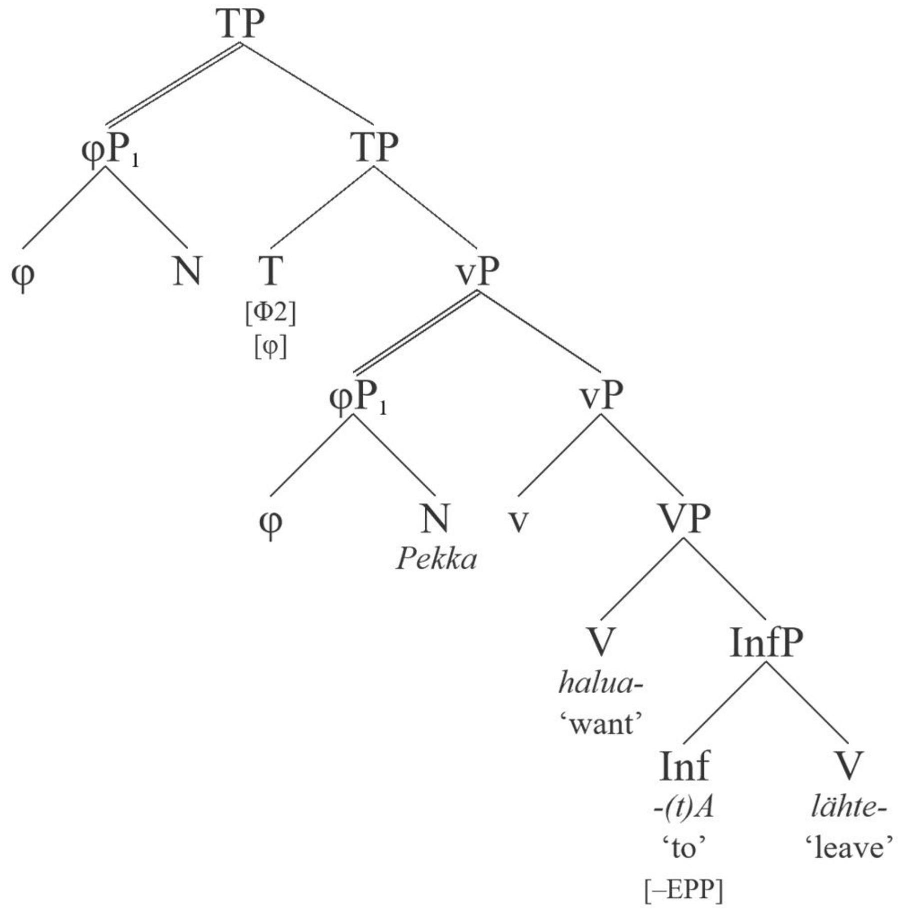
(31) Pekka halus-i lähte-ä. (#18)

*Pekka.*NOM *want-*PST.3SG *leave-*A/INF

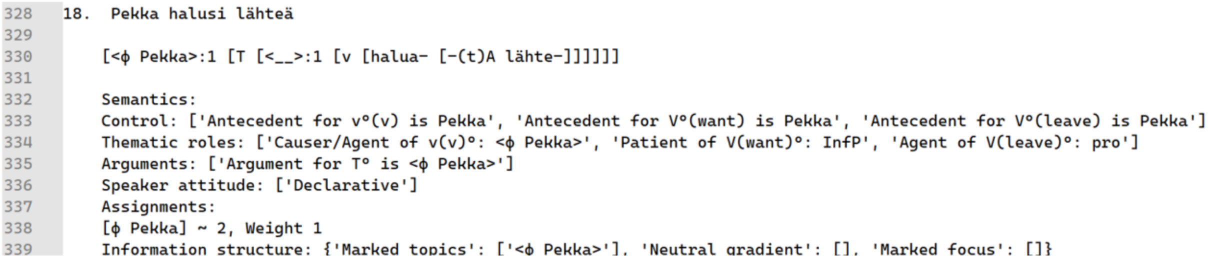
'Pekka wanted to leave.'

There is no overt embedded subject, and the main verb has been changed (*käskeä* 'to order' → *haluta* 'to want') such that it can select the subjectless A-infinitive. This input sentence is analysed by the model as (32).

(32)



The phrasal subject is missing from the embedded clause, so this time only the infinitive clause [$_{\text{InfP}}$ Inf V] is generated. This clause is projected from the isolated infinitival predicate *lähte-ä* 'leave-A/INF' by using the principles elucidated in Section 4.1. The missing subject triggers scanning, which targets Pekka as the thematic agent of leaving. This information must be checked from the output files generated by the model. The relevant entry is show in Figure 6. Notice the appearance of the new field 'Control' (line 333) which reports that scanning was activated and targeted *Pekka*.[12]



```
328 18.  Pekka halusi lähteä
329
330      [<ϕ Pekka>:1 [T [<__>:1 [v [halua- [-(t)A lähte-]]]]]]
331
332      Semantics:
333      Control: ['Antecedent for v°(v) is Pekka', 'Antecedent for V°(want) is Pekka', 'Antecedent for V°(leave) is Pekka']
334      Thematic roles: ['Causer/Agent of v(v)°: <ϕ Pekka>', 'Patient of V(want)°: InfP', 'Agent of V(leave)°: pro']
335      Arguments: ['Argument for T° is <ϕ Pekka>']
336      Speaker attitude: ['Declarative']
337      Assignments:
338      [ϕ Pekka] ~ 2, Weight 1
339      Information structure: {'Marked topics': ['<ϕ Pekka>'], 'Neutral gradient': [], 'Marked focus': []}
```

Figure 6. A screenshot from the results file generated by the algorithm, showing the entry for the input sentence *Pekka halusi lähte-ä* 'Pekka wanted leave-A/INF' (#18). The thematic roles and control dependencies, which are generated on the basis of the syntax-semantic interface representations, are on lines 334 and 333, respectively.

The A-infinitive head has feature –EPP, which explains why overt phrasal subjects cannot be projected inside the nonfinite clause (33).

(33) Pekka halus-i (*hän-en) lähte-ä. (#16, 18)
Pekka.NOM want-PST.3SG he-GEN leave-A/INF
Intended: 'Pekka wanted him to leave.'

As first observed by Vainikka (1989:283–287), the EPP behavior of the A-infinitive depends on the selecting verb: *käskeä* 'to order' requires that a phrasal subject is present, *haluta* 'to want' blocks it. In the algorithm used in this study the selecting verb determines the EPP-behavior directly; an alternative is to assume that the lexicon contains two A-infinitival predicates of which only one can occur in this environment.

Example (34) shows how the model handles selection restrictions. The main verb 'believe' cannot select for the A-infinitive and the sentence is judged ungrammatical.

(34) a. *Pekka usko-i [(hän-en) lähte-ä(-nsä)]. (#20–23)
Pekka.NOM believe-PST.3SG (he.GEN) leave-A/INF(-PX/3P)

Intended: ’Pekka believed that he will leave.’

b. *Pekka nukku-i [(hän-en) lähte-ä(-nsä)]. (#28–31)

*Pekka.*NOM *sleep-*PST.3SG *(he-*GEN*) leave-*A/INF*(-*PX/3P*)*

Sentence (34a) is ruled out because the main verb *uskoa* ‘believe’ cannot select A-infinitives (specifically, the A-infinitive head), it selects tensed clauses which in turn implements the hypothesis that verbs of this type require ‘propositional’ complements. The A-infinitival head does not have tense specification (it does not show overt tense alteration). Sentence (34b) is ungrammatical because the intransitive predicate cannot select an infinitival clause complement. To find out why they were judged ungrammatical, the researcher must consult the derivational log file. In the case of (34a), sentence #20 in the dataset, the derivational log file shows that the verb ‘believe’ failed a negative complement selection test (lines 5800–2).

```
5781    Transfer [[ϕ Pekka] [T(v,V) [[D hän] -(t)A(V)]]] to LF:-----------------------------------------------------
5782
5783        Head Chain(-(t)A) => [[ϕ Pekka] [T(v,V) [[D hän] [-(t)A lähte-]]]]
5784        Head Chain(T) => [[ϕ Pekka] [T [v(V) [[D hän] [-(t)A lähte-]]]]]
5785        Head Chain(v) => [[ϕ Pekka] [T [v [usko- [[D hän] [-(t)A lähte-]]]]]]
5786        T acquired ϕ-completeness.
5787        Extraposition(usko-) => [[ϕ Pekka] [T [v [usko- [[D hän] [-(t)A lähte-]]]]]]
5788        Scrambling Chain(<ϕ Pekka>) => [<ϕ Pekka>:61 [T [<__>:61 [v [usko- [[D hän] [-(t)A lähte-]]]]]]]
5789        lähte-° agrees with lähte- (lähte-) and values nothing (no useful features available).
5790        usko-° agrees with D ([D hän]) and values PHI:DET:DEF PHI:HUM:HUM PHI:NUM:SG PHI:PER:3 PHI:PRON:PRON
5791        v° agrees with D ([D hän]) and values PHI:DET:DEF PHI:HUM:HUM PHI:NUM:SG PHI:PER:3 PHI:PRON:PRON
5792        T° agrees with ϕ (<ϕ Pekka>) and values PHI:DET:DEF PHI:HUM:NONHUM PHI:NUM:SG PHI:PER:3 PHI:PRON:NONPRON
5793        usko- failed feature -COMP:A/inf
5794        usko- failed Selection test
5795
5796        Syntax-semantics interface endpoint:
5797        [<ϕ Pekka>:61 [T [<__>:61 [v [usko- [[D hän] [-(t)A lähte-]]]]]]]
5798
5799        LF-interface and postsyntactic legibility tests:
5800        usko- failed feature -COMP:A/inf
5801        usko- failed Selection test
5802        SOLUTION WAS REJECTED.
```

Figure 7. A screenshot from the derivational log file, showing that the main verb failed a selection test against the A-infinitive (lines 5793–4, 5799–5802). The input syntactic analysis is on line 5797. The operations on lines 5783–5792 describe what occurred during transfer (for the notion of transfer, see Section 4.1).

The A-infinitival cannot occur in an adjunct position (#24–27), which was captured by blocking the adjunction option.[13] Also the combination of an A-infinite with an intransitive predicate is ungrammatical (#28–31)(35). This was correctly calculated because the adjunction option was blocked, while the intransitive verb could not have a complement.

(35) *Pekka nukku-i [hän-en lähte-ä(-nsä)]. (#28, 29)
*Pekka.*NOM *sleep-*PST.*3*SG *he-*GEN *leave-*A/INF*(-*PX/3P*)*

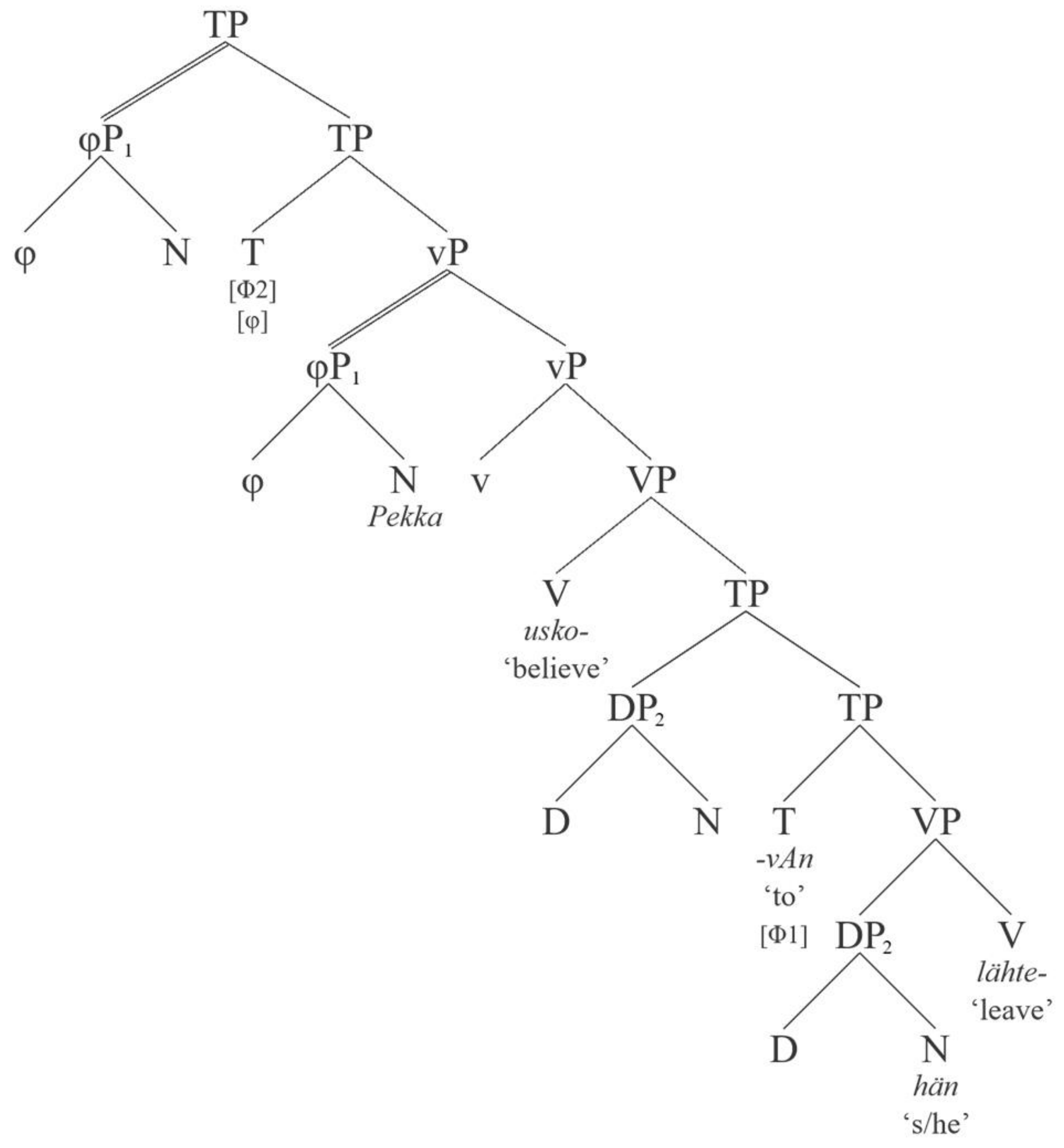
The VA-infinitive is more propositional in meaning than the A-infinitival. This was represented by assuming that the VA-infinitival head has the tense feature corresponding to overt past-present tense alteration (*lähte-vän* ~ *lähte-neen* 'leave-VA/INF.PRS ~ leave-VA/INF.PST'). We assume that 'believe' selects for finite and non-finite clauses with T (36).

(36) a. Pekka usko-i [hän-en lähte-vän]. (#40)
*Pekka.*NOM *believe-*PST.*3*SG *he-*GEN *leave-*VA/INF
'Pekka believed that he leaves.'

b. Pekka usko-i [että hän läht-e-e].
*Pekka.*NOM *believe-*PST.*3*SG *that hän.*NOM *leave-*PRS*-3*SG
'Pekka believed that he leaves.'

Sentence (36a) is calculated as (37).

(37)



Pekka was the agent of wanting, the third party was the agent of leaving. The VA-infinitive is complemented to the main verb 'believe'. The VA-infinitive differs from the A-infinitive in terms of the lexical features of the VA-infinitive head. The first difference is that the latter cannot be selected by verbs in the order-class (38).

(38) *Pekka käsk-i [(hän-en) lähte-vä(-nsä)]. (#32–35)

*Pekka.*NOM *order-*PST.*3*SG *he-*GEN *leave-*VA/INF*(-*PX/3P*)*

Intended: 'Pekka ordered him to leave.'

The second difference is that the VA-infinitive head has Φ1, which requires either an overt phrasal subject (as in the example above) or overt nonfinite agreement suffix that can substitute for a full pronoun (39).

(39) a. *Agreement substitutes for the embedded overt phrasal subject*

Pekka usko-i lähte-vä-nsä. (#43)

*Pekka.*NOM *believe-*PST.3SG *leave-*VA/INF-PX/3P

'Pekka believed that he (Pekka, *third party) will leave.'

b. *No embedded subject, no agreement*

*Pekka usko-i lähte-vän. (#42)

*Pekka.*NOM *believe-*PST.3SG *leave-*VA/INF

Intended: 'Pekka believed that he (Pekka, *third party) will leave.'

c. *Both overt subject and agreement, redundant identification*

*Pekka usko-i hän-en lähte-vä-nsä. (#41)

*Pekka.*NOM *believe-*PST.3SG *he-*GEN *leave-*VA/INF-PX/3P

Intended: 'Pekka believed that he (*Pekka, third party) will leave.'

d. *Only overt phrasal subject*

Pekka usko-i hän-en lähte-vän. (#40)

*Pekka.*NOM *believe-*PST.3SG *he-*GEN *leave-*VA/INF

'Pekka believed that he (*Pekka, third party) will leave.'

See also sentences #36–39 in the dataset which illustrate the same generalization. The A-infinitive and VA-infinitive occur only in the complement positions of transitive verbs; examples where they could only be interpreted as adjuncts or as complements of intransitive verbs (40) are judged ungrammatical (#24–27, 44–51).

(40) *Pekka nukku-i lähte-vä-nsä.

Pekka.NOM sleep-PST.3SG leave-VA/INF-PX/3P

Thus, they are both marked as being resistant to adjunction.

5.2.3 MA-infinitives (five types), E-infinitives

All MA-infinitives (41) have similar lexical entries, with most differences having to do with the type of main verbs they combine with and whether they are attached to the structure as low (VP) or high (TP) adjuncts. They do not host phrasal subjects (hence are marked with –EPP, #52–61) and can be in adjunct positions (#125–133), as indicated by the translations and the presence of the direct object argument.

(41) a. Pekka näk-i hän-et [lähte-mässä]. (#125)

Pekka.NOM see-PST.3SG he-ACC leave-MA.INE/INF

'Pekka saw him leaving.'

b. Pekka pyys-i hän-tä [lähte-mään]. (#128)

Pekka.NOM request-PST.3SG he-PAR leave-MA.ILL/INF

'Pekka asked/requested him to leave.'

c. Pekka est-i hän-tä [lähte-mästä]. (#130)

Pekka.NOM prevent-PST.3SG he-PAR leave-MA.ELA/INF

'Pekka prevented him from leaving.'

d. Pekka est-i hän-tä [lähte-mällä]. (#131)

Pekka.NOM prevent-PST.3SG he-PAR leave-MA.ADE/INF

'Pekka prevented him by leaving.'

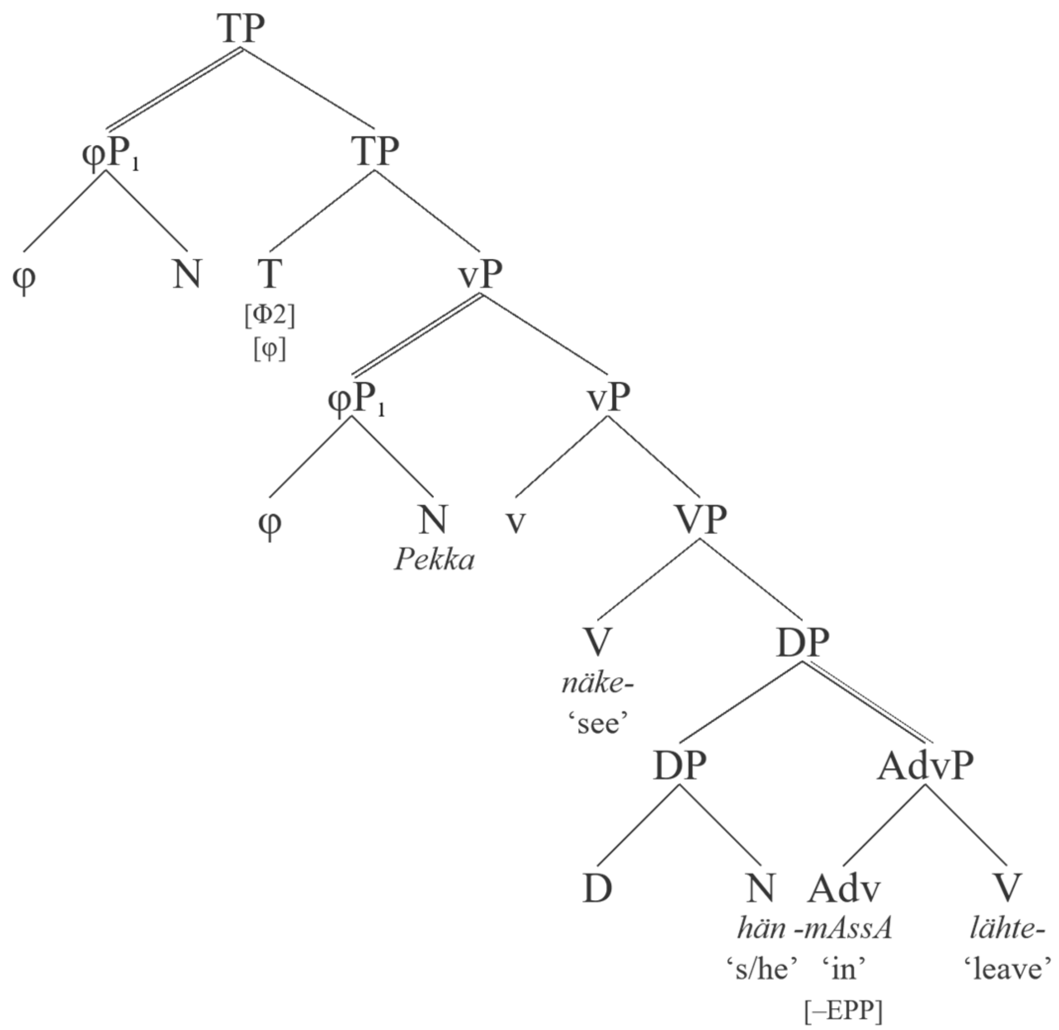
e. Pekka saavutt-i hän-et [juokse-matta]. (#133)

Pekka.NOM reach-PST.3SG he-ACC run-MA.ABE/INF

'Pekka reached him without running.'

The MA-infinitives differ from the A-infinitives and VA-infinitives in that with the exception of the MA.ILL-infinitive (#83) they can only occur in an adjunct position (#62–91). Notice that all the sentences in (41) contain a direct object in the main clause. Example (42) illustrates the output for (41a)(#125).

(42)



Adjunct attachment (as well as scrambled arguments) is marked by the double line.[14] The accusative marked object *hän-et* 'he-ACC' is in the direct object position (Vainikka 1989:261–265), while the MA-infinitival phrase has been attached to a right-adjunct position inside the VP. Because the adjunct infinitival is attached to a lower position, antecedent scanning targets the direct object as the antecedent of 'leave' and generates an interpretation where Pekka saw something while a third person was leaving. This generates object control, as shown in Figure 8.

```
1237 125.  Pekka näki hänet lähtemässä
1238
1239     [<ϕ Pekka>:1 [T [<__>:1 [v [näke- [[D hän] <-mAssA lähte->]]]]]]
1240
1241     Semantics:
1242     Control: ['Antecedent for Adv°(in) is hän', 'Antecedent for V°(leave) is hän']
1243     Thematic roles: ['Causer/Agent of v(v)°: <ϕ Pekka>', 'Patient of V(see)°: [D hän]', 'Agent of V(leave)°: pro']
1244     Arguments: ['Argument for T° is <ϕ Pekka>', 'Argument for v° is [[D hän] <-mAssA lähte->]', 'Argument for näke-° is [[D hän] <-mAssA lähte->]']
1245     Aspect: ['Aspectually bounded', 'Aspectually bounded']
1246     Speaker attitude: ['Declarative']
```

Figure 8. Object control: both the adverbial head and the verb take the direct object *hänet* ‘him’ as an antecedent (line 1242). Adjunction is marked by <, > in the symbolic notation.

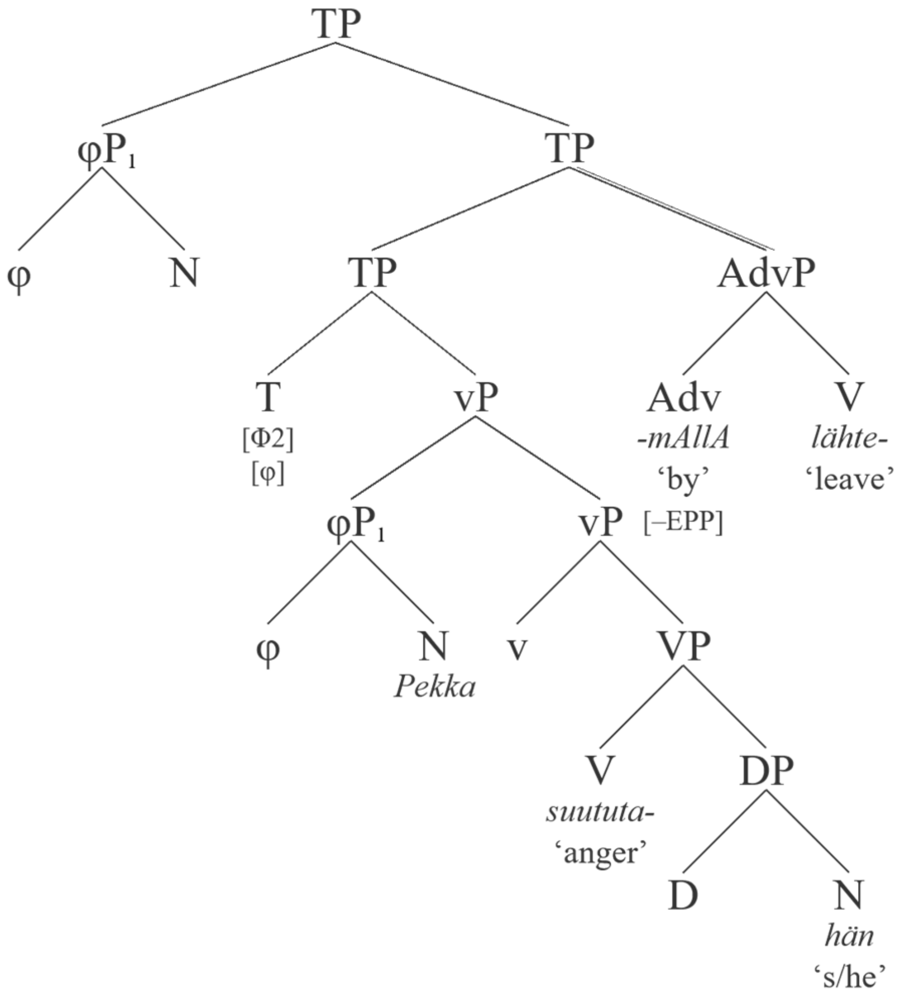
TP-adverbials are merged to a higher position inside the TP and take the matrix subject as their antecedent. This is illustrated by (43), analyzed as (44) where the MA-infinitive is right-adjoined to a higher position in the clause, accounting for subject control.

(43) Pekka suututt-i hän-et lähte-mällä. (#132)

Pekka.NOM *anger*-PST.3SG *he*-ACC *leave*-MA.ADE/INF

‘Pekka angered him by leaving.’

(44)



The antecedent scanning algorithm does not see the direct object, so the agent of 'leaving' will be the subject (line 1370 in the results file). This generates subject control. One MA-infinitival, the MA/ILL form, is able to occupy the complement position:

(45) a. ?Pekka käsk-i laula-maan. (#63)

Pekka.NOM order-PST.3SG sing-MA.ILL/INF

'Pekka ordered (one, people) to sing.'

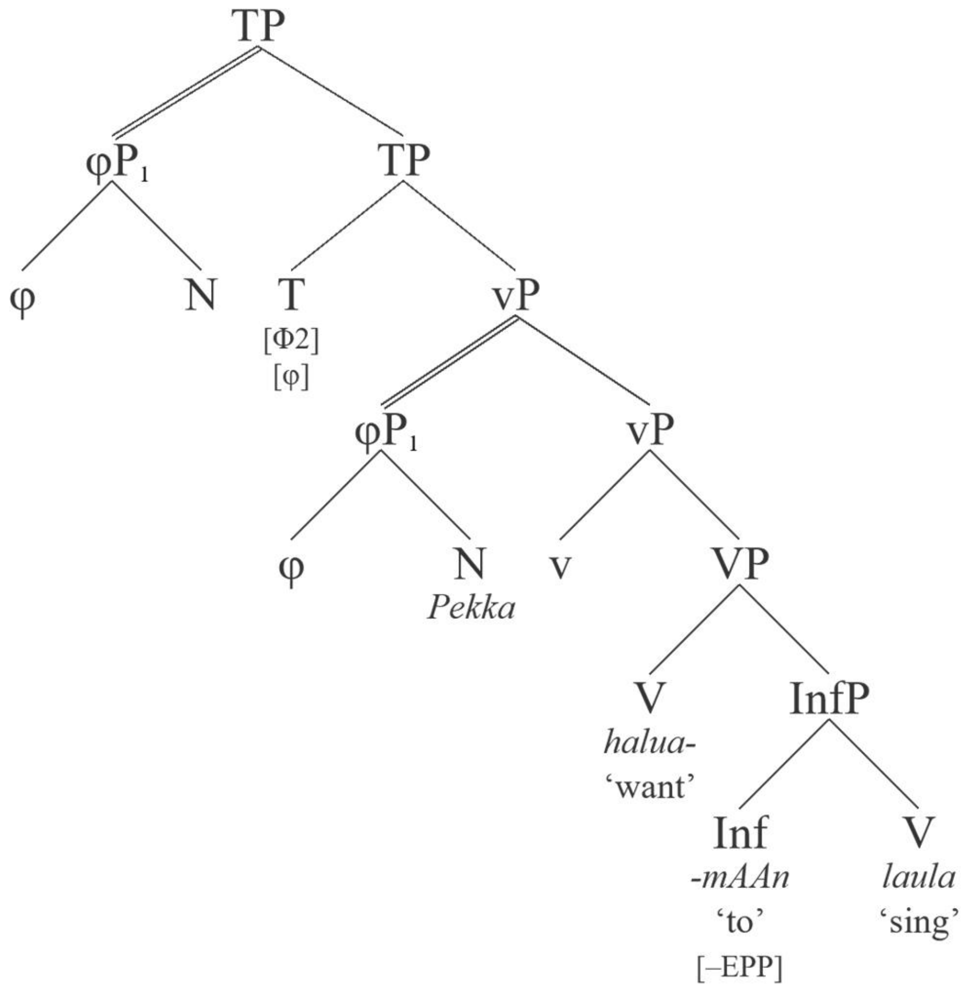
b. Pekka halus-i laula-maan. (#84)

Pekka.NOM want-PST.3SG sing-MA.ILL/INF

'Pekka wanted to sing.'

The analysis for (45b) is shown in (46).

(46)



All MA-infinitives reject agreement and were marked for –ΦPF. This rules out nonfinite agreement throughout the whole class (#57–61, 67–71, 77–81, 88–92, 98–102, 108–112, 119–124, 134–139).[15] They were also marked for –EPP, which rules out sentences with an overt genitive subject (#52–61, 72–81, 93–102, 113–124, 140–150). The MA-infinitivals have a special property in that the type of the infinitival head determines the type of main verbs that they are compatible with. For example, the MA.ILL-infinitival clause can be combined with 'request' but not with 'saw' (47).

(47) a. *Pekka näk-i hän-tä [lähte-mään].

*Pekka.*NOM *see-*PST.*3*SG *he-*PAR *leave-*MA.ILL/INF

b. Pekka pyys-i hän-tä [lähte-mään].

*Pekka.*NOM *request-*PST.*3*SG *he-*PAR *leave-*MA.ILL/INF

'Pekka requested him to leave.'

The contrast was captured by forcing the MA.ILL-infinitive to match with a verb that belongs to a specific semantic class. In this case, it was stipulated that the MA.ILL-infinitive must match with verbs that introduce a 'desired event'. Since 'see' does not belong to this class, the algorithm judges the combination 'see + him leave-MA.ILL/INF' ungrammatical.

The E-infinitival, illustrated by sentence (48) and corresponding roughly to the English 'by doing', is similar to the MA-infinitives in its syntactic properties.

(48) Pekka nukku-i [kuorsat-en.] (#181)

*Pekka.*NOM *sleep-*PST.*3*SG *snore-*E/INF

'Pekka slept while/by snoring.'

The infinitival is attached to the structure as a TP adjunct (49a), never complement (49b), taking the main subject as its antecedent. It does not have its own subject (49c) (#162–163, 166–167, 170–171, 174–175, 178–179) and never exhibits agreement (49b,d) (#163, 165, 167, 169, 171, 173, 175, 177, 179, 182).

(49) a. Pekka tavoitt-i hän-et juost-en. (#175)

*Pekka.*NOM *reach-*PST.*3*SG *he-*ACC *run-*E/INF

'Pekka reached him by running.'

b. *Pekka uskoi/halusi/käski (hän-en) lähti-e(-nsä). (#162–3, 166–7)

*Pekka.*NOM *believed/wanted/ordered* *(he-*GEN*)* *leave-*E/INF*(-*PX/3P*)*

c. *Pekka tavoitt-i hän-et hän-en juost-en. (#174)

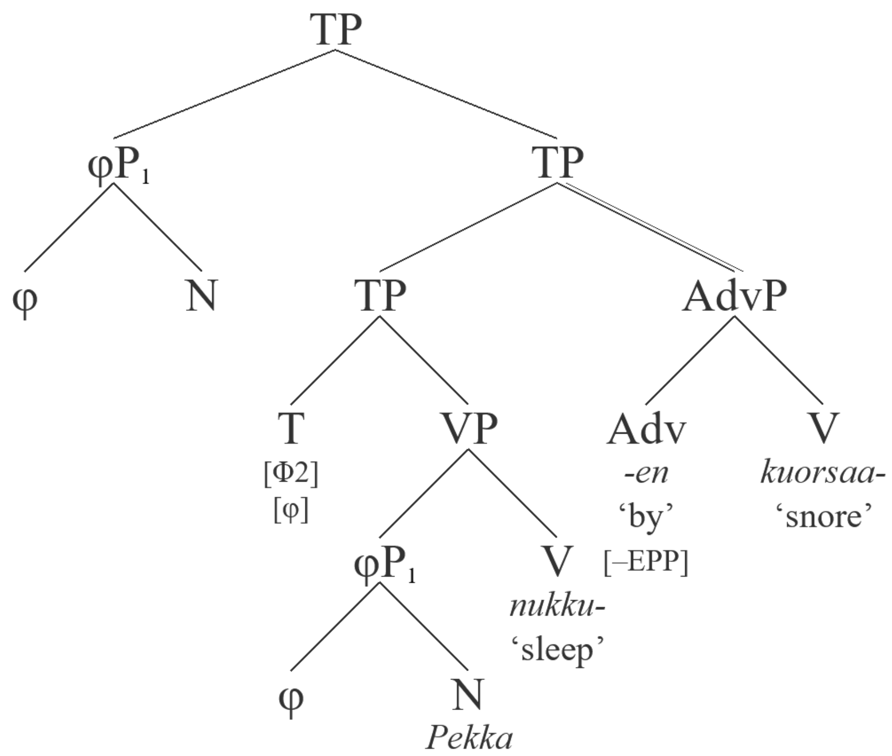
Pekka.NOM reach-PST.3SG he-ACC he-GEN run-E/INF

d. *Pekka tavoitt-i hän-et juost-e-nsa. (#177)

Pekka.NOM reach-PST.3SG he-ACC run-E/INF-PX/3P

The analysis of (48) calculated by the model is (50).

(50)



5.2.4 ESSA-, KSE- and TUA-infinitives

The ESSA-infinitival, corresponding roughly to ‘while doing’ in English, exhibits optional agreement, optional embedded subjects, and only occurs in adjunct positions (compare #183–194 and #195–202). Example (51) illustrates these properties.

(51) a. Pekka näk-i hän-et (hän-en) lähti-essä. (#195, 197)

Pekka.NOM see-PST.3SG he-ACC (he-GEN) leave-ESSA/INF

‘Pekka$_1$ saw him$_2$ while PRO$_1$/(he$_{*1, 2, 3}$) was leaving.’

b. Pekka nukku-i (hän-en) lähti-essä. (#199, 201)

Pekka.NOM sleep-PST.3SG (he.GEN) leave-ESSA/INF

‘Pekka$_1$ slept while PRO$_1$/(he$_2$) was leaving.’

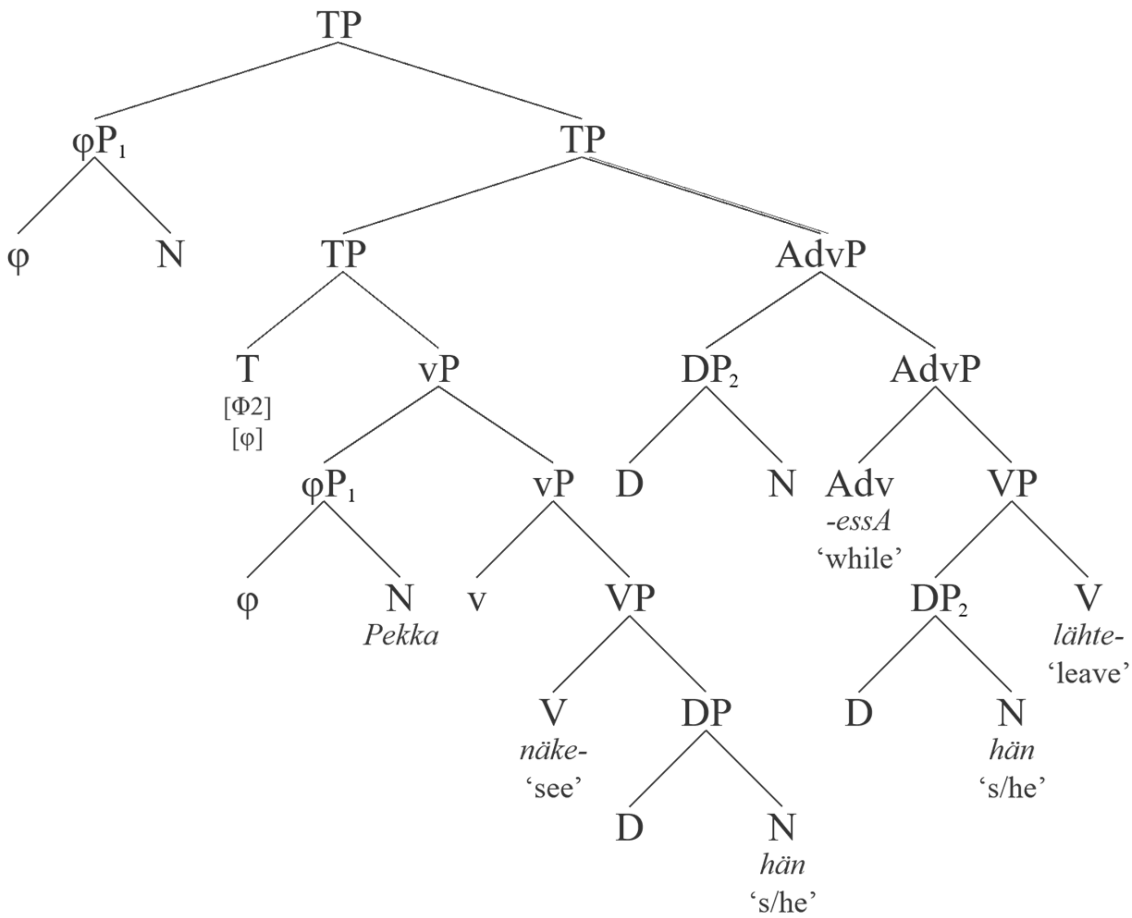
c. Pekka nukku-i (?hän-en) lähti-essä-än. (#198, 200, 202)

Pekka.NOM sleep-PST.3SG (he-GEN) leave-ESSA/INF-PX/3P

‘Pekka$_1$ slept while PRO$_1$/(he$_2$) was leaving.’

Example (51)a is analyzed as (52).

(52)



The ESSA-infinitival is adjoined to a high position inside the TP, triggering subject control. EPP is absent, which makes the subject and its identification optional. Agreement is possible, but not obligatory (51b–c).

The KSE-infinitival, roughly 'in order to do something' in English, is identical in its syntactic behavior to the ESSA-infinitival with the exception that nonfinite agreement is obligatory by +ΦPF (#235, 238) and overt phrasal subjects are illicit –EPP (#236, 238). The KSE-infinitival occurs only in adjunct positions (53)(compare #238 and 223–234). Example (55) shows the analysis of (53a).

(53) a. Pekka nukku-i lähteä-kse-en. (#242)

Pekka.NOM sleep-PST.3SG rest-KSE/INF-PX/3P

'Pekka slept in order to leave.'

b. *Pekka kannust-i hän-tä Merja-n voittaa-kse-en. (#236)

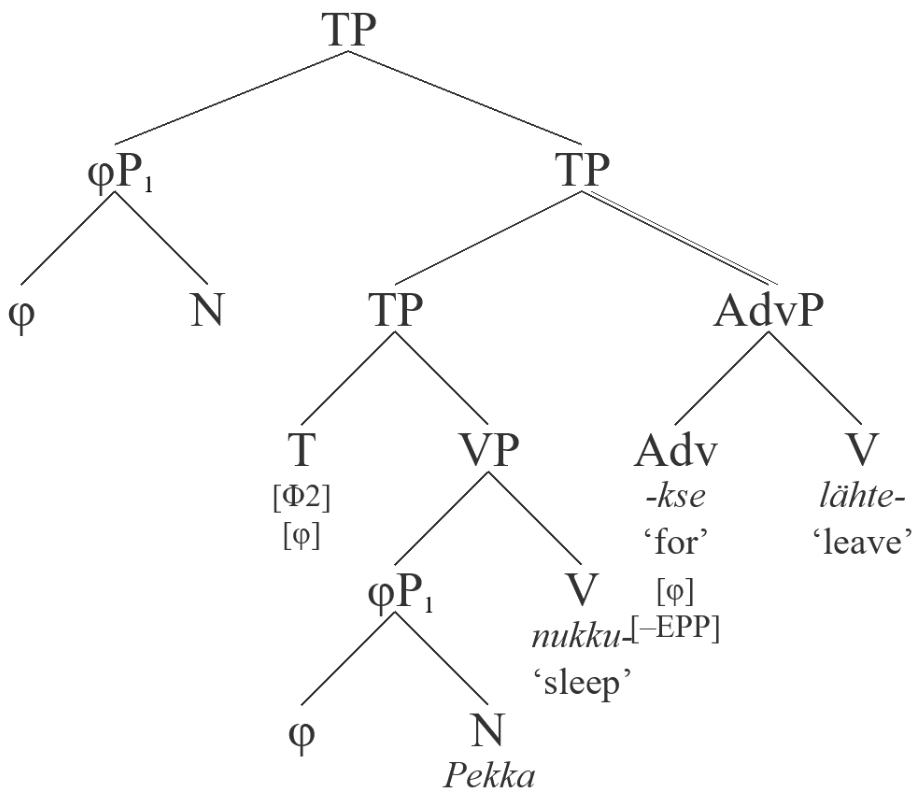
Pekka.NOM support-PST.3SG she-PAR [Merja-GEN win-KSE/INF-PX/3P]

Intended: 'Pekka supported her in order for Merja to win.'

(54) *Pekka halus-i levätä-kse-en. (#230)

Pekka.NOM want-PST.3SG rest-KSE/INF-PX/3P

(55)



Finally, the TUA-infinitive, roughly 'after doing something' in English, is an adjunct adverbial (compare #203–214 and #215, 218) that requires an overt subject (56a)(#215, 217). Nonfinite agreement is grammatical or perhaps marginal (see footnote 11) when the embedded subject is present (56b)(#216).

(56) a. Pekka$_1$ näk-i hän-et *(hän-en) lähde-ttyä. (#215, 217)

Pekka.NOM see-PST.3SG he-ACC he-GEN leave-TUA/INF

'Pekka$_1$ saw him$_2$ after PRO$_1$/(he$_{2, 3}$) left.'

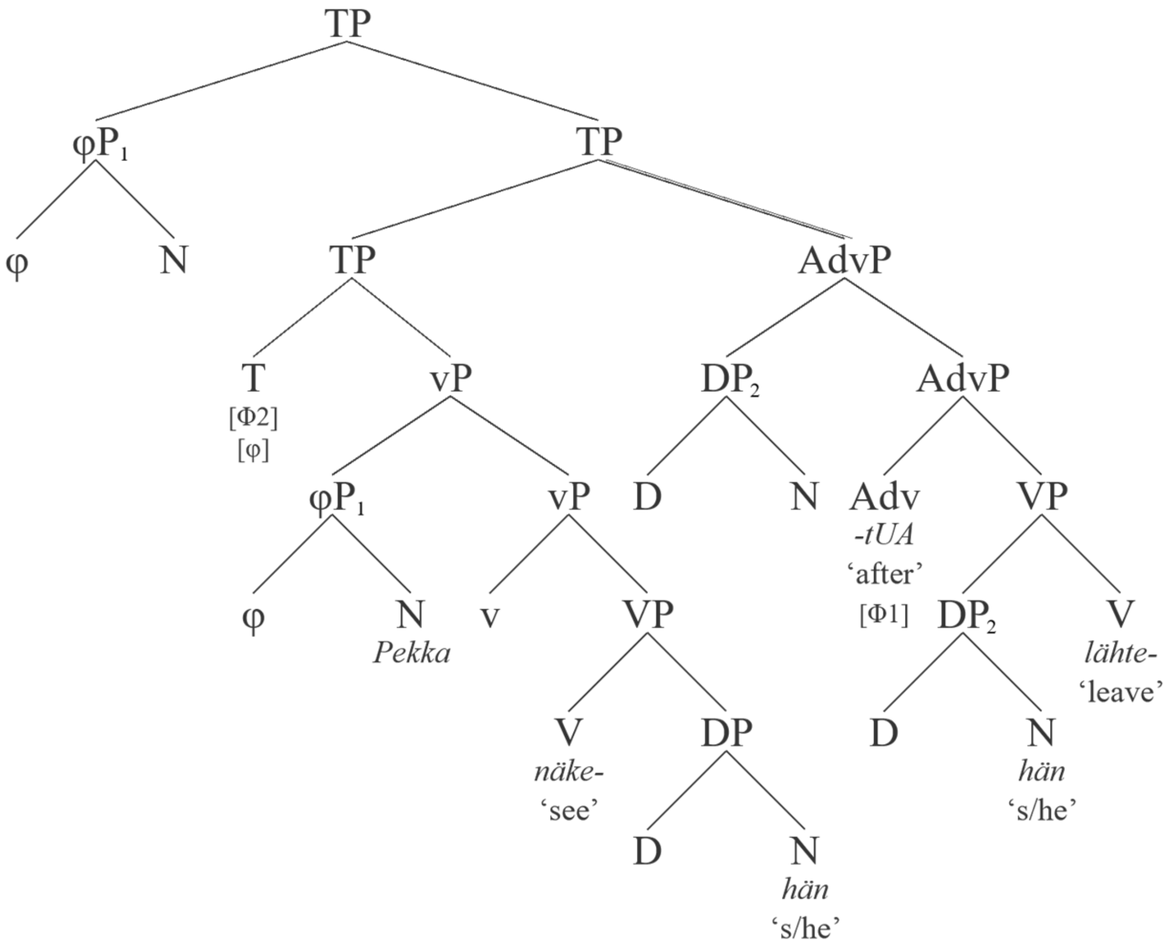
b. Pekka näk-i hän-et (??hänen) lähde-tty-än.[16] (#216, 218)

Pekka.NOM see-PST.3SG he-ACC (he-GEN) leave-TUA/INF-PX/3P

'Pekka$_1$ saw him$_2$ after PRO$_1$/(he$_{2, 3}$) left.'

Example (56)a is analyzed as (57).

(57)



The TUA-infinitive contains the feature Φ1 since, as shown by (53), it requires the occurrence of either overt agreement (#218) or overt subject (#215) but not both (#216). If the (this author's view) marginal combination of the overt subject and the nonfinite agreement (53b) is accepted, then in such grammar the TUA-infinitive head has feature Φ2. Thus, both grammars can be generated by the proposed model.

5.2.5 Participle adjective phrase

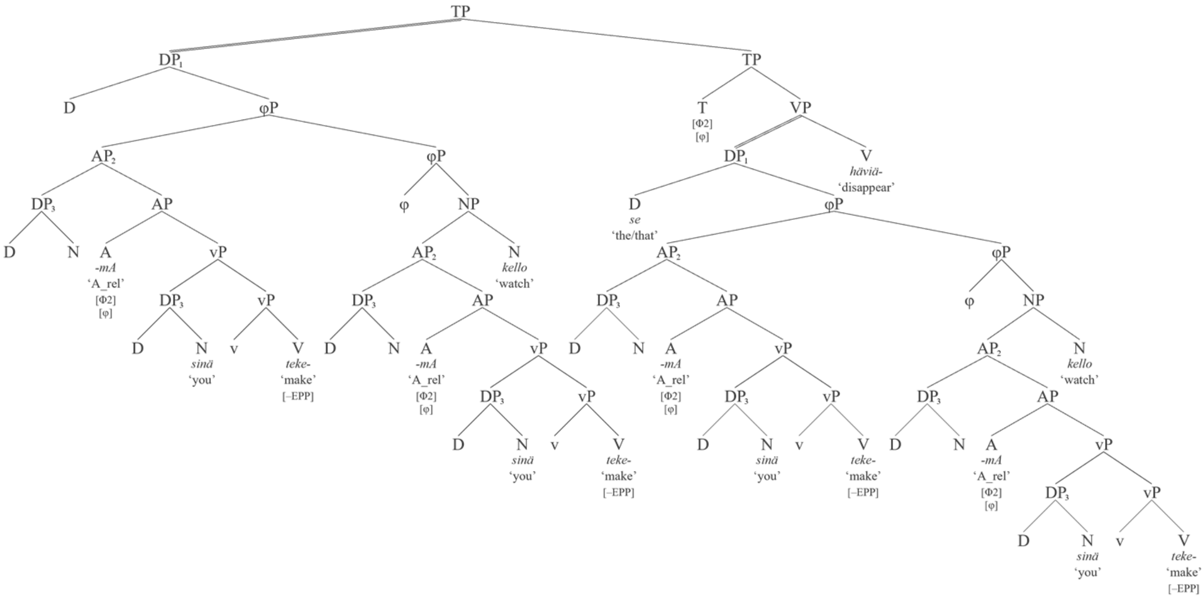
Participle adjective phrases have several nominal properties and cannot be classified unproblematically in terms of the independent syntactic variables thematized in this study. The dataset has, however, examples of both VA-participles (#257–260) and MA-participles. A MA-participle (55)(#253–256) was analyzed by the model as (58).

(58) Se [sinun teke-mä-si] kello hävis-i. (#253)

*it/that.*NOM *you.*GEN *make-*MA/A-PX/2SG *watch.*NOM *disappear-*PST.3SG

‘The/that watch made by you disappeared.’

(59)



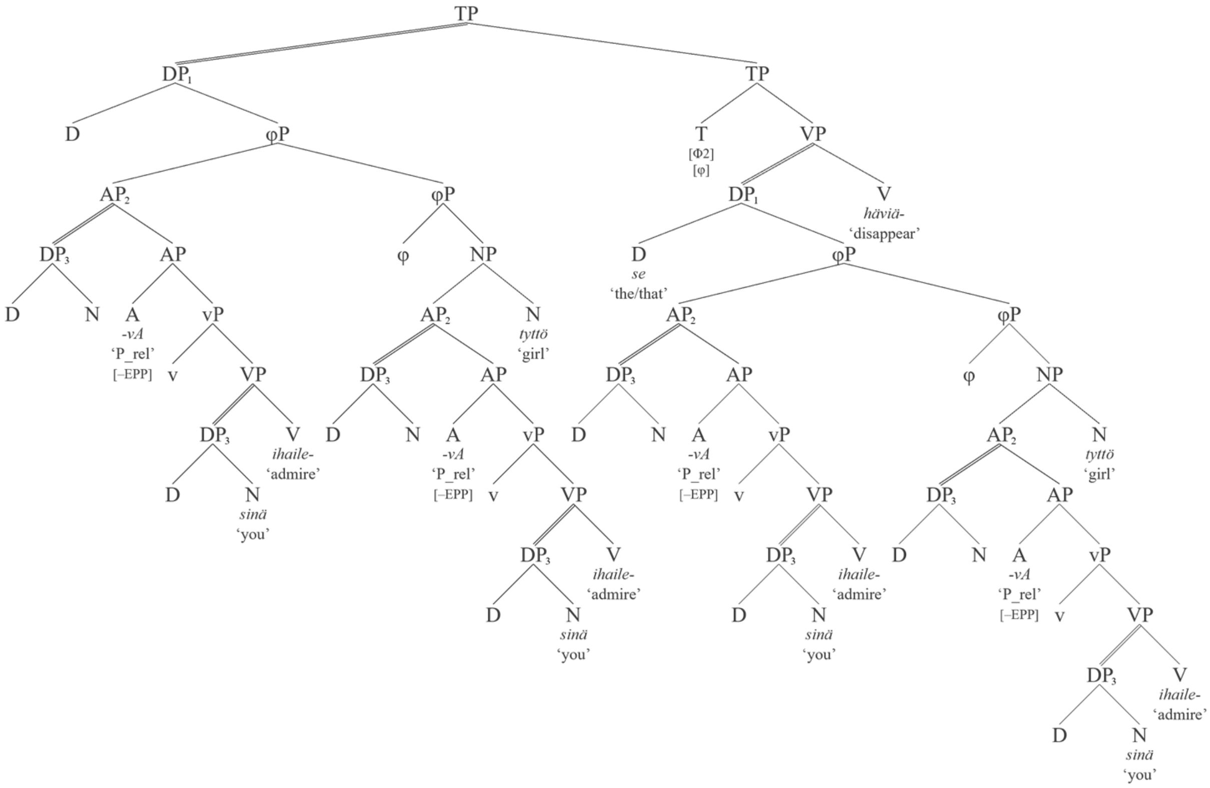
The participle adjective phrase AP = *sinun teke-mä-si* ‘you.GEN make-MA/A-PX/2SG’ meaning ‘x made by you’ is headed by the participle adjective head *-mA* taking a verb phrase complement. The genitive phrasal subject is reconstructed into SpecvP where it represents the agent ‘who made something’. The VA-participle in (60)(#257–260), meaning ‘x who admires you’, is analyzed as (61).

(60) Se [sinu-a ihaile-va] tyttö hävis-i. (#257)

*the/that.*NOM *you-*PAR *admire-*VA/A *girl.*NOM *disappear-*PST.3SG

‘The girl who admires you disappeared.’

(61)



In this case the overt argument inside the participle is reconstructed into the direct object position and becomes the patient of admiration. Thus, the participle adjective phrase is interpreted as 'x such that x admires you'.

6 DISCUSSION

The information processing model partitioned the dataset correctly into grammatical and ungrammatical sentences and provided the former with (in this author's view) plausible syntactic analyses and correct semantic interpretations. To the extent that the analyses are viewed as implausible or wrong can be assessed unambiguously since the calculations, correct or incorrect, are in the derivational log file and other raw output files. The set of lexical features used in the final simulation trial, which provided a 100 % match between theory and data, is summarized in Table 3.

Table 3. Lexical features used in the final simulation trial.

#	**Nonfinite clause**	**Label**	**EPP**	**ΦPF**	**ADV**	**TP/VP**	**SI**	**Other**
1	A-infinitive (–EPP)	A/inf	–EPP	–ΦPF	–ADV	—	1	Extra semantic feature
2	A-infinitive (+EPP)	A/inf	+EPP	–ΦPF	–ADV	—	3	Extra semantic feature
3	VA-infinitive	VA/inf	—	Φ1	–ADV	—	4	T(ense)
4	ESSA-infinitive	ESSA/inf	(both)	(both)	+ADV	TP	1, 6	
5	E-infinitive	E/inf	–EPP	–ΦPF	+ADV	TP	1	
6	TUA-infinitive	TUA/inf	—	Φ1(Φ2)	+ADV	TP	6	
7	KSE-infinitive	KSE/inf	–EPP	+ΦPF	+ADV	TP	5	
8	MA.INE-infinitive	MA.INE/inf	–EPP	–ΦPF	+ADV	VP	2	
9	MA.ABE-infinitive	MA.ABE/inf	–EPP	–ΦPF	+ADV	TP	1	
10	MA.ADE-infinitive	MA.ADE/inf	–EPP	–ΦPF	+ADV	TP	1	
11	MA.ELA-infinitive	MA.ELA/inf	–EPP	–ΦPF	+ADV	VP	2	Extra semantic feature
12	MA.ILL-infinitive	MA.ILL/inf	–EPP	–ΦPF	(both)	VP	2	Extra semantic feature
13	MA-participle	MA/A	—	Φ2	+ADV	φP	4, 6	Nominal syntax
14	VA-participle	VA/A	–EPP	–ΦPF	+ADV	φP	1 + 2	Nominal syntax

Symbols: ±EPP = whether an overt phrasal subject is mandatory (+),illicit (–) or optional (no feature); ±ΦPF = whether overt agreement is mandatory (+), illicit (–) or optional (no feature); Φ1 = either agreement or an overt subject must occur but not both; Φ2 = either agreement or overt subject must occur; ±ADV = whether the infinitival can or cannot appear in an adjunct position; TP/VP = level of adjunct attachment, leading to subject (TP) and object control (VP); SI = subject identification, referring to the numbers in Figure 9; T = tense.

There is a clear correlation between –EPP and –ΦPF, although the –EPP A-infinitive (row 2) and the KSE-infinitive (row 7) constitute exceptions. Whether this correlation holds in Finnish in general, or crosslinguistically, will be left for future research. Some of the features appearing in Table 3 can be arranged into a functional hierarchy shown in Figure 9. Intuitively the hierarchy determines how the subject of the nonfinite predicate is identified on the basis of the overt elements appearing in the input sentence.

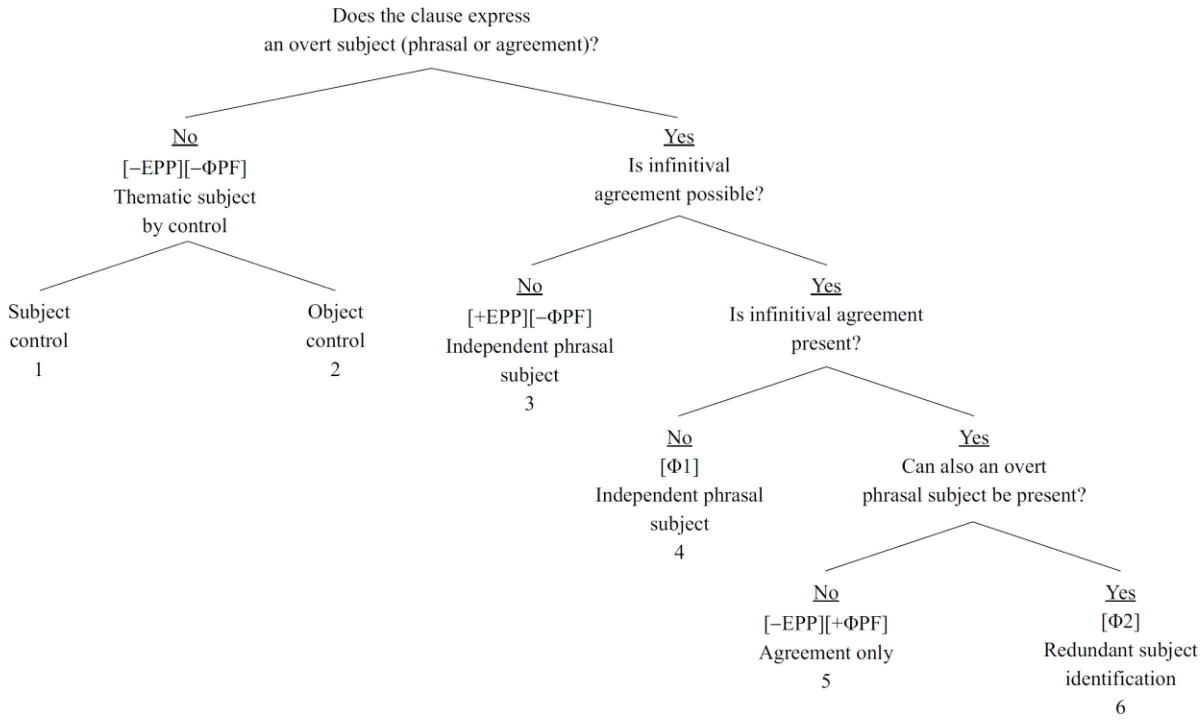


Figure 9. Hierarchical dependencies between the lexical features positing in this study. See the main text for explanation.

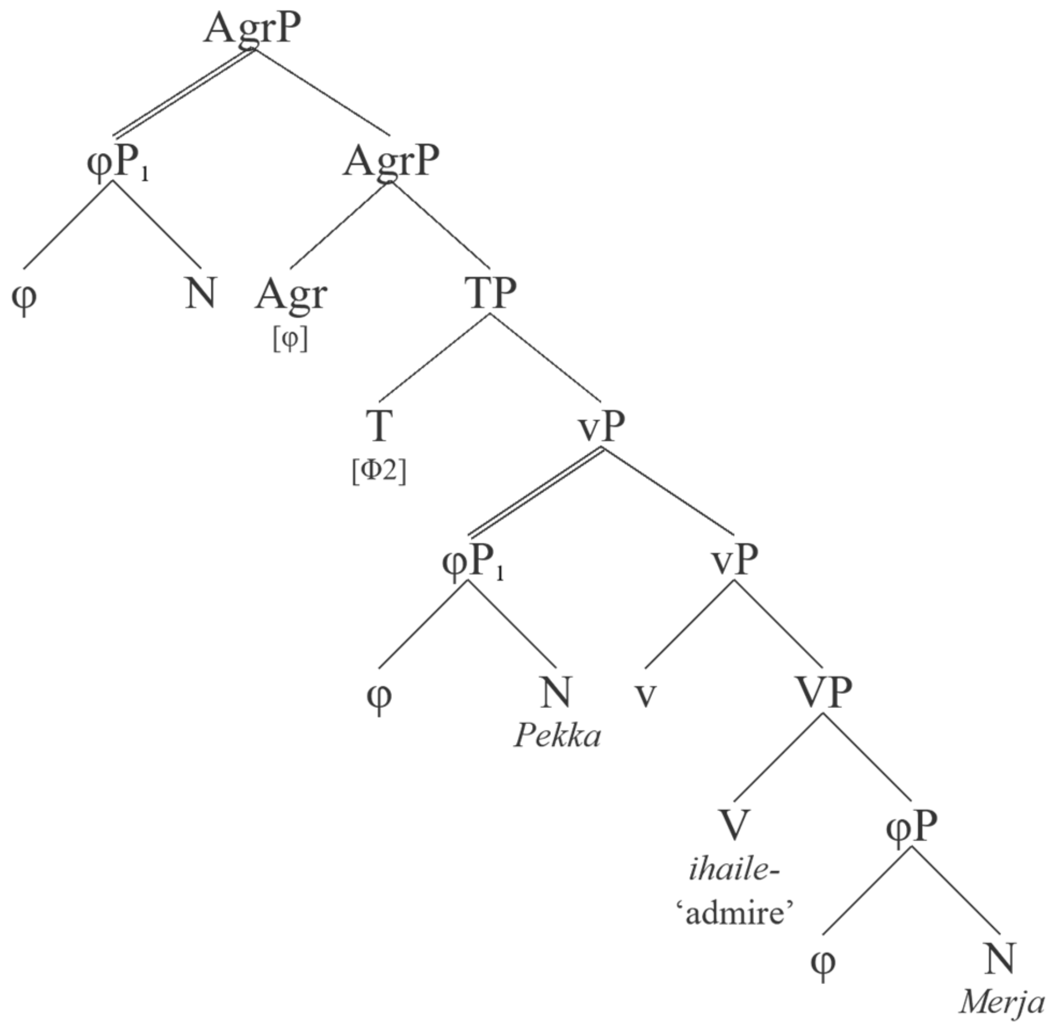
Beginning from the top, a distinction is first established between clauses which can and cannot project a subject (if subject projection is optional, we branch accordingly). If subject projection does not occur, the subject is determined by control (left side with features –EPP, –ΦPF). This class contains several nonfinite clauses (A-infinitive, MA-infinitive, E-infinitive, bare ESSA-infinitive and the VA-participle). If the subject is projected, then the next question is whether overt nonfinite agreement is possible; if it isn't, then the only option is to use an overt phrasal argument. These choices exhaust the options in languages where only finite verbs exhibit non-concordial agreement. Since Finnish too has agreementless nonfinite predicates, it includes but is not limited to the same contrast (e.g., rows 1–2, Table 3). If the nonfinite predicate can exhibit agreement but does not do so, then the subject must be expressed by means of an overt phrasal subject. This situation is exemplified by the VA-infinitive (row 3) and the TUA-infinitive (row 6) which project an overt phrasal subject if agreement is absent. This behavior was captured by Φ1. If agreement is present, the remaining question is whether an overt subject can occur redundantly. If not, we have an agreement-only nonfinite predicate (KSE-infinitive); if yes, then redundant subject identification is possible (ESSA-infinitive, MA-participle).[17] There was considerable variation between native speakers with respect to the grammaticality and/or acceptability judgments concerning redundant subject identification.

Once we have a model that calculates the dataset and provides the input sentences with correct semantic interpretations and plausible syntactic analyses, we can regard the resulting model as a baseline hypothesis and pose further questions, such as whether the model is able to calculate everything if challenged by a larger dataset. We could include binding, word order variations, deverbal nominals, more data concerning the participles, layered adverbials, selection by noun heads, and so on. It is also possible to include data from other languages, since the algorithm can change the speaker model based on the language it recognizes in the input sentence. This could force changes to the model, for example an extra CP to the VA-infinitive as proposed by Kiparsky (2019). Additional data concerning clause-internal scrambling could require projecting more elaborate information structural representations on the top of the structures assumed here, following Koskinen (1998) and Ylinärä (2018). It is important though to make sure that any possible revision calculates both the new and the old data, and that the demonstration is fully rigorous.

We can also ask if it is possible to make the model simpler. Is the feature system elucidated in Figure 8 as simple as possible? Could the phrase structure apparatus be replaced by a connectionist model or by dependency trees with a simpler structure? It goes without saying that all attempts at reducing the model into more primitive components must preserve observational, descriptive and explanatory adequacy and process at least the same dataset, preferably a larger one. Furthermore, everything must be provided in a complete unambiguous form so that the hypothesis can be tested rigorously and the results compared with the present approach.

There are several alternative hypotheses that can be explored by using the baseline model by making changes into the lexicon. Let us assume that the third person agreement suffix in the verb *ihaile-e* ‘admire-PRS.3SG’ maps to its own grammatical Agr^0 head and not to an inflectional feature bundle. We can create an experimental verb for this purpose and run the simulation. The sentence is *Pekka ihailee* Merjaa* (sentence #262), where *ihailee** denotes the new verb with the decomposition *admire*#v#T#Agr containing a separate Agr^0 head. The sentence is analyzed by the baseline model as shown in (62).

(62)



The agreement suffix was automatically expanded into its own Agr^0 head by the reversed mirror principle. Once we have made sure that reasonable output is generated, we can map all third person suffixes experimentally into a separate Agr^0 head (finite agreement with an additional finiteness feature, nonfinite agreement without) and run the simulation over the whole dataset. This experiment resulted in 40 errors in grammaticality judgments = 15% error rate. For example, the model wrongly accepts (63).

(63) *Pekka halus-i hän-en lähte-ä-nsä.
Pekka.NOM want-PRS.3SG he-GEN leave-A/INF-PX/3P
Inteded: 'Pekka wanted him to leave.'

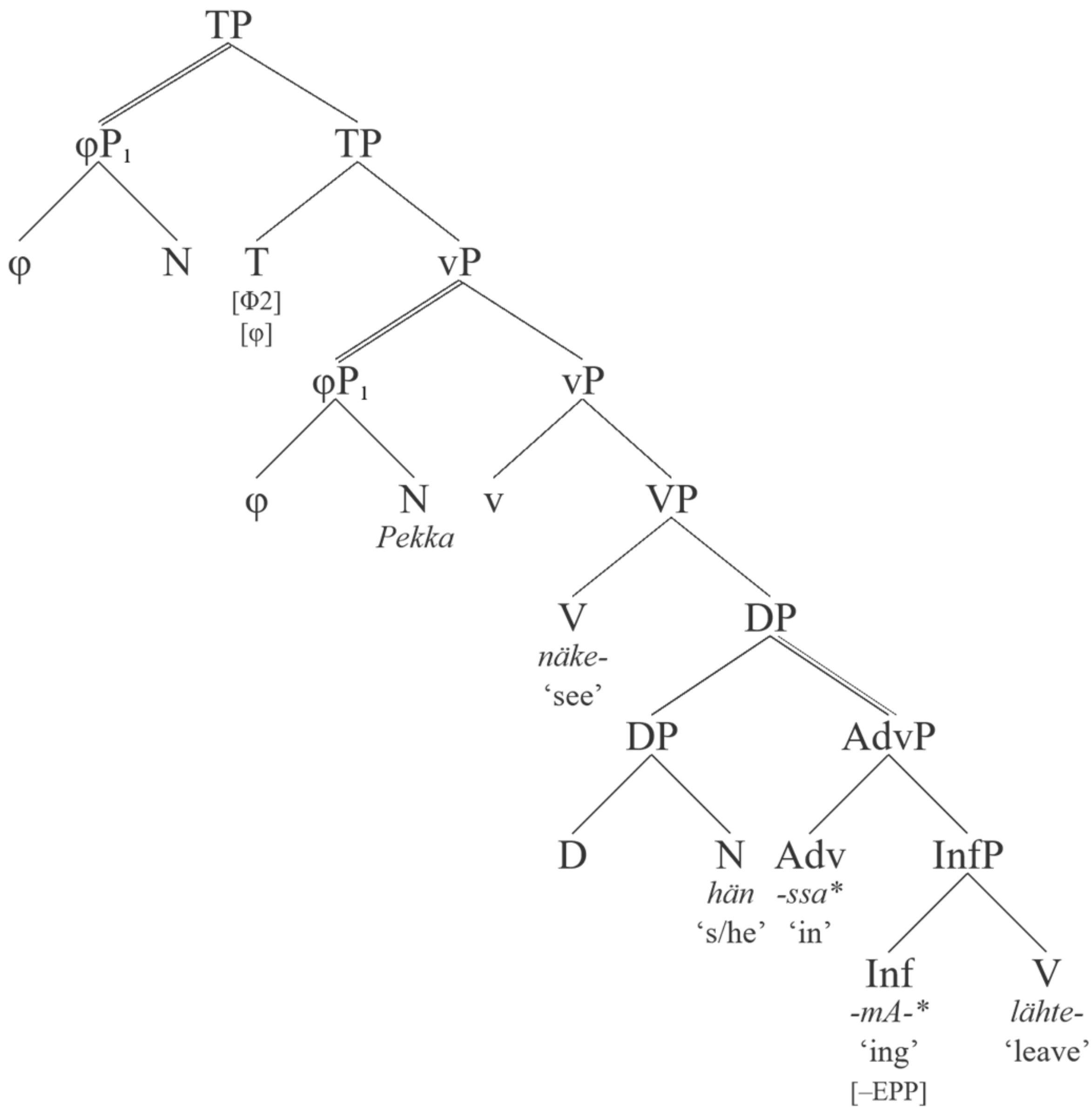
The problem is that the nonfinite agreement cluster projected an extra AgrP over the infinitive, no agreement was left for the infinitival head, and the feature conflict with –ΦPF no longer materializes. To fix this problem, we could add a rule which prevents Agr from selecting the infinitive and run the whole simulation anew. Thus, once we introduce Agr^0 into the theory its selection properties must be modelled and tested over the whole dataset. The issue is not whether all these errors can be fixed – they can be fixed because the implementation is written in a general-purpose programming language – but whether there is data that forces us to posit the more complex analysis.

Let us consider another hypothesis which decomposes the MA-infinitives into two morphemes, the MA-affix and a semantic case form. Let us assume, following Nikanne (1993), that semantic case forms are represented by covert prepositions such that '*lähte-mä-ssä**, for example, is decomposed as V + MA/inf + P(inessive). First we test the analysis with a single item (64).

(64) Pekka näk-i hän-et lähte-mä-ssä*. (#262)
Pekka.NOM see-PST.3SG he-ACC leave-MA-INE
'Pekka saw him leaving.'

After a few adjustments,[18] the model calculated (65) where P* and mA* designate the new experimental morphs.

(65)



The analysis is not implausible and closely resembles the one proposed by Vainikka (1989). The issue, however, is that it requires a special inessive preposition P* which is adjoined obligatorily into a low position, is excluded from many regular prepositional phrase positions (66a) and does not select for regular noun phrases (66b).

(66) a. Juhli-ssa oli hauskaa/ *Lähte-mä-ssä oli hauskaa

*party-*INE *was fun* *leave-*MA-INE *was fun*

'It was fun in the party.' Intended: 'It was fun to leave.'

b. nopea-ssa auto-ssa/ *nopea-ssa lähte-mä-ssä

fast-INE *car*-INE *fast*-INE *leave*-MA-INE

'in the fast car' Intended: 'in the fast leaving'

Once we are forced to create special infinitival adpositions, the question of whether this alternative is more elegant than the one that does not posit them becomes much less clear. Both models require special heads corresponding to the infinitival morphemes in the input.

7 CONCLUSIONS

Finnish nonfinite clauses were examined from the point of view of human information processing. Selection, control, syntactic role, the syntax of embedded subjects and agreement were calculated successfully. The model analyzed nonfinite sentences as truncated clause structures [α VP] with one functional layer α above the verb phrase. They can be viewed as reduced finite clauses such that the finite projections (e.g., force, finite T, mood, complex tense, finite negation) were replaced by just one projection α with more reduced feature content but with agreement and EPP. Nonfinite predicates were analyzed as bimorphemic verbs with the structure V#α (e.g., *lähte-vän* 'leave-VA/INF'). Their special properties were captured by relying on the lexical content of α. The proposed analysis can be contrasted with more complex hypotheses projecting several nominal and verbal functional projections above the verb phrase (Vainikka 1994, Koskinen 1998, Ylinärä 2018, Kiparsky 2019) and with the simplest possible (null) hypothesis which claims that the nonfinite clauses are regular noun, verb or adposition phrases (Vainikka 1989). If the present proposal is correct, the truth falls somewhere between these two extremes.

References

Baker, Mark. 1985. The Mirror Principle and Morphosyntactic Explanation. *Linguistic Inquiry*, 16, 373–415.

Brattico, Pauli. 2019. *Computational implementation of a linear phase parser. Framework and technical documentation (version 15.0)*, Pavia. https://github.com/pajubrat/parser-grammar/blob/master/docs/Old%20documentation/Computational%20implementation%20of%20a%20linear%20phase%20parser%20(version%2015).pdf

Brattico, Pauli. 2022. Predicate clefting and long head movement in Finnish. *Linguistic Inquiry*, 54(4), 663–692.

Brattico, Pauli. 2023. Across the board agreement in Finnish. *Ms*.

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.

Hakulinen, Auli & Fred Karlsson. 1979. *Nykysuomen lauseoppia* [On the grammar of modern Finnish]. Helsinki: SKS.

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

Holmberg, Anders, Urpo Nikanne, Irmeli Oraviita, Hannu Reime & Trod Trosterud. 1993. The Structure of INFL and the Finite Clause in Finnish. In Anders Holmberg & Urpo Nikanne (eds.), *Case and other functional categories in Finnish syntax*, 177–206. Berlin: Mouton de Gruyter.

Ikola, Osmo. 1974. *Lauseenvastikeoppia* [A study of nonfinite clauses]. Helsinki: SKS.

Julien, Marit. 2002. *Syntactic Heads and Word Formation*. Oxford: Oxford University Press.

Jussila, Aapo. 2020. *Suomen ei-finiittiset verbimuodot typologisesta näkökulmasta* [Finnish nonfinite forms from the typological perspective]. Master's Thesis, University of Helsinki.

Kayne, Richard. 1983. Connectedness. *Linguistic Inquiry*, 14.2, 223–249.

Kayne, Richard. 1984. *Connectedness and Binary Branching*. Berlin: De Gruyter Mouton.

Kiparsky, Paul. 2019. Notes on nonfinite clauses in Finnish. In Cleo Condoravdi & Tracy Holloway King (eds.), *Tokens of Meaning: Essays in Honor of Lauri Karttunen*. Stanford, CA: CSLI Publications.

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

Marr, David. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA.: MIT Press.

Mitchell, Erika. 1991. Evidence from Finnish for Pollock's theory of IP. *Linguistic Inquiry*, 22, 373–379.

Nikanne, Urpo. 1993. On assigning semantic cases in Finnish. In Anders Holmberg & Urpo Nikanne (eds.), *Case and other functional categories in Finnish syntax*, 75–87. Berlin: Mouton de Gruyter.

Phillips, Colin. 1996. *Order and structure*. Ph.D. dissertation, MIT.

Pollock, Jean-Yves. 1989. Verb movement, Universal Grammar and the structure of IP. *Linguistic Inquiry*, 20, 365–424.

Toivonen, Ida. 1995. *A study of Finnish infinitives*. Senior thesis, Brandeis University.

Vainikka, Anne. 1989. *Deriving Syntactic Representations in Finnish*. Ph.D. dissertation, University of Massachusetts Amherst.

Vainikka, Anne. 1994. Functional Projections in Finnish Non-Finite Constructions. *Penn Working Papers in Linguistics*, 2, 141–146.

Vainikka, Anne. 1995. Functional Projections in Finnish Non-Finite Constructions. *University of Pennsylvania Working Papers in Linguistics (PWPL)*, 2(1).

Vilkuna, M. 2000. *Suomen lauseopin perusteet* [Introduction to Finnish grammar]. Helsinki: Edita.

Visapää, L. 2008. *Infinitiivi ja sen infiniittisyys: Tutkimus suomen kielen itsenäisistä A-infinitiivikonstruktioista*. Ph.D. dissertation, University of Helsinki.

Visapää, L. 2022. Infinitives of affect and intersubjectivity: on the indexical interpretation of the Finnish independent infinitives. *Cognitive Linguistics*, 33(3), 521–551.

Wiik, K. 1981. Mika lauseenvastike on? *Virittäjä*, 85(1), 21–39.

Ylikoski, J. 2003. Defining non-finites: Action nominals, Converbs and Infinitives. *SKY Journal of Linguistics*, 16, 185–237.

Ylinärä, E. 2018. *Forme infinitive in finlandese*. Munchen: LINCOM.

---

[1] The source code as well as the raw input/output files are available at https://github.com/pajubrat/parser-grammar/. The raw data is in the folder */language data working directory/study-13-c-infinitivals*. The source code used in this study, and which should be used for replication, is contained in the branch *Finnish-infinitivals-(Study-13c)*. The main master branch contains the latest version of the model.

[2] Abbreviations: 1, 2, 3P = first, second and third person; A/INF = A-infinitive; ACC = accusative case; ABE = abessive case; ALL = allative case; ADE = adessive case; E/INF = E-infinitive; ESSA/INF = ESSA-infinitive; ILL = illative case; INE = inessive case; NOM = nominative case; GEN = genitive case; MA/INF = MA-infinitive, any form; MA.X/INF = MA-infinitive marked with one of the five semantic cases X = ABE, ADE, ELA, ILL or INE; KSE/INF = KSE-infinitive; PAR = partitive case; PL = plural; PRS = present tense; PST = past tense; PX = nonfinite agreement (possessive suffix); SG = singular; TUA/INF = TUA-infinitive; VA/INF = VA-infinitive.

[3] The VA-infinitive is sometimes referred to as the 'referative construction' (Vilkuna 2000:Chapter 9.5) or 'clausal complement infinitival' (Vainikka 1989). We use morphological forms as a basis for naming the Finnish nonfinite clauses examined in this article. The term 'VA-infinitive', for example, comes from the *-vAn* morph which characterizes this infinitival predicate. The approach is theory-neutral and makes glossing easier to use and understand.

[4] Finnish finite agreement must be distinguished from nonfinite agreement. The former occurs in connection with finite elements such as the negation, auxiliaries and finite verbs, the latter covers almost all of the remaining major lexical categories and is expressed by means of a special morph called the possessive suffix, glossed as PX. Concordial agreement forms a third agreement category.

[5] Sentence *kysy-i-n häne-ltä mitä teh-dä* 'ask-PST-1SG he-ELA [what.PAR do-A/INF]' contains a nonfinite clause headed by an interrogative operator and constitutes an exception to the claim presented in the main text. This exceptional pattern is only possible in connection with the A-infinitive and only concerns the interrogative operator: infinitival relativization, for example, is impossible (**suunnitelma jonka sinä teh-dä oli huono* 'plan which.ACC you.NOM make-A/INF was bad').

[6] Vainikka's work was preceded by Hakulinen and Karlsson (1979), the first serious syntactic analysis of Finnish as a whole and therefore also an important progenitor for the work discussed in this article. The work was based on the pre-GB-theoretical model, however, and will not be reviewed here.
[7] Kiparsky (2019) proposes another more recent variation of the same idea, thus according to his analysis the "functional syntactic structure of Finnish nonfinite clauses is a transparent reflection of the overt morphological makeup of their participial and infinitival lexical heads" (p. 22).
[8] Visapää (2008, 2022) shows that the A-infinitive (see Table 1) has non-elliptical standalone uses. The computational analysis proposed in this article does not rule out standalone uses of infinitives that occur in complement positions. Sentence (7) is marginally grammatical in elliptical contexts, for example when it represents an answer to the question 'what did Pekka knew he would do?'. This topic merits a study of its own and will be put aside here.
[9] The notions of complement and adjunct correlate with the notions of (object) argument and adverbial modifier, respectively, the latter which are sometimes used in the literature to capture the distinction discussed in the main text. The two notions correlate, but are not the same. We will focus on the formal complement/adjunct distinction in this study. The semantic notions of argument and modifier are composite properties and do not therefore function as primitives in the formal model proposed later in this article.
[10] We assume the overall information processing framework of Marr (1982). It decomposes the cognitive phenomenon of interest (e.g., vision, language) into three levels of explanation: computational, algorithmic and neural implementation. The first two are discussed in this article.
[11] An anonymous *NJL* reviewer disagreed with the author on whether the agreeing nonfinite predicate can co-occur with an overt phrasal subject in a sentence such as (i) *Pekka ilahtui hän-en lähti-essä-än* 'Pekka celebrated he-GEN leave-ESSA/INF-PX/3P'. Six additional speakers were consulted (2 linguists + 4 nonlinguists). Five speakers (incl. the author) considered (i) marginal or ungrammatical (call it grammar A), three (incl. the reviewer) grammatical (grammar B). The model proposed in this article allows one to represent both grammars (A = $\Phi 1$, B = $\Phi 2$). The sentence was marked grammatical for the purposes of this study and the speaker model used in the simulations instantiated B. To simulate a speaker with a different grammar, a corresponding speaker model must be selected. Whether this variation represents different grammars, noise or some other factor(s) must be established in a separate study, however. This issue concerned only a few sentences, specifically highlighted later in this article.
[12] Technically the antecedent/control algorithm uses an approach inspired by Kayne's connectedness analysis (Kayne 1983, 1984), where the idea is to establish an upward path from the predicate to the antecedent. We can imagine the predicate scanning for a suitable antecedent within the active syntactic working memory by exploring the phrase structure upwards/leftwards. See the supplementary §2.5.
[13] There are a few marginal exceptions to this generalization, namely expressions such as *te-i-n ruoka-a laste-n syö-dä* 'make-PST-1SG food-PAR children-GEN eat-A/INF' which to me feel old-fashioned, marginal, and which are not part of my standard spoken or written Finnish. If, however, we wanted to include them, then a speaker model must be used where the lexical features of A/Inf^0 are provided with the (possibly restricted) adjunction option.
[14] Adjunct attachment, which is licensed in the underlying algorithm by +ADV, creates regular geometrical left or right daughter constituents with the special property that they are pulled into a separate syntactic working memory and became invisible to many grammatical operations and dependencies (e.g., sisterhood, labelling, selection, reconstruction) in the hosting structure. They increase the dimensionality of the syntactic structure by creating connected parallel structures. For a more detailed description see the supplementary §2.6.
[15] The claim is refuted by examples such as *huomaa-ma-tta-si* 'notice-MA-ABE-3SG' which means 'without your noticing', as observed by an anonymous *NJL* reviewer. While true, this form cannot be used with a direct object: *sinä lähdit huomaa-ma-tta(*-si) minua* 'you left notice-MA-ABE-(*3SG) me'. The rule is not general: *sinä olit koko yön nukku-ma-tta(?*-si)* 'you were all night sleep-MA-ABE(-2SG)'. Perhaps the form is a frozen adverb. If so, its exceptional properties can be represented in the lexicon in connection with the individual lexical item.
[16] Whether the pronoun can co-occur with nonfinite agreement in this construction is subject to variation among native speakers. ~50 % of the speakers consulted for the purposes of this study judged

it ungrammatical (including the author), the rest grammatical. The difference can be represented by features Φ1 (former grammar) and Φ2 (latter grammar).

[17] Redundant subject identification is also possible in canonical finite clauses, nouns and adpositions which were not discussed in this study.

[18] One immediate problem was that the model was able to separate the adposition *-ssA* and the infinitival *lähte-mä-* from each other, grouped the former with the direct object, and then interpreted the sentence analogously to *Pekka istui* [$_{\text{PP}}$ *minut -ssA*] [$_{\text{InfP}}$ *laulama*] 'Pekka sat me near singing' where it analyzed the DO + P complex wrongly as a postposition comparable to *minua lähellä* 'I.PAR near'. To prevent this, the adjunction option was disabled from the infinitival head, the preposition was allowed to take it as a complement, and furthermore the PP was provided with the same adjunction options as the original MA-infinitival to generate the correct control dependencies.