Agreement and chains

Pauli Brattio, Cristiano Chesi, Balasz Suranyi[1]

Abstract

This article explores a previously undocumented relationship between secondary Ā-chains and agreement. We show by examining data from English, Finnish, Hungarian and Italian that a functional head with label K projects a second merge position for both A- and Ā-chains if some lexical items with K exhibit overt agreement. Since it is already well-known that there is a correlation between A-chains and agreement, these results suggest that agreement is connected to chain formation in a more general way. An analysis is proposed according to which agreement and A/Ā-chains are two sides of a more abstract edge mechanism. The hypothesis is tested by calculation against a linguistic dataset. Rigorous computational testing revealed that the full unification of scrambling/topicalization with the edge mechanism and A/Ā-chain formation presents a further challenge.

Keywords: EPP, phi-agreement, Agree, wh-movement, pied-piping, secondary wh-movement, Finnish, Hungarian, Italian

1 Introduction

Compare the following sentences from English and Finnish (1).[2]

(1)

a. [$_{AdvP}$ In order to travel [$_{PP}$ towards which city]] did John buy the tickets _? (*English*)

b. [$_{AdvP}$ [$_{PP}$ Mitä kaupunkia$_1$ kohti __$_1$]$_2$ matkustaakseen __$_2$]$_3$ Jari ost-i lipu-t__$_3$? (*Finnish*)
what city.PAR towards to.travel Jari.NOM buy-PST.3SG ticket-ACC.PL
'In order to travel towards which city did Jari buy the tickets?'

[1] This is currently (2025) an unpublished but relatively polished draft that was written and researched when the first and second authors were working at IUSS, Italy.

[2] Abbreviations: 0 = no agreement or default agreement; 1, 2, 3 = first, second and the third person, respectively; φ = number and person features; ACC = accusative case; ALL = allative case; DAT = dative case; DEF = definiteness; EF = edge feature; F = feminine; GEN = genitive case; IMPASS = impersonal passive verb form; KSE = KSE-infinitival, roughly 'in order to V'; MA = ma-infinitival (several forms); M = masculine; NOM = nominative case; NUT = past participle adjective; PAR = partitive case; POSS = Hungarian possessive; PL = plural; PRS = present tense; PRTCPL = participle; PST = past tense; PX = Finnish possessive suffix; Q = yes/no question particle; SG = singular; SUP = suppressive; uφ = uninterpretable φ-set; VA = va-infinitival or va-participle adjective

In Finnish (a Finno-Ugric language), the *wh*-element forms a secondary *wh*-chain inside the chain containing the criterial scope position (Huhmarniemi 2012). This is not possible in English (2).

(2) *[[Which city$_1$ towards __$_1$]$_2$ in order to travel __$_2$]$_3$ did John buy the tickets __$_3$?

The reason for the Finnish pattern, which applies throughout the language, remains unknown. We argue that the relevant crosslinguistic difference has to do with agreement. In Finnish, prepositions and adverbs, both domains for pied-piping and secondary Ā-chains, exhibit agreement in number and person (3), whereas in English or Italian neither does.

(3) *Finnish*

a. *Adpositions and agreement*

i.	minun	lähellä-ni,	ii.	sinun	lähellä-si,	iii.	hänen lähellä-än
	I.GEN	near-2SG		you.GEN	near-2SG		s/he.GEN near-3P
	‘near me’			‘near you’			‘near her/him’

b. *Adverbs and agreement*

i. Jari harjoittel-i [voittaa-kse-en kilpailu-n.]
Jari.NOM practice-PST.3SG win-KSE-3P competition-ACC
‘Jari practised in order to win a/the competition.’

ii. Minä harjoitte-i-n [voittaa-kse-ni kilpailu-n.]
I.NOM practice-PST-1SG win-KSE-1SG competition-ACC
‘I practiced in order to win a/the competition.’

We show that overt agreement in number and person characterizes every category that triggers secondary Ā-chains in Finnish and Hungarian, and none of the corresponding categories in English or Italian. This suggests that there is a connection between a head’s ability to project members for Ā-chains and its ability to agree with a local argument. We propose that agreement and the projection of specifiers are based on the same edge mechanism which relates heads to local noncomplement arguments; negative specification generates agreementless heads that merge with complements only. Because specifiers projected by the edge feature are not licensed thematically, they must be chain members, generating the inversion phenomenon (1)b.

The argument is organized in the following way. Section 2 provides the required empirical and theoretical background concerning the phenomena and the languages targeted for modelling. It defines key terminology used throughout the paper. Section 3 cites the empirical data in support of the correlation between secondary Ā-chains and agreement. We discuss adpositions, four types of adverbials, participle adjectives, noun phrases, relative clauses and finite clauses, and cite data from Finnish, English, Italian and Hungarian. An analysis, which is described first in informal English, is provided in Section 3.8. Section 4 revisits the data

by replicating it with a fully formal Python implementation of the analysis. Python is a general-purpose programming language that allows the researcher to define the analysis rigorously and use the formalization to calculate/replicate the data. This section begins with the formalization (Section 4.1), elucidates the computational methodology and data processing (Section 4.2) and reports the results (Section 4.3). Section 5 discusses certain further issues and problems that emerged during the computational study. Section 6 contains the conclusions.

Almost all empirical examples cited in this article are provided with a numerical identifier #N which refers to the corresponding sentence in the dataset processed by the formalized analysis proposed in Section 4 of this article. A "corresponding sentence" is either the same sentence as the one appearing in this article, a syntactic equivalent, or a set of sentences illustrating the syntactic phenomenon more broadly. These numbers are not relevant when examining the data as such, but useful when inspecting how the analysis derives them. Section 4.2.1 explains how to use these numbers.

2 Background. W*h*-movement and agreement in Finnish

2.1 Secondary wh-chains in pied piping

Exactly one *wh*-pronoun must occur at the left edge of an interrogative clause in Finnish. If a larger phrase containing the *wh*-feature is pied-piped, the word carrying the *wh*-feature must occur at the edge of the pied-piped phrase (4)a-b. If the *wh*-operator is not at the edge of the pied-piped phrase, only the echo-question is available (4)(c).

(4)

a. Jari asu-u [lähellä Pariisi-a.] (Sentence #1 in the dataset)
Jari.NOM live-PRS.3SG near Paris.PAR
'Jari lived near Paris.'

b. [Mitä$_1$ kaupuki-a lähellä __$_1$]$_2$ Jari asu-u __$_2$? (#65)
what.PAR city-PAR near Jari.NOM live-PRS.3SG
'Near which city does Jari live?'

c *[lähellä mitä kaupunki-a]$_1$ Jari asu-u __$_1$. (#64)
near which.PAR city-PAR Jari.NOM live-PRS.3SG
Intended: 'Near which city does Jari live?'

Finnish therefore obeys Heck's (2008; 2009) edge generalization: pied-piping is possible only if the *wh*-feature occurs at the edge of the pied-piped phrase. Secondary *wh*-chains inside pied-piped phrases follow the same regimen, generating recursively nested Ā-chains (5)a. If any of these chains are missing, the sentence reads as an echo question (5)b.

(5)

a. [[Mitä$_1$ kaupunki-a kohti __$_1$]$_2$matkustaakseen __$_2$]$_3$ Jari ost-i lipu-t __$_3$? (#89)
which.PAR city-PAR towards to.travel Jari.NOM buy-PST.3SG ticket-PL.ACC
'In order to travel towards which city did Jari buy the tickets?'

b. [[Matkustakseen kohti [mitä kaupunki-a]] Jari ost-i lipu-t __.(#87)
to.travel towards which.PAR city-PAR Jari.NOM buy-PST.3SG ticket-PL.ACC
(Only the echo-reading)

The Finnish data were first documented and systematized by Huhmarniemi (2012). She demonstrated that the edge generalization applies to determiner/noun phrases, adposition phrases (6a), several types of adverbials (6b) and participle adjectives (6c). Relative clauses and interrogative clauses (6d) satisfy the edge generalization but cannot be pied-pied.

(6)

a. [$_{PP}$[$_{DP}$ Mitä kaupunki-a]$_1$ lähellä __$_1$] $_2$ Jari asu-u __$_2$? (#65)
which.PAR city-PAR near Jari.NOM live-PRS.3SG
'Near which city does Jari live?'

b. [[$_{PP}$ Mitä kaupunki-a kohti]$_1$ matkustaa-kse-en __$_1$] Jari ost-i lipu-t? (#89)
what.PAR city-PAR towards travel-KSE-3P Jari.NOM buy-PST.3SG tickets-ACC.PL
'In order to travel towards which city did Jari bought the tickets?'

c. [Mitä$_1$ tutki-va __$_1$] professori sa-i vira-n? (#411)
what.PAR research-PRTCPL professor.NOM get-PST.3SG tenure-ACC
'A professor that studies what got the position?'

d. Mitä$_1$ professori tutk-i __$_1$? (#55, 58)
what.PAR professor.NOM study-PST.3SG
'What did the professor study?'

e. ratkaisu, [jota$_1$ kohti __$_1$]$_2$ professori eten-i __$_2$ (#108-110)
solution which.PAR towards professor.NOM progress-PST.3SG
'a solution towards which the professor progressed'

Secondary Ā-chains satisfy the standard, defining properties of Ā-chains (Huhmarniemi and Brattico 2013). The pattern holds for *wh*-chains and chains headed by relative pronouns, but it does not hold for all operators. The Finnish yes/no operator pied-pipes material to the scope position but does not need to occur (although it can occur) at the edge of the pied-piped phrase (7).

(7)

a. [Lähellä sitä puisto-a-ko]$_1$ Pekka asu-u __$_1$? (#141)
near this.PAR park-PAR-Q Pekka.NOM live-PRS.3SG

‘Is it near that <u>park</u> (and not the river) that Pekka lives?’

b. [tätä-kö puisto-a_2 lähellä $\_\_{}_2]_1$ Pekka asu-u $\_\_{}_1$? (#142)
this.PAR-Q park-PAR near Pekka.NOM live-PRS.3SG
‘Is it near <u>this</u> (and not the other) park that Pekka lives?’

c. *Pekka asu-u [lähellä tätä kaupunki-a-ko?] (#132-140)
Pekka.NOM live-PRS.3SG near this.PAR city-PAR-Q
Intended: ‘Does Pekka live near this <u>city</u>?’

We distinguish *strong operators*, which fall under edge generalization, from *weak operators*, which do not. Thus defined, Finnish interrogative pronouns are strong operators, while the yes/no operator is a weak operator. We also distinguish *primary Ā-chains* from *secondary Ā-chains* such that the former are involved in scope-marking and clause typing, whereas the latter have no semantic- or scope-related motivation. Weak operators are exempted from the edge generalization when it comes to secondary Ā-chains, but they are required to occur at an appropriate scope position if they lack the in situ option.

Nonoperators can occupy any position operators can occupy with a few exceptions discussed later in this article. They trigger chains, but are not associated with changes in force and/or clause typing and, as one would expect, do not create operator readings (8).[3],[4]

(8)

a. Seine virta-a [kohti Pariisi-a]. (#9)
Seine.NOM flow-PRS.3SG towards Paris-PAR
‘Seine flows towards Paris.’

b. Seine virta-a [Pariisi-a_1 kohti $\_\_{}_1$]. (#10)
Seine.NOM flow-PRS.3SG Paris-PAR towards
‘Seine flows towards Paris.’

c. [kohti Pariisi-$a]_1$ Seine virta-a $\_\_{}_1$. (#11)
towards Paris-PAR Seine.NOM flow-PRS.3SG
‘Seine flows towards Paris.’

d. [Pariisi-a_1 kohti $\_\_{}_1]_2$ Seine virta-a $\_\_{}_2$. (#12)
Paris-PAR towards Seine.NOM flow-PRS.3SG
‘Seine flows towards Paris.’

---

[3] We assume that the nonoperators under consideration in the main text do not bear prosodic emphasis; if they do, then they count as contrastive focus operators and must be analyzed differently. They are discussed in Section 4.3.

[4] The term “trigger” used here and throughout the text has a precise meaning within the formal implementation we provide later in this article. It refers to a formal property of the underlying phrase structure representation which activates the chain formation algorithm.

Comparison between (8) and (4)-(7) suggests that Ā-chains are at stake. The claim that nonoperators trigger Ā-chains should not be viewed as uncontroversial, however, and indeed this issue turns out to be nontrivial. Descriptively, though, the word orders and chains in (8) by and large mirror those made by operators and obey the same or at the very least very similar limitations. Any position occupied by *Paris* in (8) can be occupied by an operator, and vice versa. Operators and nonoperators differ in how they interact with the final scope position: sentence-initial nonoperators are interpreted as topics, while sentence-initial operators create operator readings.[5]

2.2 Agreement in Finnish

Agreement in Finnish canonical finite clauses holds between the finite element (finite verb, auxiliary or the auxiliary-like negation) and the grammatical subject. Standard finite subject-verb agreement is exemplified by almost all sentence examples cited in this article. Several nonfinite words also exhibit agreement in Finnish. Consider the behavior of two adpositions *lähellä* 'near' and *kohti* 'towards'.

(9) a. lähellä minu-a (#280)
near I-PAR
'near me'

b. minun$_1$ lähellä(ni) __$_1$ (#284)
I.GEN near(1SG)
'near me'

c. minu-a$_1$ lähellä __$_1$ (#281)
I.PAR near
'near me'

d. *lähellä(-ni) minun (#18)
near(-1SG) I.GEN

e. *lähellä-ni minua (#287)
near-1SG I.PAR

f. *minua lähellä(-ni) (#288)
I.par near(-1SG)

g. lähellä-ni (#467)
near-1SG
'near me'

h. lähellä (#286)
near
'near'

(10) a. kohti minu-a (#268)
towards I-PAR
'towards me'

b. *minun kohti(-ni) (#273)
I.GEN towards(-1SG)

c. minu-a$_1$ kohti __$_1$ (#224)
I-PAR towards
'towards me'

[5] Finnish preverbal syntax is traditionally partitioned into two segments, the operator segment and the topic segment (Vilkuna 1989; Vilkuna 1995). Any constituent that is moved to or generated at the operator position must be an operator or be interpreted as one; all fronted nonoperators must therefore be in the topic area, which results in the topic interpretations mentioned in the text.

The adposition 'near' exhibits agreement in number and person with its argument, whereas the adposition *kohti* 'towards' does not.[6] A number of other grammatical properties are implied, shown by the data above. First, the argument of an agreeing adposition must take the genitive case; partitive or any other case is ungrammatical. Second, the genitive argument, which may also exhibit agreement with the adposition, must precede the adposition for most adpositions, while partitive arguments can precede and follow it. Third, both the agreeing and the nonagreeing forms can occur alone.

3 Agreement and Ā-chains

3.1 The Agreement-Remerge Correlation

The data above leads one to suspect that there might be a correlation between agreement and secondary Ā-chains. Here we show that the correlation exists and takes the form (11).

(11) *Agreement-Remerge Correlation (ARC)*

Any lexical item in a major lexical category K licenses a nonthematic/non-criterial specifier position for A/Ā-chains (in language L) if and only if at least some lexical items in K agree (in L).

Principle (11) is descriptive and is meant to capture the apparent correlation between Ā-chains and agreement illustrated in Section 1. For the time being we can use the following informal definitions for the terms occurring in this principle. The specifier of head H refers to any noncomplement phrase inside HP (ignoring the status of adjuncts); all positions created by secondary Ā-chains in Section 1 are of this type. A "nonthematic/non-criterial specifier position" is one that is associated neither with a canonical (base-generated) thematic interpretation nor with a scope interpretation; we put aside the status of discourse (topic/focus) positions for now and return to them later. We never see movement into a thematic position in our data, and criterial positions such as SpecCP are not associated with overt agreement, so they are excluded (for now) on such basis. The notions of A- and Ā-chains retain their usual meanings. Any individual movement step that brings the operator into a criterial scope position is a *primary chain*. The secondary Ā-chains exemplified by the data in Section 1 are not associated with scope readings; they appear to be idle with respect to semantic interpretation. Agreement in (11) denotes systematic and overt covariation in phi-features (complete or partial) between a head and a full argument DP. Phi-concord, pronominal clitic adjunction, frozen default agreement and all forms of covert agreement are excluded. Let us next look at the data we think supports (11).

---

[6] Nonfinite agreement is expressed by the possessive suffix (PX) which reflects the number and person properties (excluding number in the third person) of the local argument pronoun. The whole paradigm is *minun lähellä-ni* 'I.GEN near-1SG', *sinun lähellä-si* 'you.GEN near-2SG', *hänen lähellä-än* 'he.GEN near-3P', *meidän lähellä-mme* 'we.GEN near-1PL', *teidän lähellä-nne* 'you.GEN near.2PL', *heidän lähellä-än* 'they.GEN near-3P'. Nonhuman pronouns (**sen lähellä-än* 'it.GEN near-3P'), full DPs (**koiran lähellä-än* 'dog.GEN near-3P) and relative and interrogative pronouns do not induce possessive agreement. See (Kanerva 1987; van Steenbergen 1987; Vainikka 1989; Toivonen 2000; Huhmarniemi and Brattico 2015; Brattico 2017).

3.2 Adpositions

As already mentioned, the Finnish adposition head projects a specifier position for an operator (12)a and can agree (12)b with a genitive argument preceding the adposition. Example (12)c shows the canonical word order.

(12)

a. [PPmitä$_1$ lähellä __$_1$]$_2$ Jari asu-u __$_2$? (#65)
what.PAR near Jari.NOM live-PRS.3SG
'Near what does Jari live?'

b. minun lähellä(-ni) (#222)
I.GEN near(-1SG)
'near me', lit. 'my near'

c. Jari asu-u [lähellä Pariisi-a.] (#1)
Jari.NOM live-PRS.3SG near Paris-PAR
'Jari lives near Paris.'

Notice that the adposition does not agree with the operator; it agrees with a genitive-marked argument, if any. Principle (11) does not require operators to trigger agreement; it requires agreement to be possible in the major lexical class projecting the specifier position for the operator. If the genitive-marked argument itself is an operator, it remains at the specifier position and agrees with the adposition.[7] English (13) and Italian (14) prepositions and adverbial participles, on the other hand, exhibit neither agreement nor project specifiers.

(13)

a. John lives near Paris. (#25)
b. *John lives [Paris$_1$ near __$_1$.] (#27)
c. *[Near Paris] John lives. (#26)
d. *[Paris$_1$ near __$_1$]$_2$ John lives __$_2$. (#26)
e. *[What$_1$ near __$_1$]$_2$ John lives __$_2$ (#81)
f. *John lives near#[3sg] Paris.[8] (#305-308)

(14) *Italian nonoperators*

a. Giuseppe vive vicino a Parigi. (#309)
Giuseppe live.PRS.3SG near at Paris
'Giuseppe lives near Paris.'

b. *Giuseppe vive vicino Parigi$_1$ a __$_1$. (#312)

---

[7] An example: *Minun-ko lähellä-ni Pekka asui* 'I.GEN-Q near-1SG Pekka.NOM lived?' which means 'was it near me (and not you) that Pekka lived?'. The *-ko* suffix represents the Finnish yes/no operator, glossed as Q. Interrogative and relative pronouns do not trigger agreement with the noun head.

[8] The symbol #[3sg] forces the formal model presented later in this article to insert the third person agreement features inside the English adposition. No such adposition exists in English; the purpose was to examine what would happen in our model were such Finnish-type forms generated in English.

c. *Giuseppe vive [a Parigi]$_1$ vicino __$_1$. (#313)
d. *Giuseppe vive [Parigi$_1$ a__$_1$]$_2$ vicino __$_2$. (#314)
e. *Giuseppe vive vicino Parigi$_1$ a#[3sg] __$_1$. (#315)

(15) *Italian operators*

a. Giuseppe vive vicino a cosa. (#310)
Giuseppe live.PRS.3SG near at what
'Giuseppe lives near what? (echo-reading only)'
b. [Vicino a cosa]$_1$ Giuseppe vive __$_1$? (#311)
c. *[[Cosa$_1$ a __$_1$]$_2$ vicino __$_2$]$_3$ Giuseppe live __$_3$? (#318)

Adpositional agreement is also attested in Hungarian (16), a Finno-Ugric language distantly related to Finnish.

(16)

a. én mellett-em
I.NOM near-1SG
'near me'
b. mi mellett-ünk (#319, 320)(É. Kiss, 2002, p. 186, ex. 13)
we.NOM near-1PL
'near us'

A nominative pronominal argument agrees with the adposition and is located at its specifier position. Nonagreeing adpositions, on the other hand, permit an argument on either side of the adposition (17).

(17)

a. Út nem vezet [át [az erdőn.]] (#321)
path not leads through the forest.SUP
'No path leads through the forest.'
b. Út nem vezet [[az erdőn]$_1$ át __$_1$.] (#322)
path not leads the forest through
'No path leads through the forest.'

When the NP argument is a *wh*-phrase, it can only occur at the left of adposition within the PP (18), exactly as in Finnish:

(18)

a. [[Melyik erdőn]$_1$ át __$_1$]$_2$ nem vezet út __$_2$? (#323)
which forest.SUP through not leads path.NOM
'Through which forest does no path lead?'
b. [Át [$_{DP}$ melyik erdőn]]$_1$ nem vezet út __$_1$? (#324)

through which forest.SUP not runs path.NOM
(Only echo-interpretation)

3.3 Adverbials

Finnish adverbial participles exhibit (11). Example (19) shows that the adverbial participle agrees and obeys the edge generalization. Leaving the interrogative pronoun at any other legitimate position (marked by __) creates an echo-interpretation.

(19)

a. [[Mitä kaupunki-a$_1$ kohti __$_1$]$_2$ matkustaa-kse-en __$_2$]$_3$ Jari ost-i lipu-t __$_3$?
what.PAR city-PAR towards travel-KSE-3SG Jari.NOM buy-PST.3SG ticke-ACC.PL
'In order to travel towards which city did Jari buy the tickets?'

b. [[Mitä kaupunki-a$_1$ kohti __$_1$]$_2$ matkustaa-kse-ni __$_2$]$_3$ minä ost-i-n lipu-t __$_3$?
which.PAR city-PAR towards travel-KSE-1SG I.NOM buy-PST-1SG ticket-ACC.PL
'In order to travel towards which city did I buy the tickets?'

The canonical position of the pied-piping PP is illustrated in (20)a. A version where a nonoperator PP is moved optionally to the edge of the adverbial clause is also possible (20)b. This pattern replicates the data from adpositions and shows that nonoperators can occupy (all else being equal) the same positions as operators.

(20)

a. Minä ost-i-n lipu-t [matkustaa-kse-ni [kohti Pariisi-a.]] (#29)
I.NOM buy-PST-1SG ticket-PL.ACC travel-KSE-1SG towards Paris-PAR
'I bought the tickets in order to travel towards Paris.'

b. Minä ost-i-n lipu-t [[kohti Pariisi-a]$_1$ matkustaa-kse-ni __$_1$.] (#31)
I.NOM buy-PST-1SG ticket-PL.ACC towards Paris-PAR travel-KSE-1SG
'I bought the tickets in order to travel towards Paris.'

The specifier position can be filled in by a phrase independent of its label (21), but there can be only one filler (22). These adverbials, therefore, project a specifier position that can hold a member of exactly one chain.

(21)

a. Matkust-i-n Pariisiin [auto-n$_1$ ostaa-kse-ni __$_1$]. (#352)
travelled-PST-1SG to.Paris car-ACC buy-KSE-1SG
'I travelled to Paris in order to buy a car.'

b. Harjoittel-i-n joka ilta [nopeasti$_1$ juosta-kse-ni __$_1$].

practised-PST-1SG every evening fast run-KSE-1SG
'I practised every evening in order to run fast.'

(22)

a. *?Matkust-i-n Pariisiin [[DP auto-n]$_1$ [PP ystävältä]$_2$ ostaa-kse-ni _$_1$ _$_2$]. (#367-372)
travelled-PST-1SG to.Paris car-ACC from.friend buy-KSE-1SG
'I travelled to Paris in order to buy a car from a friend.'

b. *?Matkust-i-n Pariisiin [[PP ystävältä]$_2$[DP auto-n]$_1$ ostaa-kse-ni _$_1$ _$_2$].
travelled-PST-1SG to.Paris from.friend car-ACC buy-KSE-1G

In English (23) and Italian (24), adverbials neither license specifier positions nor agree with anything (adverbials in these languages are morphologically invariant).

(23)

a. Seine reaches the ocean [by flowing towards Paris]. (#373)
b. *Seine reaches the ocean [Paris$_1$ by flowing towards __$_1$]. (#377)
c. *Seine reaches the ocean [[towards Paris]$_1$ by flowing __$_1$]. (#378)
d. [By flowing towards what]$_1$ did Seine reach the ocean __$_1$? (#375)
e. *Seine reaches the ocean [by flowing#[3sg] towards Paris]. (#376)

(24)

a. Giuseppe ha attirate attenzione [per aver vinto Luisa]. (#387)
Giuseppe has attracted attention for having win.PST.PRTCPL Luisa
b. [per aver vinto cosa] Giuseppe ha attirate attenzione? (#207)
for having win what Giuseppe has attract attention
c. *[Cosa$_1$ per aver vinto __$_1$]$_2$ Giuseppe ha attirato attenzione __$_2$? (#389)
what for having won Giuseppe has attracted attention
Intended: 'Having won what did Giuseppe attracted attention?'

The Finnish KSE-infinitival cited above is only one of the Finnish adverbials. Finnish has nine nonfinite clause types (also called "infinitives" and "infinitival clauses") formed from a verbal stem and an infinitival suffix that project adverbial clauses. The edge generalization holds for all participial adverbials (Huhmarniemi 2012). It is demonstrated below for four infinitival predicates (26)-(28) selected for this study to represent the range of grammatical options that were found to be relevant to our analysis.

(25)

a. Pekka väsy-i tehd-essä-än talo-a. (#331-340)
Pekka.NOM get.tired-PST.3SG make-ESSA/INF-3P home-PAR
'Pekka got tired while making a house.'

b. *Pekka väsy-i tehd-essä-än mitä. (#333)
Pekka.NOM get.tired-PST.3SG make-ESSA/INF-3P what.PAR
Intended: ‘While making what house did Pekka get tired?’

c. [Mitä$_1$ tehd-essä-än __$_1$]$_2$ Pekka väsy-i __$_2$? (#334)
what.PAR make-ESSA/INF-3P Pekka.NOM get-tired-pst.3SG
‘While making what house did Pekka get tired?’

(26)

a. Pekka läht-i teh-tyä-än talo-a. (#341-350)
Pekka.NOM leave-PST.3SG make-TUA/INF-3P house-PAR
‘Pekka left after building (a little bit of the) house.’

b. *Pekka läht-i teh-tyä-än mitä. (#347)
Pekka.NOM leave-PST.3SG make-TUA/INF-3P house.PAR
Intended: ‘After making what did Pekka leave?’

c. [Mitä$_1$ teh-tyä-än __$_1$]$_2$ Pekka läht-i __$_2$? (#344)
what.PAR make-ESSA/INF-3P Pekka.NOM leave-PST.3SG
‘After making what did Pekka leave?’

(27)

a. Pekka läht-i tehdä-ksee-en talo-a. (#351-356)
Pekka.NOM leave-PST.3SG make-KSE/INF-3P house-PAR
‘Pekka left in order to build (a little bit of the) house.’

b. *Pekka läht-i tehdä-ksee-en mitä. (#353)
Pekka.NOM leave-PST.3SG make-KSE/INF-3P what-PAR
Intended: ‘In order to do what did Pekka leave?’

c. [mitä$_1$ tehdä-kse-en __$_1$]$_2$ Pekka läht-i __$_2$? (#354)
what.PAR make-KSE/INF-3P Pekka.NOM leave-PST.3SG
‘In order to do what did Pekka leave?’

(28)

a. Pekka läht-i teke-mättä talo-a. (#357-366)
Pekka.NOM leave-PST.3SG make-MA/INF house-PAR
‘Pekka left without building the house.’

b. *Pekka läht-i teke-mättä mitä? (#359)
Pekka.NOM leave-PST.3SG make-MA/INF what.PAR
Intended: ‘Without doing what did Pekka leave?’

c. [mitä$_1$ teke-mättä __$_1$]$_2$ Pekka läht-i __$_2$? (#360)

what.PAR make-MA/INF Pekka.NOM leave-PST.3SG
'Without doing what did Pekka leave?'

The relevant properties of these constructions that we would like to note here are the following. The KSE-adverbial (27) must exhibit agreement under all grammatically possible contexts, while it cannot have an overt subject. This creates a class where agreement is obligatory, but an overt specifier subject is ungrammatical. The TUA-adverbial (26) must exhibit either agreement or have an overt subject, but not both. The ESSA-adverbial (25) exhibits agreement and/or projects a subject, each of them optionally, but never both. The MA-infinitival (28) never exhibits agreement nor an overt subject, and thus requires control by a main clause argument. It behaves like the English and Italian adpositions and adverbs. When an adverbial clause projects an overt subject or agreement, it identifies a thematic subject of the adverbial clause; otherwise control is relinquished to a main clause participant, resulting in subject or object control, depending on the main verb and the infinitival predicate.

Let us consider what happens when the adverbial has its own subject while we try to satisfy the edge generalization by pied-piping a direct object operator to its edge. Sentence (29)a illustrates the canonical word order, (29)b tests pied-piping. Example (29)c shows that pied-piping is possible if the embedded thematic subject is absent.

(29)

a. Lapsi nukaht-i [isä-n luet-tua iltasatu-a]. (#349)
child.NOM fall.asleep-PST.3SG father-GEN read-TUA/INF good.night.story-PAR
'The child fell asleep after (her) father read a good night story.'

b. *[Mitä$_1$ isä-n luet-tua __$_1$] $_2$ lapsi nukaht-i __$_2$? (#370)
what.PAR father-GEN read-TUA/INF child.NOM fall.asleep-PST.3SG
Intended: After (her) father read what did the child fell asleep?'

c. [Mitä$_1$ luet-tua-an __$_1$]$_2$ lapsi nukaht-i __$_2$? (#344)
what.PAR read-TUA/INF-3SG child.NOM fall.asleep-PST.3SG
'After reading what did the child fell asleep?'

The presence of the subject blocks chain formation and pied-piping. These data suggest that the embedded subject occupies the same specifier position as the operator, making the subject position a "mixed" A- and Ā-position that sustains both subjects (29)a and operators (29)c but not both simultaneously (29)b. This feature will become one of the key assumptions of our unified chain formation mechanism.

3.4 Determiner phrases

Finnish noun heads can agree with a pronominal possessor (30).

(30) a. ne kolme minun ystävä-ä(-ni) (#393)
those three I.GEN friend.SG-PAR(-1SG)
'those three friends of mine'

b. ne kolme sinun ystävä-ä(-si) (#394)
those three you.GEN friend.SG-PAR(-2SG)
'those three friends of yours'

The position of the possessive argument within the noun phrase (or DP) is free (see Gröndahl 2015)(31).

(31) (?Pekan) ne (Pekan) kaikki (Pekan) kolme (Pekan) hyvää (Pekan) ystävää
Pekka.GEN those Pekka.GEN all three good friends
'all those three good friends of Pekka'

W*h*-elements must, as predicted by (11), occupy the specifier position (32).

(32)

a. kenen$_1$ kolme __$_1$ ystävä-ä
who.GEN three friend-PAR
'whose three friends'

b. *ne kolme kenen ystävä-ä (#395, 396)
those three who.GEN friend-PAR
'those three whose friends'

Moreover, filling in the edge position of the noun phrase is not sensitive to the category of the operator (33).

(33)

a. [$_{DP}$ Miten monta ystävä-ä] sinulla o-n?
how many friend-PAR you be-PRS-3SG
'How many friends do you have?'

b. [$_{DP}$ Millainen auto] o-n paras?
what.kind.of car be-PRS.3SG best
'What type of car is the best?'

c. [$_{DP}$ Tämä-kö auto] o-n rikki?
this-Q car be-PRS.3SG broken
'Is it this (and not that) car that is broken?'

These constructions satisfy the edge generalization and exhibit agreement. In Hungarian, as in Finnish, noun heads can agree in number and person with a local non-case-marked pronominal possessor (34).[9]

[9] The Hungarian dative possessor can occur higher than both the demonstrative pronoun and the definite article, as in *Jánosnak ez a könyve* 'John.DAT this the book'. Hungarian has two possessor constructions, one where the possessor is in the dative and appears in the highest specifier position inside the noun phrase, as in

(34) (É. Kiss, 2002: 158, ex. 16) (#401, 404)

a.	az	én	diák-ja-i-m		a	mi	diák-ja-i-nk
	the	I	student-POSS-PL-1SG		the	we	student-POSS-PL-1PL
	‘my students’				‘our students’		
b.	a	te	diák-ja-i-d		a	ti	diák-ja-i-tok
	the	you	student-POSS-PL-2SG		the	you	student-POSS-PL-2PL
	‘your students’				‘your students’		
c.	az	ő	diák-ja-i-0		az	ő	diák-ja-i-k
	the	(s)he	student-POSS-PL-3SG		the	(s)he	student-POSS-PL-3PL
	‘his students’				‘their students’		

Possessor *wh*-elements must occupy the edge position (35) (Szabolcsi 1994).

(35)

a.	ki-nek	a	könyv-e (#402)
	who-DAT	the	book-POSS
	’Whose book?’		
b.	*a	ki	könyv-e (#403)
	the	who	book-POSS

In English, possessor arguments do not agree with the noun head, but they occur canonically at the the edge, which makes testing ARC a challenge. Furthermore, if the possessor is base-generated at the edge, the construction is irrelevant to (11).

3.5 Participle adjectives

Example (36) shows that the Finnish prenominal participle adjective phrase (glossed as AP based on its distribution) projects a specifier position and exhibits agreement with the thematic argument.

(36)

a.	[DP	se	[AP	minun	Jarille	anta-ma(-ni)]	lahja]	men-i	rikki. (#410)
		that		I.GEN	to.Jari	give-MA(-1SG)	present	go-PST.3SG	broken
	‘The/that present that I have given to Jari broke down.’								

b.	[DP [AP	Kenelle	anta-ma-ni]	lahja]	men-i	rikki? (#411)
		to.who	give-MA-1SG	present	go-PST.3SG	broken

János-nak a könyve ‘John-DAT the book-POSS’, and another where the possessor takes the nominative case and occurs below the determiner (Szabolcsi 1983; Szabolcsi 1994). The choice between non-case-marked and dative possessor construction depends in part on the definiteness of the DP. For discussion, see (Szabolcsi 1994; Dikken 1999; Kiss 2002).

‘The present that I gave to who broke down?’

c. *[$_{DP}$ [$_{AP}$ Minun kenelle anta-ma-ni] lahja] men-i rikki? (#412)
I.gen to.who give-ma-1sg present go-PST.3SG broken
Intended: ‘The present that I gave to who broke down?’

The participle adjective segment *minun Jarille antamani* ‘I.GEN to.Jari give-MA-1SG’ best corresponds to a relative clause in English: it contains a predicate (here *antama* ‘given’) together with one or more arguments, which modifies the denotation of the whole noun phrase. The predicate agrees with the thematic subject, here the participant who gave the present to a third party. In Hungarian, the corresponding adjectival participle neither agrees (37) nor creates secondary Ā-chains (38).

(37)

a. a [Mari által Jánosnak adott] ajándék (#413, 414)
the Mary by John.DAT given present
‘the present given by Mary to John’

b. (?)a Jánosnak Mari által adott ajándék (#415)
the John.DAT Mary by given present
’the present given by Mary to John.’

(38)

a. ?a ki által Jánosnak adott ajándék? (#417)
the [who by John.DAT given] present
‘The present given by who to John’

b. ?a Jánosnak ki által adott ajándék?
the [John.DAT who by given] present
’the present givent by who to John’

c. ?a Mari által kinek adott ajándék? (#418)
the [Mary by who.DAT given] present
‘The present given by Mary to who’

d. *a kinek Mari által adott ajándék (#419)
the [who.DAT Mary by given] present

These constructions do not exist in English or Italian, but note that the Finnish and Hungarian constructions already provide the contrast we are looking for.

3.6 Relative clauses

Finnish (39) relative pronouns are strong operators, but they lack the in-situ option; English (40) relative pronouns are weak, while they too lack the in-situ alternative.

(39)

a.	Kaupunki	[[jota$_1$	kohti	__$_1$]	Seine	virta-a]	o-n	kuuluisa.(#108)
	city.NOM	which.PAR	towards		Seine.NOM	flow-PRS.3SG	be-PRS.3SG	famous
	'The city, towards which Seine flows, is famous.'							
b.	*Kaupunki	[[PP kohti	jota]		Seine	virta-a]	o-n	kuuluisa.(#111)
	city.NOM	towards	which.PAR		Seine.NOM	flow-PRS.3SG	be-PRS.3SG	famous
c.	*Kaupunki	[Seine	virta-a	[PP kohti		jota]]	o-n	kuuluisa.(#113)
	city.NOM	Seine.NOM	flow-PRS.3SG	towards		which.PAR	be-PRS.3SG	famous
d.	*Kaupunki	[Seine	virta-a	[PP jota		kohti]]	o-n	kuuluisa.(#112)
	city.NOM	Seine.NOM	flow-PRS.3SG		which.PAR	towards	be-PRT.3SG	famous

(40)

a. The city [[PP towards which] Seine flows] is famous. (#126)
c. *The city [[PP which towards __] Seine flows] is famous. (#127)
b. *The city [Seine flows [PP towards which]] is famous. (#128)
c. *The city [Seine flows [PP which towards __]] is famous. (#129)

The Hungarian relative clause follows the Finnish template with the exception that a separate topic phrase, occurring in a high position above operators in the Hungarian left periphery, can optionally (and marginally) precede the relative pronoun (see Section 3.7)(Kiss 2002: ch. 10.3), while Italian follows the English template (Cinque 1982). In all four languages relative pronouns pied-pipe material to the edge of the relative clause. The English pattern is captured by (11), since adpositions in English/Italian do not exhibit agreement. In all languages selected for a close examination, an in situ 'echo-relativization' is not possible.

3.7 Finite clauses

The finite element of the Finnish finite clause, which can be a tensed finite verb, auxiliary or negation (Mitchell 1991; Holmberg et al. 1993), agrees with a nominative grammatical subject, if present, and projects an additional specifier position that is occupied by the topic (not necessarily the grammatical subject) (Vainikka 1989; Vilkuna 1989; Vilkuna 1995; Holmberg and Nikanne 2002; Brattico 2019a; Huhmarniemi 2019). Topicalization derives part of the Finnish free word-order profile, illustrated in (41).

(41)

a.	Pekka	anto-i	kirja-n	Merjalle. (#420)
	Pekka.NOM	give-PST.3SG	book-ACC	Merja-ALL
	'Pekka gave a/the book to Merja.'			
b.	Merja-lle$_1$	anto-i	Pekka	kirja-n __$_1$. (#423)
	Merja-ALL	give-PST.3SG	Pekka.NOM	book-ACC
	'When it comes to Merja (=topic), Pekka gave her a/the book.'			

c. Merja-lle anto-i kirja-n Pekka __1 (#424)
Merja-ALL give-PST.3SG book-ACC Pekka.NOM
'When it comes to Merja, it was Pekka who gave her the book.'

d. Kirja-n1 anto-i __1 Merja-lle Pekka. (#425)
book-ACC give-PST.3SG Merja-ALL Pekka.NOM
'When it comes to this book (=topic), Pekka gave it to Merja.'

In English, the SpecTP position is occupied by the grammatical subject. With the exception of a few verbs, finite agreement is mandatory in both languages. There is therefore a correlation between agreement at T and the projection of a nonthematic and non-criterial preverbal specifier SpecTP, and both in English and Finnish.[10] In Hungarian, the operator (focus) follows a dedicated topic position (a) (or several) projected by an abstract topic head Top^0 (42) that triggers a topic Ā-chain, according to (Kiss 2002: ch. 2).

(42) a. [TopicP János1 [FP . . . [vP __1 keresi Marit]]. (#432)
János.NOM seek.PST.3SG Marit.ACC
'As for Janos, he seeks Marit.'

b. [TopicP1 Marit [FP . . . [vP __1 keresi János]]. (#433)
Marit.ACC seek.PST.3SG János.NOM
'As for Marit, Janos seeks her.'

c. [TopicP János1 [TopicP Marit2 [FP . . . [vP __1 keresi __2]]. (#435)
János.NOM Marit.ACC seek.PST.3SG
'As for Janos and Marit, he seeks her.'

Hungarian topic positions are not associated with overt agreement, but criterial scope positions were excluded from (11). We return to the topic/subject issue in Section 5.2.

3.7.1 Summary

Evidence was provided from Finnish, English, Italian and Hungarian that there might be a correlation between the projection of a nonthematic/non-criterial specifier position for an operator and the ability of the head to exhibit agreement with an argument. The correlation is realized through the major lexical categories such that if and only if some heads of a major lexical category K exhibit agreement, then all heads in K project formal specifier positions.

[10] According to (Holmberg and Nikanne 2002), the Finnish finite head (Fin^0 in their analysis) has an additional [NONFOCUS] feature checked against the topic phrase at SpecFinP. This analysis combines a semantically motivated topic position with agreement, something the Agreement-Remerge Correlation does not rule out. We believe, however, that the Finnish preverbal topic position is a formal position whose behavior is subsumed by ARC. See Section 5.2.

3.8 Analysis

3.8.1 Agreement and A-chains

It is well-known that agreement and A-chains are related to each other. For one, A-movement to SpecXP is often followed by agreement, various kinds of raising constructions being a clear demonstration of this effect. Another possible relationship was proposed by Alexiadou & Anagnostopoulou (1998), henceforth A&A, who, following (Taraldsen 1978; Chomsky 1981; Rizzi 1982), argued that in VSO languages verbal agreement can obviate phrasal movement to SpecTP.[11] This suggests that the positive correlation between agreement and A-chains concerns weak agreement; if agreement is rich, it can 'substitute' for phrasal movement. Both claims rely, however, on the notion that agreement and the projection of extended specifiers are related to each other. The data cited in the present article suggests that the pattern generalizes to Ā-chains.

3.8.2 Agreement and chain formation

We propose that chain formation and agreement are part of an abstract edge mechanism (43) licensing or mandating the projection of extended subjects (and later specifiers more generally).

(43) Edge feature

The edge feature at head H can either

a. license [EF_{φ}],

b. force [!EF_{φ}] or

c. block [−EF_{φ}] an extended subject at its noncomplement position.[12]

The "extended subject" can be either a phi-carrying overt phrase at the specifier position or overt phi-features at the head, following the main contours of A&A's analysis. The three options (a-c) in (43) can be illustrated by the Finnish adverbials reviewed in Section 3.3, examples (25)-(28). The licensing option (47a) is exercised when the adverbial predicate manifests overt specifiers and/or agreement optionally (ESSA-adverbial (25)). The forced regime (47b) characterizes those predicates which require that an extended subject is present (KSE-adverbials (27), TUA-adverbial (26)), either in the form of an overt subject or agreement (or both, redundantly). The blocking behavior (47c) characterizes the MA-infinitival (28), which exhibits neither agreement nor overt extended subjects. English and Italian adpositions and adverbials have

[11] Under their analysis, agreement obviated phrasal movement by allowing the rich verbal agreement features of V to check the uninterpretable phi-features at T via head movement. We follow essentially the same idea, but without taking a strong stance on how the rich agreement features end up at T. In our formalization, which is based on a recognition grammar, they come from the input. Agreement is discussed further in Section 5.1.

[12] In addition to A&A, the edge feature approach was inspired by (Chomsky 2008), where the term refers to a more general property of a lexical item LI permitting Merge to its edge. The edge feature posited in this study handles the special case of Internal Merge and the second edge feature in Chomsky's system.

the negative edge feature. When the edge feature is checked by both filling the specifier position and by agreement, the two must match. A positive edge feature triggers Agree, which checks that the features of the specifier and those at the head do not conflict with each other. We remain neutral concerning other possible features or conditions triggering agreement.

The data cited in this article has the property that the projected specifier positions are nonthematic and non-criterial. We assume that (exactly) one member of any chain containing the specifier element licensed or mandated by the edge feature must occur at a canonical thematic position. It is in this sense that the edge elements can be thought of as "extended" subjects: they replicate the contents of a thematic position. We further assume that an Ā-chain is formed if the specifier contains an operator. It follows that 'extended phrasal subject + operator → Ā-chain', which covers subject Ā-chains in our dataset. Whether the connection between operators and A-chains can be strengthened into an equivalence is complicated by the problem of topic chains, meriting a separate section (5.2).

These assumptions do not capture (11), which requires Ā-chain formation to apply to all instances of lexical items of a major lexical category K if it applies to some subset K′ of K in virtue of the positive edge feature. Some kind of generalization or inductive abstraction must occur. To capture it, we assume that the edge feature as such without any diacritic is responsible for projecting the extended specifier position and triggering chain formation, while the diacritic φ requires the presence of φ-features and restricts the operation to extended subjects. This generalizes the edge behavior from extended subjects into specifiers. It leaves room for a class which hosts extended specifiers but does not require extended subjects, and allows us to reformulate (11) by stating that when some K′ ⊆ K acquire [EF_{φ}], the edge behavior generalizes to K. We propose that this is because the edge behavior is a property of major lexical categories.[13] Crucially, this assumption allows us to incorporate criterial Ā-positions and primary Ā-chains into the analysis: they are generated by heads which project an edge but do not require φ-matching.[14]

[13] This statement of the law is compatible with several formalizations. In this study, it was implemented by lexical redundancy rules connecting major lexical categories directly with the edge behavior. An alternative is to assume that the major lexical categories are defined by or decompose into feature sets which include the edge feature. Another possibility is that the connection is hardwired into the algorithms processing lexical category information. A third option is force it into the system that performs lexical retrieval. It is difficult to find data that could separate these options from each other, so we decided on a strategy that was easiest to experiment with. These alternatives are not notational variants. The lexical redundancy strategy, for example, still leaves room for lexical exceptions.

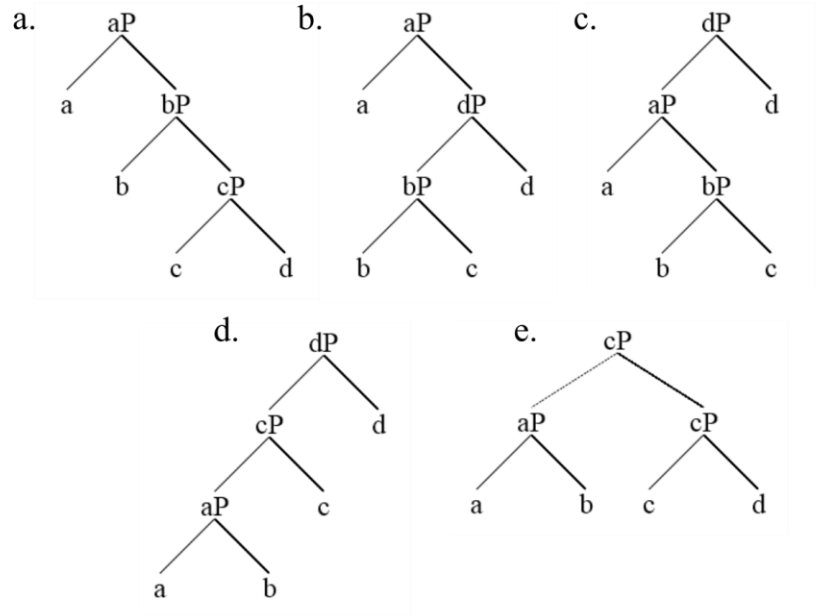
[14] The existence of nonagreeing criterial heads such as C projecting extended operator specifiers does not violate ARC (11), because ARC excludes thematic and criterial specifiers. Thematic and criterial specifier positions have an external semantic motivation, perhaps they involve a separate licensing lexical feature that differs from the edge mechanism. Nevertheless, this assumption introduces a class of nonagreeing lexical items with extended specifiers into the model; agreeing heads without the edge feature are still impossible.

4 Formalization and testing

4.1 Formalization

We subjected the analysis to a rigorous test by formalizing it and using the formalization to replicate the data. We used a Python-based minimalist grammar (Brattico 2019b; Brattico and Chesi 2020; Brattico 2022) as a platform for building the analysis. The algorithm, which implements a minimalist recognition grammar, creates minimalist asymmetric binary-branching bare phrase structure representations from linear lists of phonological words that it reads from left to right, following the original idea by (Phillips 1996).[15] Given a list of monomorphemic and empty *ad hoc* words $a * b * c * d$, the algorithm returns all phrase structures that are consistent with left-to-right depth-first linearization of the original string (44)a-e (#436). All phrase structure images, here and elsewhere, were produced by the algorithm.

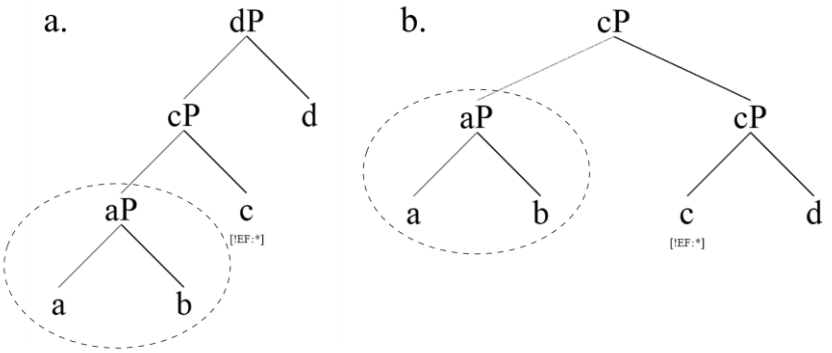
(44)



[15] The recognition grammar algorithm has been influenced by two additional theoretical approaches: top-down grammars (Chesi 2012) and the dynamic syntax framework (Kempson et al. 2001; Cann et al. 2005). Top-down rules such as rightward/top-down expansion at the bottom node are applied by the algorithm, though not exclusively. The dynamic syntax framework conceptualizes grammar as a dynamic parser engine. While we do not assume that the parser is literally the same thing as grammar, parsing operations are part of the derivations generated by our model. We ignore parsing in this article, however.

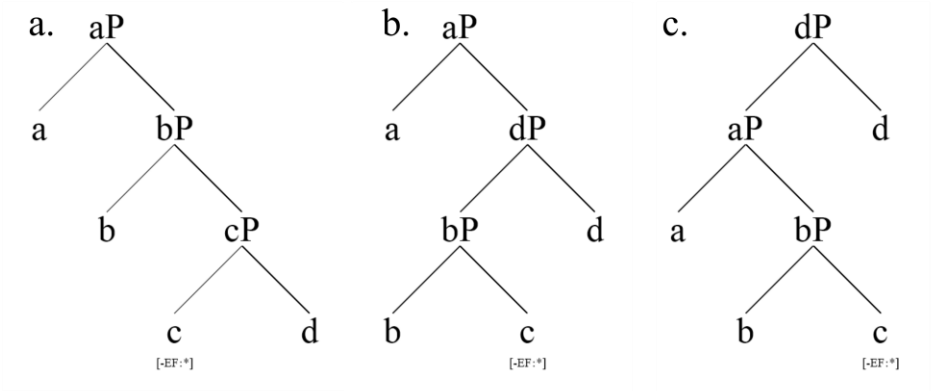
We define the notion of *specifier* of X to refer to any left phrase inside XP, which allows us to add the edge feature to the system. The definition presupposes a labeling algorithm.[16] Assuming $a * b * c_{[EF]} * d$ as a test input, the algorithm returns (45) out of the five logically possible outputs (#437).

(45)



These are the structures where the edge head has a phrasal specifier at its edge, given the current definitions for specifiers and labeling.[17] Setting [−EF] for *c* returns the three missing solutions by blocking (45)(#438).

(46)



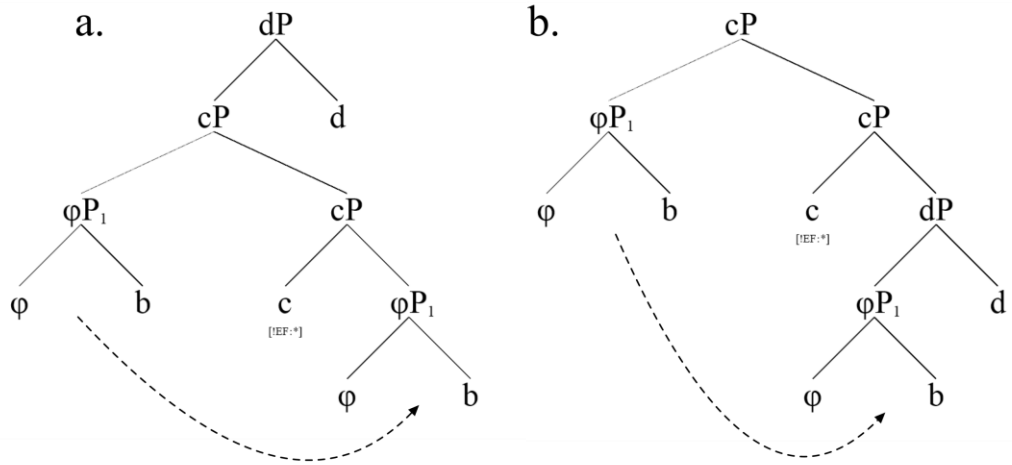
Only structures where *c* does not have an edge are now accepted. They correspond to English and Italian adpositions, for example. Adding agreement to *c* returns all five solutions due to the agreement cluster at the edge (#439), per our analysis. Recall that the edge of H includes both the left phrases inside HP and the agreement clusters at the head itself. The phi-features of the specifier must match with the phi-features

[16] Expressed in informal English, the labeling algorithm works as follows, ignoring the role of adjuncts: Suppose we target α for labelling. If α is primitive, it will be the label; if α is complex, then if the left constituent X of α = [X Y] is primitive, it will be the label; if X is complex but Y is primitive, Y will be the label; otherwise apply labeling recursively to Y. The labels generated by this algorithm are visible in all phrase structure trees generated by the algorithm.

[17] Feature [!EF:*] was created only for the purposes of performing these tests. It requires the presence of a mandatory specifier of any king (* = diacritic denoting a phrase of any kind).

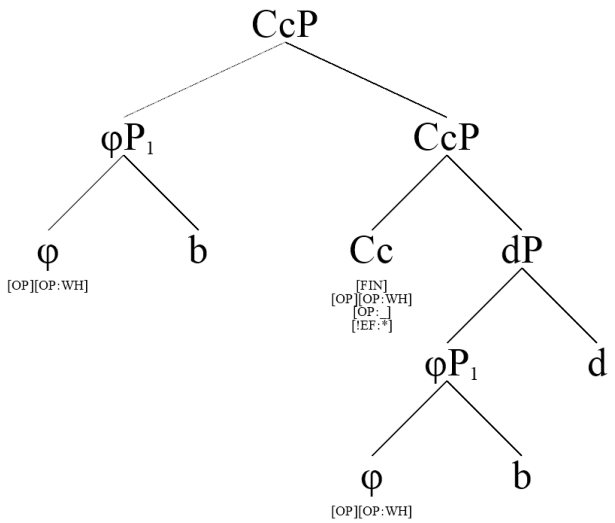
provided in the input sentence. If there is a mismatch, the specifier solutions are rejected while (46) are accepted (#440, 441). The agreement mechanisms were tested by several expressions in our dataset (#220-267). A further property is that the edge feature triggers chain creation. Letting the edge feature to control the activation of chains creates (47) for *a* ∗ *b* ∗ c ∗ *d* (#441). The tree drawing algorithm marks chains by subscripts; the arrows were drawn by the authors.

(47)



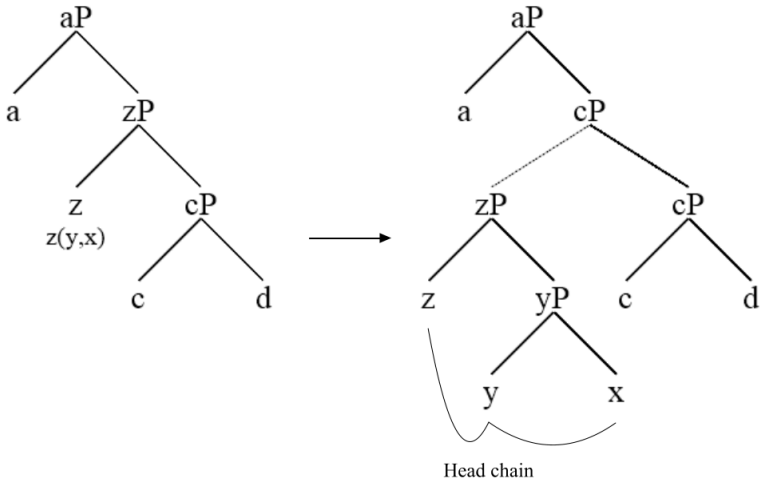
Notice that these chains need not to be realistic: the lexical items *a*, *b*, *c* and *d* used in these examples are unrealistic and used only to simplify the exposition and testing of the formal mechanisms. We examine realistic chains later. Ā-chains are generated if the specifier is an operator (48)(#442).

(48)



The presence of the operator at φ triggers additional feature checking mechanisms involved in the scope and clause type calculations, discussed later.[18] Secondary Ā-chains are generated in the same way, but the lack of finiteness at the edge heads renders the resulting operator positions inert with respect to scope and force.[19] Secondary Ā-chains are not accepted in English or Italian due to the [−EF] feature of the relevant heads. We will examine secondary chain creation below in connection with more realistic examples. In order to apply these assumptions to more realistic examples, the system must handle morphologically complex words. The background model maps derivationally complex words into syntactically complex heads and expands them into reversed head chains (Brattico 2022). Suppose *z* is a morphologically complex word with structure x#y#z, where x, y and z are primitive heads (e.g., *admires* = T + v + V). Example (50)e generates (52) with the head chain {z, y, x}(#444).

(49)

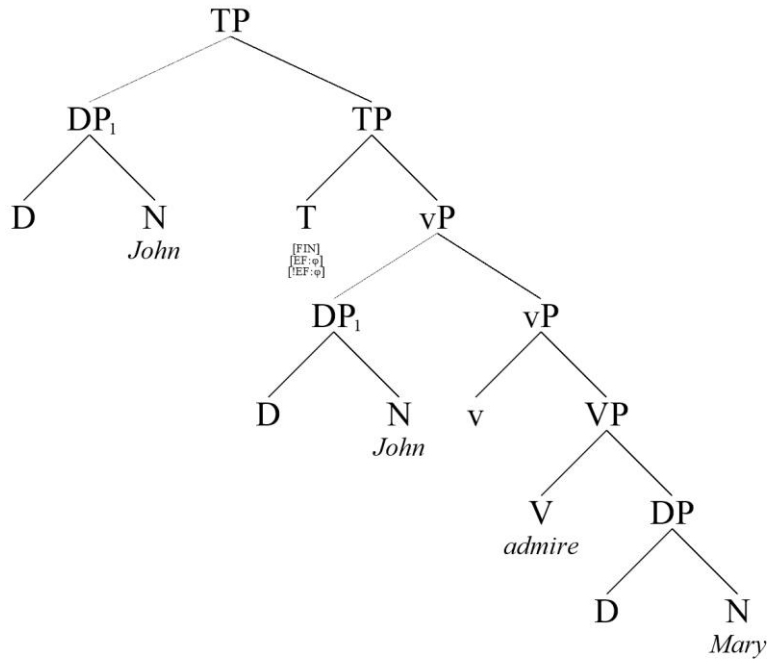


We can now apply the model to more realistic examples. A simple canonical transitive clause *John* ∗ *admires* ∗ *Mary* (#444) calculates (53). Proper names decompose into DPs = [$_{DP}$ D N] and the finite verb contains T (tense), v (transitivity) and V (verbal stem), creating reversed head chains D[…N…] and T […v…V…], respectively. The model generates (50).

[18] The three drawing algorithm labels the edge head as "Cc" because it thinks that the *c* head has properties of C; symbol "Cc" does not exist in the theory.

[19] The syntactic background algorithm calculates scope information by matching the reconstructed operator with a c-commanding finiteness head with the same operator feature (e.g., between (~~*what*~~$_{wh}$, $C_{wh,fin}$) in a sentence *what* $C_{wh,fin}$ *John saw* ~~*what*~~$_{wh}$). This produces wrong results with successive-cyclic Ā-chains with intermediate SpecCP positions as landing sites, creating spurious scope sites, and infinitival interrogatives such as *John wondered who to trust*, which remain without scope. These phenomena were not present in our dataset, so we did not attempt to correct these issues. Something else than 'finiteness + operator' must determine the legible scope positions.

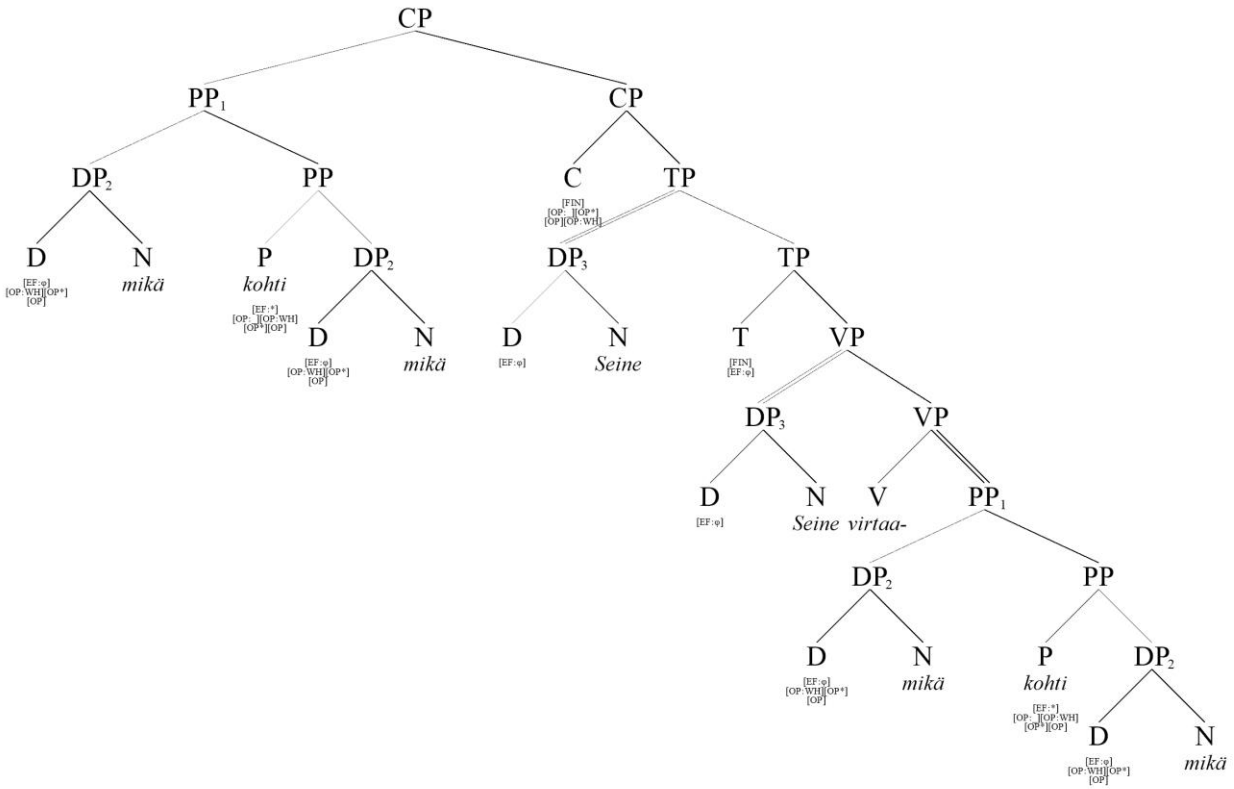
(50)



English finite T has the edge feature which generates the A-chain {*John*, *John*}, with the lowest member at the thematic position. The thematic position is defined by v, which assigns a theta role to its specifier. An Ā-chain is generated if the specifier is an operator. The model calculates Finnish (51) as (52).

(51) [Mitä$_1$ kohti __$_1$]$_2$ Seine virta-a __$_2$? (#71, 448)
what.PAR towards Seine.NOM flow-PRS.3SG
'Towards what does Seine flow?'

(52)

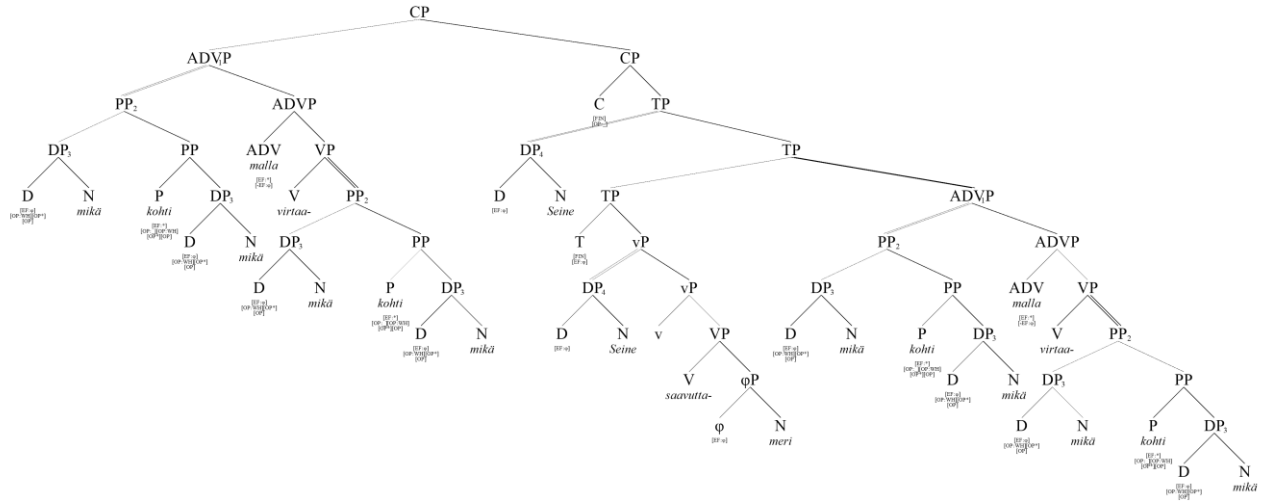


The adposition *kohti* 'towards' has the edge feature, while the operator triggers a secondary Ā-chain. The operation is repeated at the clause level, generating an embedded chain. T with [EF_{φ}] creates an A-chain, assuming that the Finnish preverbal position is a formal edge position, not a topic position (see Section 5.2). The operator is bound by $C_{[wh,\ fin]}$ at the LF-interface. See the supplementary document* for how to read this information from the output generated by the model.

Heck's edge generalization requires that the operator occurs at the edge of the pied-piped phrase in order to generate interrogative force and scope, which holds for some but not all operators. We assumed that Finnish operators are partitioned into strong and weak, the former associated with the edge generalization. To model the distinction we assume that the checking dependency between a strong operator and its edge head must hold between the *head* of the operator phrase and the head of the hosting phrase. Weak operators, such as the Finnish yes/no operator, do not require local head-to-head checking. If we leave any of the Ā-chains unimplemented in the (51), the algorithm derives an echo-question. This is shown in (53)/(54).

(53) [[Kohti mitä]$_1$ virtaamalla __$_1$]$_2$ C Seine saavutta-a valtamere-n __$_2$?(#450)
towards what.PAR flow-MA/INF Seine.NOM reach-PST.3SG ocean-ACC
'By flowing towards what does Seine reach the ocean? (only echo-interpretaton)'

(54)



Notice that the main clause C now lacks all operator features. Adv^0 can only check [*wh*] from the head of the PP, which needs to check the feature from the head of the DP at its specifier position, now missing. The output is grammatical, but since the strong operator cannot be paired with a finite head with the same operator feature, a regular interrogative is not derived.

To summarize, the analytic principles proposed in 3.8 were formalized in Python. This setup allows us to explore their logical implications in a rigorous way. Section 4.2 elucidates the testing procedure, Section 4.3 summarizes the results.

4.2 Methods

4.2.1 Design and procedure

The data reviewed in this article together with several other sentence variants were collected into a dataset file (Section 4.2.2). The algorithm processed each input sentence from that file† by using the assumptions posited in our analysis, and then provided the output in the form of several files and images. Specifically, it paired each input sentence with a *grammaticality judgement* and a *derivation*, the latter containing all linguistically meaningful computational steps executed during the processing (corresponding to runtime logging). If the sentence was judged grammatical, it was further provided with a *syntactic analysis* and *semantic interpretation* (several if the input was ambiguous). The output was written by the algorithm into text files and phrase structure images that constitute the raw data of this study. The sentences in the dataset as well as in the output files were identified by a numerical identifier (#N) that can be used to locate the same items from different files and from the main article. The *lexicon*, which contained a lexical entry for all phonological words appearing in the input, was provided as an external file. The lexicons for different languages were handled by constructing a separate parser instantiation at runtime for the speakers of all languages present in the dataset, and then by processing sentences in language L by the language-specific recognition grammar for L. We can imagine the grammar instantiations as ‘idealized brain models’ for the

speakers of different languages, as they incorporate experience- and acquisition-based differences between speakers. The algorithm provided all grammatical sentences with phrase structure images, showing the calculated structure at the LF interface. All raw input and output files, as well as the source code, are available in the public domain.[‡] The design of the testing regime is illustrate in Figure 1.

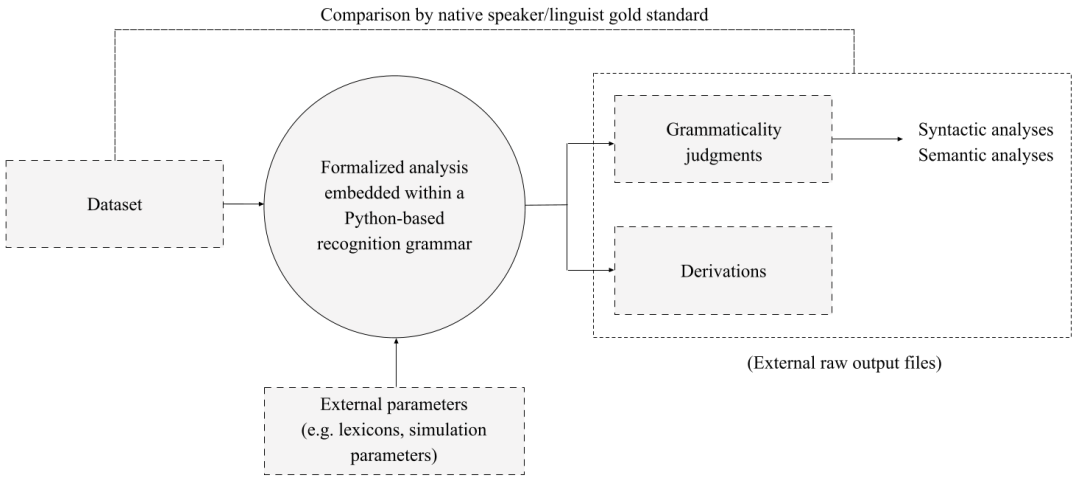


Fig. 1. Design of the study. The model, when provided with a few external parameters and a linguistic dataset, maps the input sentences in the dataset into grammaticality judgments and derivations, and then provides the grammatical sentences with syntactic and semantic analyses. The output is compared with a native speaker/linguist gold standard.

4.2.2 Stimuli (dataset)

The dataset was composed from the sentences (or syntactic equivalents) cited in this article plus several variants so that all word orders/construction variants of interest were included. Sentences from four languages were included: Finnish, Hungarian, Italian and English. The input sentences were tokenized and normalized linear strings of phonological words with no morphosyntactic tagging or decompositions apart from a few cases where we wanted to disambiguate some word in order to make the output easier to analyze. Table 1 summarizes the contents of the dataset, complete dataset is available online.[§] There were a total of 467 test sentences in the dataset.

Table 1. Summary of the dataset.

#	Class	Subclass	Representative examples
1-24	Secondary Ā-chains and topicalization	Secondary Ā-chains, topicalization, all adposition types (I, IIa-b)	*Parisia lähellä Pekka asuu* '[Paris-par near __] Pekka lives __]' *Kohti Pariisia Seine virtaa* '[towards Paris-par] Seine flows __' **Pekka asuu lähellä minun* 'Pekka lives [near I.par __]'
25-28		Secondary Ā-chains, topicalization, in English	*John lives near Paris* **John lives Paris near __*

			**[Near Paris] John lives __*
29-52		Secondary A-chains, topicalization, adpositions inside adverbials, all adpositions	*Seine saavuttaa meren virtaamalla kohti Pariisia* 'Seine reaches ocean by.flowing towards Paris' *Pariisia kohti virtaamalla S. saavuttaa meren* '[[Paris.par towards __] by.flowing __] Seine reaches ocean _ **Seine saavuttaa meren virtaamalla lähellä Pariisin* 'Seine reaches ocean by.flowing near Paris.gen'
53-61	Strong operators with the in situ option	Ordinary interrogative sentences, Finnish and English; basic operator mechanisms	*Who does John admire?* *Ketä Pekka ihailee?* 'who.par Pekka admires __?' *Pekka ihailee ketä* 'Pekka admires who' (echo-interrogative)
62-77		Adpositions and strong operators in Finnish	*Pekka asuu lähellä mitä* 'Pekka lives near what' (echo) *Kenen lähellä Pekka asuu* '[who.gen near __] Pekka lives __? **Pekka asuu lähellä kenen* 'Pekka lives near who.gen'
78-81		Adpositions in English	*John lives near what* **[What near __] John lives __*
82-101		Adpositions inside adverbials in Finnish, all adpositions and word orders	*Seine saavuttaa meren virtaamalla kohti mitä* 'Seine reaches ocean by.flowing towards what' *Mitä kohti __ virtaamalla __ Seine saavuttaa meren __?* 'What towards __ by.flowing __ Seine reaches ocean'
102-119	Strong operators without the in situ option	Finnish relative clauses	*Kaupunki joka kohti Seine virtaa hävisi* 'city [which towards __] Seine flows __ disappeared' **Kaupunki [kohti jota] Seine virtaa __ hävisi* 'city towards which Seine flows disappeared'
120-131		English relative clauses	**The city which towards Seine flows disappeared* *The city towards which Seine flows disappeared*
132-167	Weak operators without the in situ option	Yes/no operator in Finnish, all word orders	**Pekka asuu lähellä Pariisia-ko* 'Pekka lives near Paris-Q' *Lähellä Pariisia-ko Pekka asuu __?* '[Near Paris-Q] Pekka lives __'
		Yes/no operator in Finnish in adposition and adverbial constructions	**Seine saavuttaa meren virtaamalla kohti Pariisia-ko* 'Seine reaches ocean by.flowing towards Paris-Q' *Pariisia-ko kohti virtaamalla Seine saavuttaa meren?* '[[Paris-Q towards __] by.flowing __] Seine reaches ocean'
168-181	Weak operators with the in situ option	Finnish contrastive focus operator, adpositions	*Pekka asuu lähellä PARIISIA* 'Pekka lives near PARIS (not Lyon)!' *PARIISIA lähellä Pekka asuu* '[PARIS (not Lyon) near __] Pekka lives __!'
182-203		Contrastive focus operator in adposition and adverbial constructions	*Seine saavuttaa meren virtaamalla kohti PARIISIA* 'Seine reaches ocean by.flowing towards PARIS' *PARIISIA kohti virtaamalla Seine saavuttaa meren* '[[PARIS towards __] by.flowing __] Seine reaches ocean __
204-219	Combinations of strong and weak operator	Interrogative and yes/no operators in the same operator	*Mitä-kö lähellä Pekka asuu?* '[what-Q near __] Pekka lives __' *Pekka asuu lähellä mitä-kö?* 'Pekka lives near what-Q'
220-249	Agreement and chain formation	Agreement and word order in Finnish adpositions	*lähellä minua* 'near I.par' **minua lähellä-ni* 'I.par near-1sg __'
250-255		Finite clauses	*Pekka ihaile-e Merjaa* 'Pekka admire-3sg Merja' *Pekka ihaile-n Merjaa* 'Pekka admire-1sg Merja'
256-261		Noun phrases	*Minun kello-ni hävisi* 'I.gen watch-1sg disappeared' **Meidän kello-ni hävisi* 'We.gen watch-1sg disappeared'
262-267		Participial adjective clauses	*Se minun tekemä-ni kello hävisi* 'that I.gen made-1sg watch disappeared' **Se meidän tekemä-ni kello hävisi* 'that we.gen made-1sg watch disappeared'
268-300		Agreement and secondary Ā-chains	**Seine virtaa kohti-ni minua* 'Seine flows towards-1sg I.par'

			**Minun kohti-ni Seine virtaa*
			'[I.gen towards-1sg __] Seine flows __]
301-308		Agreement and secondary A-chains in English	**John lives near-3sg Paris*
			**Paris near-3sg __ John lives __*
309-318		Agreement and secondary A-chains in Italian	*Giuseppe vive vicino a Parigi*
			'Giuseppe.nom live.prs.3sg near at Paris'
			**Cose a vicino Giuseppe vive*
			'[[what at __] near __] Giuseppe live.prs.3sg __'
319-326		Agreement and secondary A-chains in Hungarian	*Út nem vezet át az erdőn*
			'path not lead through the forest'
			Át melyik erdőn út nem vezet
			'[through [which forest]] path not lead __'
327-336	Infinitivals and secondary chains	Finnish ESSA-adverbial	*Pekka lähti tehdessä-än taloa*
			'Pekka left while making.essa/inf-3sg house'
			Mitä tehdessä-än Pekka lähti?
			'[what making.essa/inf-3sg __] Pekka left __'
337-346		Finnish TUA-adverbial	*Pekka lähti tehtyä-än taloa*
			'Pekka left made.tua/inf-3sg house'
			Mitä tehtyä-än Pekka lähti?
			'[what made-tua/inf-3sg __] Pekka left __'
347-352		Finnish KSE-adverbial	*Pekka lähti tehdäkse-en taloa*
			'Pekka left make.kse/inf-3sg house'
			mitä tehdäkse-en Pekka lähti
			'[what make.kse/inf-3sg __] Pekka left __'
353-362		Finnish MA-adverbial	*Pekka lähti tekemättä taloa*
			'Pekka left make.ma/inf house'
			Mitä tekemättä Pekka lähti?
			'[what make.ma/inf __] Pekka left __'
363-368		Double fronting in Finnish adverbials	**Pekka lähti taloa minun tehtyä*
			'Pekka left house$_1$ I.gen made __$_1$'
			**Pekka lähti minun taloa tehtyä*
			'Pekka left I.gen$_1$ house.par made __$_1$'
369-382		English adverbials and secondary Ā-chains	*Seine reaches the ocean by flowing towards Paris*
			By flowing towards what Seine reaches the ocean
			**Paris by flowing towards Seine reaches the ocean*
			**What towards by flowing Seine reaches the ocean*
383-385		Italian adverbials and secondary Ā-chains	*Giuseppe ha attirato attenzione per aver vinto Luisa*
			**Cosa per aver vinto Giuseppe ha attirato attenzione*
386-396	Determiner and noun phrases	Finnish noun phrases (DPs, φPs)	*Merjan kello hävisi*
			'Merja.gen watch disappeared'
			Kenen kello hävisi?
			'who.gen watch disappeared?'
397-400		Hungarian noun phrases (DPs, φPs)	*az én' diákjaim vezet*
			'the my book runs'
			**az mi diákjaim*
			'the I.nom student.pl'
401-405		English determiner phrases (DPs)	*John admires his sister*
			Whose sister admires him?
406-408		Participial adjective phrases in Finnish	*Se minun tekemä-ni kello hävisi*
			'that I.gen made-1sg watch disappeared'
			Se kenen tekemä kello hävisi
			'that who.gen made watch disappeared'
409-415		Participial adjective phrases in Hungarian	*a Mari által adott ajándék vezet*
			'the [[Mari by __] given] present runs'
			**a kinek Mari által adott ajándék vezet*
			'the [who.dat [Mary by __] given] present'
416-421	Finite clause	Finnish finite clause and word order variations	*Pekka antoi kirjan Merjalle*
			'Pekka.nom gave book to.Merja'
			Merjalle antoi kirjan Pekka
			'to.Merja gave book Pekka.nom'
422-427		Special constructions	*Minun täytyy ihailla Merjaa*
			'I.gen must admire Merja'
			**Merjaa täytyy ihailla minun*
			'Merja.par must admire I.gen'
428-432		Hungarian finite clause and topic constructions	*János keresi Marit*
			'János seeks Mari'
			Marit János keresi

			‘Mari.top János.top seeks
433-435		English finite clause and word order	*John admires Mary* *John Mary admires* **Admires John Mary*
436-467	Formalization	Sentences from the main article, sections §4, 5	

The dataset was composed with the aim of verifying that the analysis does calculate the data. The dataset therefore aims to cover all sentences that ‘operationalize’ the phenomenon (11) targeted for an analysis. Within that narrowly defined domain, the dataset then aims to be exhaustive; this explains how the original 54 sentences cited in this article expanded tenfold.

4.3 Results

First we examined if the analysis replicated native speaker grammaticality judgements (i.e., whether it was observationally adequate). Success rate was 100%. It judged all ungrammatical sentences as ungrammatical, grammatical as grammatical. Thus, the analysis captures the dataset when it comes to predicting grammaticality.[20]

Next we verified that the semantic interpretations (e.g., echo-readings, interrogativization, thematic roles, scope information) and syntactic analyses were correct and/or at least plausible. We will examine the core results in this article; there is a technical supplementary document[**] which examines the complete output and a few additional details we ignore here. We begin by considering the derivation of the Agreement-Remerge Correlation (11), repeated as (55).

(55) *Agreement-Remerge Correlation (ARC)*

Any lexical item in a major lexical category K licenses a nonthematic and/or non-criterial specifier position for A/Ā-chains (in language L) if and only if at least some lexical items in K agree (in L).

The basic pattern captured by (55) is repeated in (56).

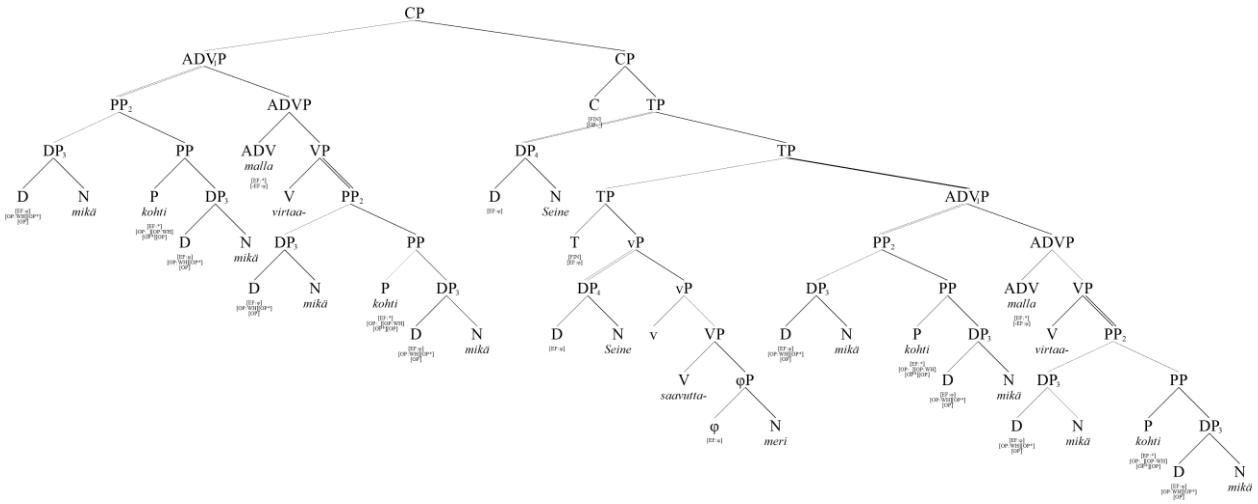
(56) [[Mitä$_1$ kohti __$_1$] virtaa-malla __$_2$]$_3$ Seine saavutta-a __$_3$ mere-n ?(#89)
what.PAR towards flow-MA/INF Seine.NOM reach-PST.3SG ocean-ACC
‘By flowing towards what (city) does Seine reach the ocean?’

Heads participating in the agreement/chain creation have a positive edge feature which either licenses or mandates the presence of an extended subject at the edge of the head and triggers chain creation. The feature

[20] The algorithm was inefficient, creating an average of 10.05 garden paths (range 0-249) and 594.5 elementary Merge operations (range: 2-6248) for the grammatical sentences before finding the first acceptable solution. The parser was overwhelmed by the nested Ā-chains. We ignore parsing efficiency in this study.

is present in Finnish adpositions, adverbials and the criterial operator heads and creates the embedded chains visible in (56). The model derives embedded Ā-chains (57).[21]

(57)



Sentences #62-77, 82-101 contains variations of Finnish sentences exhibiting this pattern. Inverted word orders are rejected in English and Italian: lack of the positive edge feature leaves extended specifiers at the uninterpretable surface positions, and the structure is rejected at the LF-interface (#78-81, 301-318). Only base-generated positions (e.g., *towards which city*) are accepted. The fact that the Finnish adposition *kohti* 'towards' exhibits this behavior even though it does not itself exhibit agreement is captured by the generalization according to which the edge feature behavior creates (morpho)syntactic paradigms for the major categories.[22] The preposition *kohti* 'towards' in (57) has [EF] even though it does not agree itself; all English or Italian prepositions have [–EF]. The full dataset contained examples from three types of Finnish adpositions which have a slightly different syntactic properties; detailed examination of them can be found from the supplementary.

While all inverted word orders are ungrammatical in English and Italian (#369-385), most of the logically possible word order possibilities are grammatical in Finnish when it comes to *wh*-operators and other operators exhibiting the in situ alternative. The positive edge feature of the Finnish adverbials and

[21] The tree drawing algorithm can be instructed to show selected features. The trees produced in this study contain the original words, labels (for lexical items; the rest are determined by the labelling algorithm) and features that were relevant for the analysis proposed here. For a complete list of these features, see the supplementary document. High-resolution versions of the most important phrase structure images shown in this section are available for download, and are cited in the endnotes.

[22] We left the model with an in-principle option for creating lexical items that are exempt from this rule, although no such items were present in our dataset or in the languages we examined. Such exceptions can be created by attaching the edge feature directly to a specific lexical item. This follows from our decision to model the generalization by using lexical redundancy rules. There are other, theoretically stronger and more interesting options; see footnote 13.

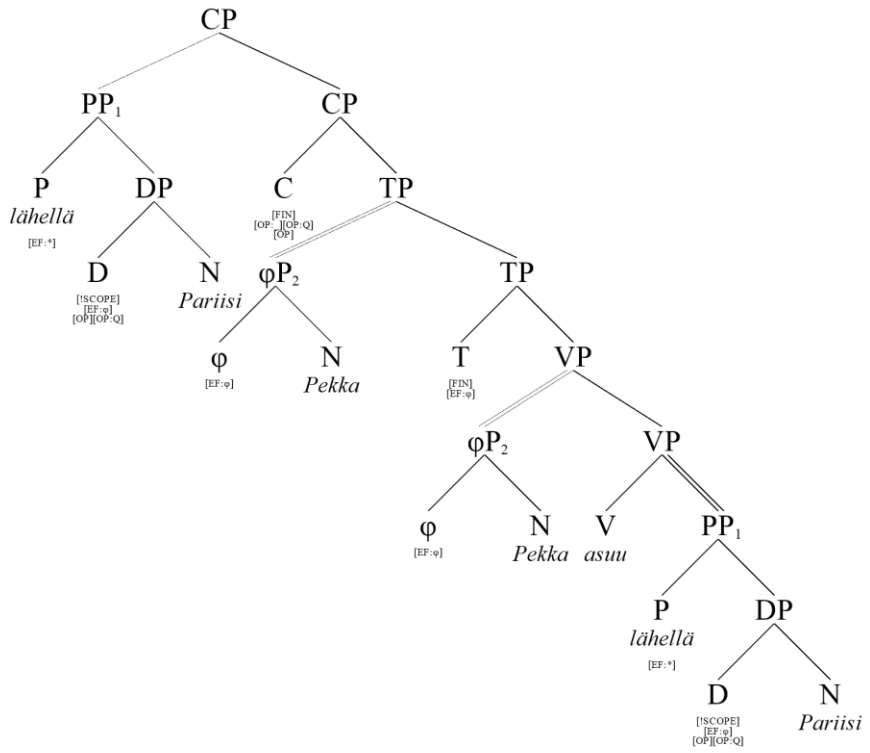
adpositions does not force chain creation, only licenses it. If the argument of the adposition occurs at the complement position in the input (i.e. the order is *towards which city*), it will be generated directly to the complement position on the basis of the input, no chains are generated, and the canonical P-DP passes at the LF-interface. All logically possible word orders are captured by the same mechanism: edge specifiers trigger chains, complements do not. If the edge behavior is mandated, then either the specifier position must be filled with an overt phrase or rich agreement must occur at the head. See the Finnish adverbial data (Section 3.3) for an example of each grammatical profile. This derives sentences in our dataset where the presence of an extended subject is required, including canonical finite clauses in English, deriving its strict EPP profile.

All other word order variations result in an echo-question reading, since strong operators were required to occur at the edges of their pied-piping phrases to check the corresponding feature at the finite operator head generating criterial scope positions, in compliance with Heck's edge generalization (see (53)).[23] Consequently, the scope position will not be generated if the word order differs from (56), and the semantic component will report an echo-question. Weak operators, such as the Finnish yes/no question operator, are exempted from the edge generalization. While weak operators must still occur at the left peripheral scope positions, they are not required to occur at the edges of the pied-piped phrases (#132-167). Sentence (58), for example, is analyzed as (59).

(58) Lähellä Pariisi-a-ko Pekka asu-u? (#141)
near Paris-PAR-Q Pekka.NOM live-PRS.3SG
'Was it near Paris (and not Lyon) that Pekka lives?'

[23] Recall that strong operators were required to check the corresponding operator feature at the hosting head by means of dependency that connects the hosting head with the head of the pied-piped phrase, between H_{wh} and C_{wh} under [[$_{HP}$ …H_{wh} XP] C_{wh} . . .]. This forces the operator to the outer edge position at the final spellout configuration. Weak operators were exempted by definition. These mechanisms provide an analysis of Heck's edge generalization.

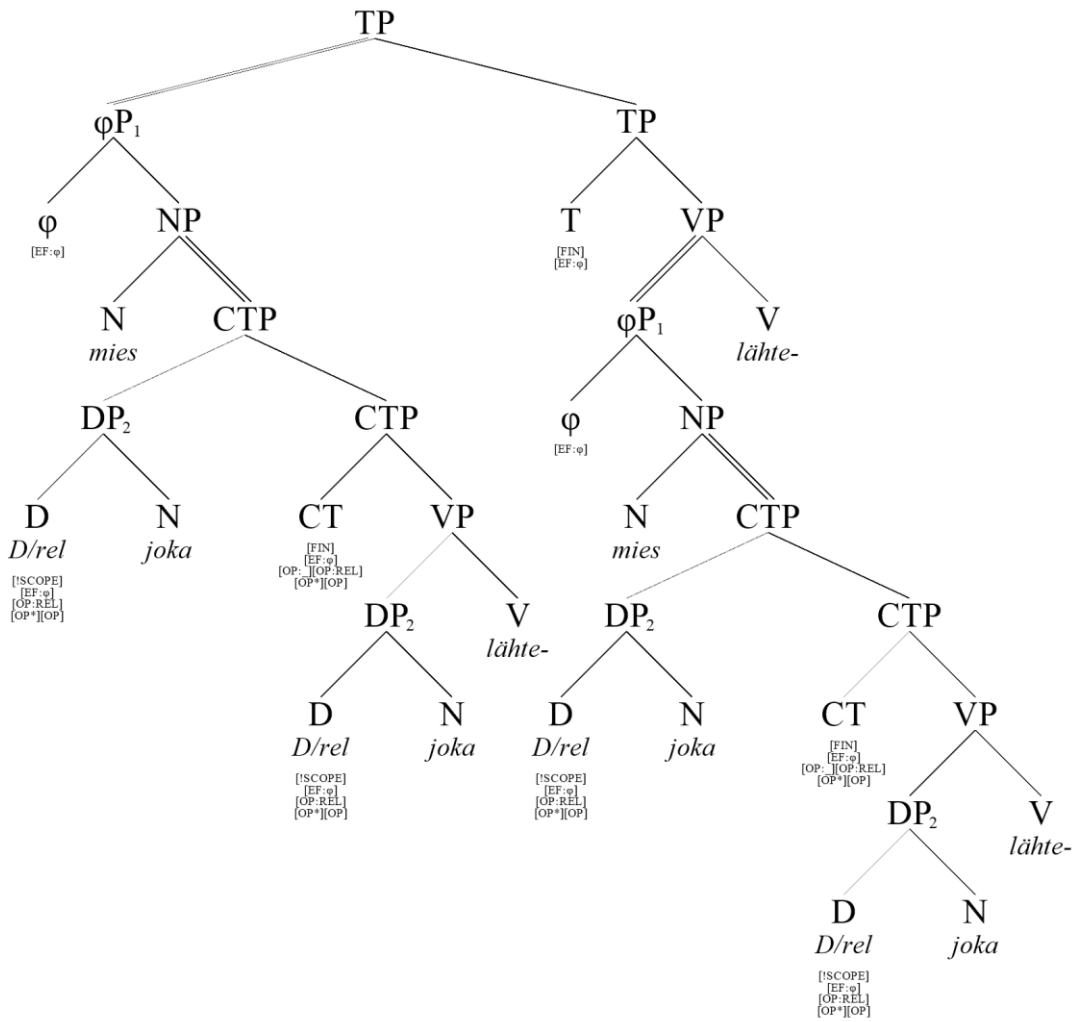
(59)



The yes/no Q-feature at C can be checked by the DP inside the PP at the scope position SpecCP because the Q operator is weak. The in situ alternatives, which are available for *wh*-interrogatives and contrastive focus operators in Finnish, were specifically licensed by a lexical feature in this study.[24] Finnish relative clauses (#102-119) exhibit the combination of strong operators without the in situ option and were correctly reproduced. Sentence *mies joka tuli lähti* 'a man who came left' (#102) is derived by the analysis as (60).

[24] The feature is [!SCOPE], which requires that the operator is linked with its scope position by an overt chain. See the supplementary document for a complete list of the lexical features posited in this study.

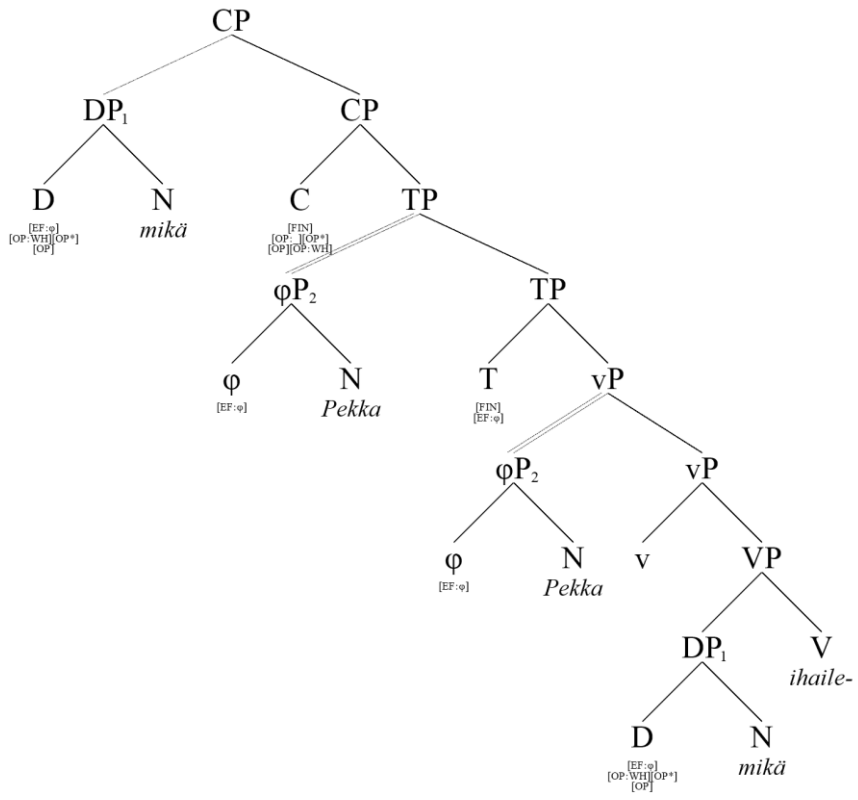
(60)



All variations where the relative pronoun remain in situ with respect to any of the required Ā-chains are correctly judged ungrammatical (#111-119): the model first generates an 'echo-relative clause' which is judged illegitimate when the relative pronoun cannot be paired with a finite scope marker at the LF interface. When the operator does occur at the criterial scope position, and the required operator checking configuration has been established between the moved operator and the finite head, the operator is bound by a finite scope marker (typically C) at the LF-interface. In English and Italian, the relative clauses lack the in situ option, but the operator itself is weak and does not (and cannot) satisfy the edge generalization (#120-131). A standard Finnish interrogative (61) is analysed as (62).

(61) Mitä Pekka ihaile-e? (#55)
what.PAR Pekka.NOM admire-PRS.3SG
'What does Pekka admire?'

(62)



Only the base position is targeted for binding by the scope marker, the copies are not. The sentence is judged to be an interrogative. If the operator appears in its canonical position in the input, C will not have [*wh*] and the sentence will be interpreted as an echo-question (#57).

The ±strong and ±in situ taxonomy leaves room for weak in situ operators. Finnish contrastive focus operator, expressed by prosodic stress, fit this description. Both primary and secondary chains are optional (#168-203). Prosodic stress is glossed by as "foc" and represented by the same feature in the input sentences.

(63) a. Pekka asu-u PARIISIA lähellä. (#169)
Pekka.NOM live-PRS.3SG Paris.PAR.FOC near
'Pekka lives near PARIS (not Lyon)!'

b. Lähellä PARIISIA Pekka asu-u! (#175)
near Paris.PAR.FOC Pekka.NOM live-PRS.3SG
'Pekka lives near PARIS (not Lyon)!'

c. PARIISIA lähellä Pekka asu-u! (#176)
Paris.PAR.FOC near Pekka.NOM live-PRS.3SG
'Pekka lives near PARIS (not Lyon)!'

Notice that the model does distinguish (a) from (b, c): in the latter, but not in the former, the contrastive focus operator is bound by an explicit scope marker. This difference is reported in the output, but not linked with any specific semantic attribute.

Finnish makes room for certain combinations of operators, such as interrogative pronouns with the yes/no operator. Testing all combinations systematically would increase the size of the dataset enormously, so only few examples were tested here, as discussed in the supplementary (#204-219). As a general principle, however, whenever a lexical item contains several operator features, the semantic system processes them independently. The model does not specifically rule out operator interaction, however.

The edge feature controls Agree, which checks that the agreement features at the head and a local subject match (if they are both present). Grammatical and ungrammatical agreement combinations were tested for adpositions (#220-249), finite clauses (#250-255), noun heads (#256-261) and participle adjectives (#262-267) and were correctly derived.

Secondary Ā-chains as such are never obligatory: failure to implement a secondary Ā-chain leads to an echo-interrogative, not ungrammaticality. Obligatory Ā-chains, as exemplified by pied-piped relative operators, are an indirect consequence of the scope marking dynamics (i.e., strong operator without the in situ option). When it comes to nonoperator subject arguments, the situation is different. Replicating the Finnish infinitival data requires that we use obligatory, optional and blocked edge features (#327-362). Obligatory edge feature forces the infinitival to project an extended subject, either in the form of an overt subject or agreement; optional edge feature licenses them. The blocking behaviour generates infinitivals where the overt subject (always the thematic agent in our dataset) is not present and must be controlled from outside by an antecedent. The standard English EPP principle, which requires the presence of grammatical subject in a finite clause, represents a special case of the same mechanism. [EF_{φ}] mandates φ-features to the edge of TP, but because agreement features are weak (Sections 3.8.1, 5.1), the grammatical subject occupies SpecTP.[25]

In sum, the semantic interpretations and syntactic analyses generated by the model did not run into obvious problems. The model was observationally adequate. We claim on such grounds that the proposed analysis suffices to capture the data, and that it provides one possible avenue for incorporating the Finnish/Hungarian word order inversion phenomenon into the theory of UG.

[25] The edge mechanism cited here constitutes a necessary but not sufficient criterion for the standard EPP. An additional assumption is that in English the subject chain must be a local A-chain, preventing direct objects of transitive clauses from satisfying the EPP. This restriction does not work in Finnish, where all clausal arguments can be topicalized to SpecTP. See example (41) and Section 5.2.

5 Extensions and problems

5.1 Strong and weak agreement

The analysis presupposes that [EF_{φ}] can be checked by agreement (Agree). The operation can be illustrated by Italian, which is a complete pro-drop language and allows subject drop in all person and number configurations (64)(#461-467).

(64)

a.	(Io) adoro Luisa	b.	(tu) adori Luisa	c.	(lei/lui) adora Luisa
d.	(Noi) adoriamo Luisa	e.	(voi) adoriate Luisa	f.	(Loro) Adorano Luisa

It therefore constitutes a prototypical 'rich agreement language' where all verb suffixes check [EF_{φ}], in agreement with A&A. In English, finite agreement is too weak and requires the presence of an overt grammatical subject (65).

(65) *Admires John Mary. (#455)

English therefore represents a prototypical 'weak agreement language', under our analysis. Absent a more detailed crosslinguistic dataset on agreement, these two profiles could be handled by stipulating a strong/weak feature affecting edge feature checking. Although a stipulation, the present dataset cannot motivate anything more complex. The picture is complicated by Finnish, however, which patterns with Italian with the exception of the third person, where an overt subject must be present (66)a-b (Vainikka and Levy 1999; Holmberg 2005; Holmberg et al. 2009; Holmberg and Sheehan 2010).

(66) a. *Ihaile-e (Pekka) Merja-a. (#456, 457)
admire-PRS.3SG (Pekka.NOM) Merja-PAR
b. Ihaile-n Merja-a. (#458)
admire-PRS.1SG Merja-PAR
'I admire Merja'

In the third person, a phrase must occupy SpecTP. There is a long and persistent literature arguing that the phrase occupying SpecTP in Finnish must be the topic (Vainikka 1989; Vilkuna 1989; Vilkuna 1995; Holmberg and Nikanne 2002; Miyagawa 2010; Brattico 2019a; Huhmarniemi 2019; Brattico 2021a). A possible model where the edge feature is checked against a topic runs into two problems. First, the condition that the specifier and head must agree with each other can no longer be maintained, since the topic does not have to be the agreeing grammatical subject. One could perhaps develop an analysis positing a topic edge feature [EF_{topic}], but there is a further issue: the existence of a main clause antecedent can render the third person pro-drop construction grammatical (Vainikka and Levy 1999) (67).

(67) Pekka sano-i että [ihaile-e Merja-a.] (#461)
Pekka.NOM say-PST.3SG that admire-PRS.3SG Merja-PAR
'Pekka said that he (=Pekka, main clause subject) admires Merja.'

There is no topic phrase at the preverbal subject position, but the clause is grammatical. The underlying algorithm handles (67) by assuming, following an earlier proposal by (Holmberg and Sheehan 2010), that the Finnish third person suffixes do not carry a referential definiteness (D) feature, which is repaired at LF by an antecedent mechanism that picks up a local antecedent, the clause-mate topic or the subject/topic of the next clause up, whichever comes first (Brattico 2021b). The presence of the antecedent "rescues" (67) from the EPP violation. Whether this is the most elegant way to approach this problem, the relevant notions of rich and weak agreement seem to be too coarse-grained for the purposes of a crosslinguistic theory of edge feature checking. Individual phi-features (e.g., third versus first and second person in Finnish) might provide a more realistic starting point. The second implication is that a sentence with an unchecked weak phi-feature, at least the D-feature in Finnish, can be handled by an antecedent that provides the missing value. Notice that many predicates which block the occurrence of an extended subject, such as the Finnish MA-infinitival, still require an antecedent, creating a similar control construction. In sum, while the edge mechanism regulates the occurrence of local and overt extended arguments, a checking failure can – at least in some cases – be avoided by the antecedent mechanism. How to best add this phenomenon – repair by control – to the edge mechanism remains to be studied.

5.2 Nonoperators, Ā-chains and topicalization/scrambling

Processing of topics and topicalization turned out to pose nontrivial problems, both for our analysis and for the general architecture of the UG. We noted in Section 2.1 that Finnish nonoperators seems to be able to create Ā-chains (68).

(68)

a. Seine virta-a [kohti Pariisi-a.] (#9)
Seine.NOM flow-PRS.3SG towards Paris-PAR
'Seine flows towards Paris.'

b. Seine virta-a [Pariisi-a$_1$ kohti __$_1$]. (#10)
Seine.NOM flow-PRS.3SG Paris-PAR towards
'Seine flows towards Paris.'

c. [kohti Pariisi-a]$_1$ Seine virta-a __$_1$. (#11)
towards Paris-PAR Seine.NOM flow-PRS.3SG
'Seine flows towards Paris.'

d. [Pariisi-a$_1$ kohti __$_1$]$_2$ Seine virta-a __$_2$. (#12)
Paris-PAR towards Seine.NOM flow-PRS.3SG
'Seine flows towards Paris.'

These data (#1-24, 29-44) are problematic for our analysis, which claims that only operators can create Ā-chains. Furthermore, these chains are not associated with operator interpretations. If we embed (68)c into a sentence with an independent operator, the result is (69).

(69) Milloin [kohti Pariisia]$_1$ Jari matkust-i __$_1$?
when towards Paris-PAR Jari.nom travel-PST.3SG
'When did Jari (2nd topic) travel towards Paris (1st topic)?'

This sentence is felicitous in a context where both Paris and Jari are well-known from prior discourse and constitute the topics of the discussion. Double topic readings are a well-known phenomenon of Finnish (Vilkuna 1989; Brattico 2021a). One possible line of analysis is to assume that sentences of this type involve 'topic operators' and 'topic scope calculations'. This could be supported by the observation, discussed below, that in Hungarian left peripheral topics generate Ā-chains (#428-432). Accordingly, the domain of Ā-chains would contain operators and topics. This hypothesis does not come without problems, however. If we assumed that *all* phrases in the extended specifier positions trigger Ā-chains, irrespective of whether they contains operators or not, then the modelling of A-chains becomes an issue. We must provide the algorithm with access to the distinction between topic and A-chains, a nontrivial (although not unsolvable) matter since the two do not differ in terms of any overt marking on the moved phrase. But there is a more difficult problem: Finnish exhibits a productive and systematic topicalization and long-distance rightward movement that cannot be captured by using standard A/Ā-chains, see also example (41). This presents a further (and to us essentially unsolvable) challenge to any model which proposes the handle all phrasal movement by means of Ā-chains. To handle long-distance rightward movement, the algorithm forms scrambling/topicalization chains. This opens up the possibility of modelling Finnish topicalization as scrambling. Indeed, we tested this option successfully and in the final simulation trial assumed that these sentences were, indeed, processed by scrambling. As a consequence, no operator heads/scope positions were created in (69), and the fronted PP is correctly analysed as occupying the second specifier/topic position of TP.[26] Furthermore, when it comes to the present dataset, the output of the scrambling chains were identical to what would have been generated by Ā-chains (see #1-24, 29-44).

The issue is complicated, however, by the Hungarian topic phenomenon. The Hungarian finite clause contains a dedicated topic position that occurs to the left of the operator (focus) positions. This position is absent in Finnish and English. These sentences were analysed by the algorithm as (70), where the topic was part of an Ā-chain, following the analysis by (Kiss 2002).

[26] The number of specifiers any given head can project is not limited specifically by any grammatical principle. Double specifier constructions, such as Finnish multitopic constructions, can be derived if other grammatical principles permit them. The number of specifiers can be limited by lexical features and by other principles, such as the projection principle. Several simulation trials showed that the issue is nontrivial, different assumptions leading to different trade-offs in the data coverage.

(70) [$_{TopicP}$ Subj$_1$ [Top^0 [$_{FP}$ . . . [$_{TP}$ . . . __$_1$ . . .]]]]

The chain (Subj$_1$,__$_1$) is an Ā-chain according to (Kiss 2002), which is again a problem for the present analysis because it is not triggered by an operator. Similar issues would likely arise if we had nontrivial examples of the Italian left periphery (Rizzi 1997) in our dataset. The question of what exactly characterizes all phrases that participate in Ā-chains now becomes much less straightforward, however. The simulation revealed that the topic issue has no trivial solution. There is compelling *overlap* between topicalization and Ā-chain formation, but full unification does not seem easy to accomplish. The current dataset was derived by assuming that all overt operators and Hungarian left peripheral topics create Ā-chains, while the Finnish data such as (68) were calculated by relying on topicalization/scrambling. The issue could be studied further by composing a crosslinguistic dataset specifically designed to capture the topic phenomenon.

5.3 The problem with genitive subjects

Early simulation trials with the analysis provided in Section 3.8 revealed a problem with (71)a.

(71) a.	*lähellä-ni	minun	b.	minun	lähellä-ni	(Finnish)
	near-1SG	I.GEN		I.GEN	near-1SG	
	Intended: ’near me’					

The genitive argument of an agreeing adposition *lähellä-ni* ‘near-1SG’ must be in the pre-adposition position, while the initial analysis (the one defined in Section 3.8) judged (72a-b) grammatical. After the edge feature of P has been checked by the agreement cluster at the adposition, the analysis had no way to control the position of the optional genitive argument that occurs in a canonical position in the input, leading the algorithm to accept (71)a.[27] This is a situation the present analysis clearly predicts should *not* exist. It corresponds to a language with strong finite agreement and mandatory SVO order. However, there was another class of sentences that was also calculated wrongly and had similar properties: the model judged (72) wrongly as grammatical.

(72) !*Pekka	käsk-i	__$_1$	lähte-ä	Merja-n$_1$. (#427)
Pekka.NOM	order-PST.3SG		leave-A/INF	Merja-GEN
Intended: ‘Pekka ordered Merja (=new information) to leave.’				

[27] Because the genitive argument is optional, adding [!EF$_\varphi$] to the adposition does not solve the problem. It is possible to form a grammatical sentence by using the adposition without any argument (e.g., *Pekka asui lähellä* ‘Pekka lived near’). Some adpositions allow both word orders, such as *sillan yli* ‘bridge.GEN over’ ~ *yli sillan* ‘over bridge.GEN’ (#13-16), so the problem concerns only some adpositions and must rely on a lexical feature.

In both cases, the model fails to react to what looks to be a genitive argument at an illicit rightward position. Note that nominative, partitive or accusative arguments can occur in these rightward positions, and that the genitive argument can still move to the left (73).

(73) a. *Interrogativization of the embedded genitive subject*
Kenen Pekka käsk-i $\_\__1$ lähte-ä?
who.GEN Pekka.NOM order-PST.3SG leave-A/INF
'Who did Pekka order to leave?'

b. *Topicalization of the embedded genitive subject*
Merja-n käsk-i Pekka $\_\__1$ lähte-ä.
Merja-GEN order-PST.3SG Pekka.NOM leave-A/INF
'When it comes to Merja, Pekka ordered her to leave.'

These data point to a potential weakness in our analysis. This phenomenon was ultimately captured by forcing some heads to host a copy of the extended subject at their specifier positions if it existed locally inside the complement. Since the phenomenon was marginal in the original dataset, and further examples were difficult to find from the languages targeted for analysis, it is difficult to guide further hypothesizing. To study this phenomenon in detail, one would ideally need data from languages with strong agreement and SVO order. We left also this issue for future research.

6 Conclusions

The observation that Finnish involves considerable amount of word order inversion in chain formation has puzzled researchers in this field since the original discovery by Huhmarniemi (2012). An explanation for the phenomenon as well as its broader crosslinguistic significance have remained unknown. We argued that the phenomenon is related to agreement. Lexical items that trigger the relevant word order inversions in Finnish and Hungarian are all capable of exhibiting agreement, whereas the same heads in languages like English or Italian do neither. Since it is well-known that agreement is related to the formation of A-chains, these observations suggest that agreement is related to chain formation in a more general way. We proposed an analysis according to which lexical items may have a lexical edge feature licensing, mandating or blocking the presence of extended nonthematic subjects, where the notion of "extended subject" covers both overt phrasal specifiers and (rich) agreement clusters. Control of chain formation was relinquished to the edge feature.

References

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

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

Brattico, Pauli. 2019a. Subjects, topics and definiteness in Finnish. *Studia Linguistica* 73: 1–38.

Brattico, Pauli. 2019b. *Computational implementation of a linear phase parser. Framework and technical documentation (version 14.0)*. Pavia.

Brattico, Pauli. 2021a. A dual pathway analysis of information structure. *Lingua* 103156.

Brattico, Pauli. 2021b. Null arguments and the inverse problem. *Glossa: A journal of general linguistics* 6: 1–29. https://doi.org/https://doi.org/10.5334/gjgl.1189.

Brattico, Pauli. 2022. Predicate clefting and long head movement in Finnish. *Linguistic Inquiry* 54: 663–692. https://doi.org/http://dx.doi.org/10.1162/ling_a_00431.

Brattico, Pauli, and Cristiano Chesi. 2020. A top-down, parser-friendly approach to operator movement and pied-piping. *Lingua* 233: 102760. https://doi.org/10.1016/j.lingua.2019.102760.

Cann, Ronnie, Ruth Kempson, and Lutz Marten. 2005. *The Dynamics of Language: An Introduction (Syntax and Semantics, Volume 35)*. Amsterdam: Elsevier Academic Press.

Chesi, Cristiano. 2012. *Competence and Computation: toward a processing friendly minimalist Grammar*. Padova: Unipress.

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

Chomsky, Noam. 2008. On Phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, ed. Carlos Otero, Robert Freidin, and Maria-Luisa Zubizarreta, 133–166. Cambridge, MA.: MIT Press.

Cinque, Guglielmo. 1982. On the theory of relative clauses and markedness. *Linguistic Review* 1. De Gruyter Mouton: 247–296. https://doi.org/10.1515/TLIR.1982.1.3.247/MACHINEREADABLECITATION/RIS.

Dikken, Marcel den. 1999. On the structural representation of possession and agreement: the case of (anti-)agreement in Hungarian possessed nominal phrases. In *Crossing boundaries : advances in the theory of Central and Eastern European languages*, ed. I. Kenesei, 137–178. John Benjamins.

Gröndahl, Tommi. 2015. Määräisyys funktionaalisena pääsanana suomen kielen nominilausekkeessa. University of Helsinki.

Heck, Fabian. 2008. *On Pied-Piping: Wh-Movement and Beyond*. Amsterdam: Mouton de Gruyter.

Heck, Fabian. 2009. On Certain Properties of Pied-Piping. *Linguistic Inquiry* 40: 75–111.

Holmberg, Anders. 2005. Is There a Little pro? Evidence from Finnish. *Linguistic Inquiry* 36: 533–564.

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

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

Holmberg, Anders, Urpo Nikanne, Irmeli Oraviita, Hannu Reime, and Trond Trosterud. 1993. The Structure of INFL and the Finite Clause in Finnish. In *Case and other functional categories in Finnish syntax*, ed. Anders Holmberg and Urpo Nikanne, 177–206. Mouton de Gruyter.

Holmberg, Anders, and Michelle Sheehan. 2010. Control into finite clauses in partial null-subject languages. In *Parametric Variation: Null Subjects in Minimalist Theory*, ed. Theresa Biberauer, Anders Holmberg, Ian Roberts, and Michelle Sheehan, 125–152. Cambridge: Cambridge University Press.

Huhmarniemi, Saara. 2012. Finnish A'-movement: Edges and Islands. Helsinki: University of Helsinki.

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

Huhmarniemi, Saara, and Pauli Brattico. 2013. On primary and secondary movement. *Acta Linguistica Hungarica* 60: 173–216.

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

Kanerva, Jonni. 1987. Morphological integrity and syntax: The evidence from Finnish possessive suffixes. *Language* 63: 498–501.

Kempson, R., W. Meyer-Viol, and D. M. Gabbay. 2001. *Dynamic syntax: The flow of language understanding*. UK: Wiley-Blackwell.

Kiss, Katalin É. 2002. *The Syntax of Hungarian*. Cambridge: Cambridge University Press.

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

Miyagawa, Shigeru. 2010. *Why Agree? Why Move? Unifying Agreement-Based and Discourse-Configurational Languages*. Cambridge, MA.: MIT Press.

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

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

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar*, ed. Liliane Haegemann, 289–330. Amsterdam: Kluwer.

van Steenbergen, Marlies. 1987. Binding Relations in Finnish. Univ. of Groningen.

Szabolcsi, Anna. 1983. The possessor that ran away from home. *The Linguistic Review* 3: 89–102.

Szabolcsi, Anna. 1994. The Noun Phrase. In *The Syntactic Structure of Hungarian*, ed. Ferenc Kiefer and Katalin É Kiss, 27:179–274. Cambridge, MA.: Academic Press.

Taraldsen, Knut. 1978. On the NIC, Vacuous Application and the that-trace filter. *Unpublished manuscript, MIT*.

Toivonen, Ida. 2000. The Morphosyntax of Finnish Possessives. *Natural Language and Linguistic Theory* 18: 579–609.

Vainikka, Anne. 1989. Deriving Syntactic Representations in Finnish. University of Massachusetts Amherst.

Vainikka, Anne, and Yonata Levy. 1999. Empty subjects in Finnish and Hebrew. *Natural Language & Linguistic Theory* 17: 613–671.

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

Vilkuna, Maria. 1995. Discourse Configurationality in Finnish. In *Discourse Configurational Languages*, ed. Katalin É. Kiss, 244–268. Oxford: Oxford University Press.

---

* https://we.tl/t-k901AeT8PY

† Original dataset: https://we.tl/t-jDXdILxKKm

‡ They are available for review as anonymous files and can be downloaded from the following links: the original dataset (https://we.tl/t-jDXdILxKKm), grammaticality judgments generated by the model (https://we.tl/t-Pm5e2qHm7j), results file (https://we.tl/t-JSGB1cUTp8) and the phrase structure images as a zip file (https://we.tl/t-LxssqzTGhU). The derivational log file was not uploaded.

[§] Original dataset: https://we.tl/t-jDXdILxKKm

[**] A draft supplementary document, which is not part of this initial submission and need not be reviewed or evaluated, can nevertheless be downloaded (https://we.tl/t-k901AeT8PY). It addresses technical issues and a few details concerning the underlying algorithm.