# BACKWARD DEPENDENCIES MUST BE SHORT
# A unified account of the Final-over-Final and the Right Roof Constraints and its consequences for the syntax/morphology interface

Carlo Cecchetto
University of Milan-Bicocca
carlo.cecchetto@unimib.it

(A slightly revised version of this paper will appear in a Mouton volume edited by Theresa Biberauer and Ian Roberts, which collects the papers presented at the GLOW 2008 workshop on Linearization)

# BACKWARD DEPENDENCIES MUST BE SHORT.
## A unified account of the Final-over-Final and the Right Roof Constraints and its consequences for the syntax/morphology interface*

## 1. Introduction

This paper argues for a unification of the Final-Over-Final-Constraint (FOFC) and of the Right Roof Constraint. The Right Roof Constraint is a well established condition that makes rightward movement very local, while FOFC, which has been proposed more recently, imposes a limit on how disharmonic languages can be, by stating that if α is a head-initial phrase and β is a phrase immediately dominating α, then β must be head-initial (on the other hand, if α is a head-final phrase, and β is a phrase immediately dominating α, then β can be either head-initial or head-final).
I will propose that FOFC and the Right Roof Constraint are two faces of the same coin, because they both exclude the same abstract configuration which involves backward localization, either of a trace (in the case of the Right Roof Constraint) or of the selected head by a selecting head (in the case of FOFC). I will assume that, since backward localization is costly for the syntactic parser, which proceeds left-to-right, backward localization is possible only if it is very local, namely phrase-internal. I will propose that preference for a parser-friendly structure is grammaticalized as a locality condition stating that backward dependencies must be shorter than forward dependencies. My approach has consequence for the syntax/morphology interface. I will show that some morphological properties commonly found in head-final languages can be seen as strategies to cope with the fact that head selection is backward in these languages.
This paper is organized as follows: in section 2, I briefly discuss the status of the Head Parameter in the light of recent typological findings. In section 3, I switch to disharmonic languages and summarize the motivation offered for FOFC in the existing literature. In section 4, I make explicit what FOFC and the Right Roof Constraint have in common, namely that they impose that backward dependencies must be very local. In section 5, mainly by focusing on rightward movement, I propose that the notion of locality relevant for FOFC and for the Right Roof Constraint is the same, namely phrase-boundness. In section 6, I discuss an account that has been proposed for the Right Roof Constraint and I extend it to FOFC. Section 7 is an interlude in which I propose a definition of phrase in Bare Phrase Structure Theory. This technical implementation becomes important in section 8, which is devoted to discussing some apparent counterexamples to FOFC. In section 9 I discuss some consequences of my approach for head-final languages and I comment on the syntax/morphology interface. Section 10 concludes the paper.

## 2. The Head Parameter

Harmonic languages are those in which the linear order between a head and its complement is consistent through the various categories of the sentence. Given this informal characterization, English is harmonic, since a preposition precedes a DP, the verb precedes its complement, a complementizer precedes an embedded clause, and so forth and so on for all syntactic categories. German, on the other hand, is disharmonic, as exemplified by the configuration in which a complementizer precedes an embedded clause in which the verb follows its complement.
It is fair to say that the existence of disharmonic languages is somewhat disturbing for supporters of the Principles and Parameters framework, since these languages introduce some challenge for a strong and strict version of the Head Parameter, probably the textbook example of a parameter. In fact, as the simple story goes, the child acquiring a language might take advantage of the fact that syntactic categories are exemplifications of the same abstract schema (say, Specifier-Head-Complement as captured by X-bar theory) and might generalize ordering information from one category (plausibly, the most frequent one) to all others. Given this simplistic picture, harmonic languages should be "easier", because they do not require any revision of the initial guess the child makes.
Still, disharmonic languages do exist and they even outnumber harmonic ones, according to Dryer (1992), who collected a sample of 625 languages. This may lead us to reject the Head Parameter altogether. One famous example of this attitude is Kayne's (1994) antisymmetric theory of phrase structure, in which all languages are taken to be underlyingly SVO. I won't assume Kayne (1994) theory, though. I do that for several reason. One reason has to do with the fact that a corollary of Kayne's theory is that rightward movement should not exist. I argued elsewhere that genuine cases of rightward movement are attested and that the attempt to simulate the effects of rightward movement in the antisymmetric framework (i.e. via remnant movement) fail, at least in the case of rightward wh-movement in sign languages (cf. Cecchetto, Geraci and Zucchi 2009).

* I thank the audience of the GLOW workshop on linearization held in Newcastle in April 2008 and two anonymous reviewers for useful comments and for pointing out various inadequacies that I hope to have partially remedied in this version of the paper.

Another reason that prevents me from assuming the antisymmetric framework is that there is evidence showing that the original motivation for the Head Parameter persists. Dryer, in the same typological work, asks whether it is possible to identify pairs of grammatical elements whose order correlates with the order of verb and object and proposes the following definition:

(1) If a pair of elements X and Y is such that X tends to precede Y significantly more often in VO languages than in OV languages, then <X,Y> is a Correlation Pair and X is a Verb Patterner and Y is an Object Patterner with respect to this pair.

Interestingly, Dryer could identify several Correlation Pairs (some of them had already been recognized as such in the Greenbergian tradition, cf. Greenberg 1966). The following is a representative list:

(2) Correlation Pairs from Dryer (1992)

i. Adpositions and NPs (to + John),
ii. Noun and genitive (father + of John),
iii. Verb and (subcategorized) adpositional phrases (slept + on the floor),
iv. Copula and predicate (be+ a teacher),
v. Tense/aspect auxiliary verb and VP (has + eaten dinner),
vi. Complementizer and embedded clause (that + John is sick).

The World Atlas of Language Structures Online (WALS Online), which contains updated information based on typological research, can confirm the robustness of some of these correlation pairs. For example, Dryer (2008) reports the data in (3), which is based on 892 languages in which the relevant information could be collected. It clearly emerges that a language is overwhelmingly OV if and only if it is postpositional:

(3)

| | |
|---|---|
| Object-verb and postpositional | 427 |
| Object-verb and prepositional | 10 |
| Verb-object and postpositional | 38 |
| Verb-object and prepositional | 417 |

It seems fair to summarize recent typological research by saying that it confirms the existence of a correlation between the order of verb and object and the order of other pairs of elements. Furthermore, in all cases reported in (2) it is plausible that the Object Patterner in the Correlation Pair is a complement of the Verb Patterner. Interestingly, the main difference between Dryer's (1992) investigation and the former typological literature is that Dryer can show that noun and adjective do not form a Correlation Pair[1], unlike what was initially assumed in the Greenbergian tradition (that noun and adjective are not a correlation pair is confirmed by consultation of WALS Online at the time of writing). This is telling, since it is clear that noun and adjective are not in a head-complement relation. So, correlation pairs survive scrutiny based on large samples, if the members of these pairs are plausible head-complement pairings.
We can summarize the current typological research by saying that the principles driving linearization, whatever their precise formulation turns out to be, have a bias to favour a consistent linear order across categories. After all, this is core idea underlying the Head Parameter. It seems to me that a general theory of linearization should explain why harmonic structures are favoured. In the antisymmetric framework, while harmonic head-initial languages are expected, harmonic head-final languages are much less so. I take this to be a drawback of this framework.[2]

[1] This pattern is particularly interesting since relative clauses, which, like adjectives, seem to be adjoined to nouns, do form a Correlation Pair with the noun they modify (Dryer 1992 and WALS Online). Space considerations do not allow me to deal with this interesting fact. However, I can mention that the raising analysis of relatives clauses, which assumes a tighter link between the noun and the relative clause in which it originates, might be better suited to cope with this typological finding than the more standard analysis that take relatives to adjoin to the noun, much as adjectives do (cf. Bianchi 2002 for a survey on different approaches to relative clauses).

[2] This is not to say that it is impossible to derive consistent head-final ordering in the antisymmetric framework. In fact, Kayne (1994) and supporters of the antisymmetric framework have proposed various ways to derive this type of ordering. On this, see footnote 10 below. However, all these devices involve complex movement operations that, unless independently motivated, have the flavor of ad hoc moves.

I am aware of at least two (non mutually incompatible) explanations for why harmonic structures should be favoured. The first is in terms of language acquisition. Nespor, Guasti and Christophe (1996) claim that the child develops a preference for harmonic structures due to the need to maintain a parallelism between phonological and syntactic structures.
Another approach (cf. Hawkins 1994) explains word order on the basis of processing considerations. In Hawkins's model, consistent head-initial and consistent head-final languages are favoured because they optimize the parsing of the sentence.
This paper does not aim at offering a general theory of linearization, so I will now directly switch to the core of the paper, namely disharmonic structures. I hope that, as is common from other disciplines, 'pathology' (disharmonic structures, in the case at hand) can give us some useful information about what is more physiological (harmonic word order). I will offer a final remark on linearization in the conclusion, when my account of disharmonic structures will have been spelled out.

## 3. Harmonic and disharmonic languages: the Final-over-Final Constraint (FOFC)

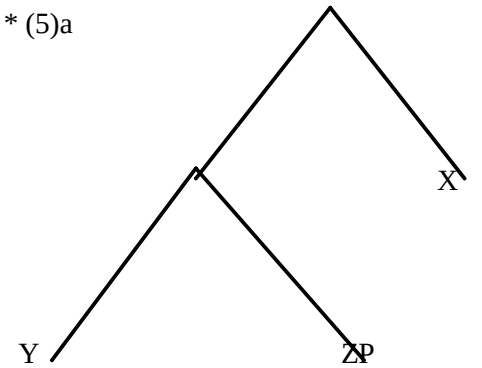
An interesting perspective on disharmonic structures is investigating how disharmonic they can be, in order to see if there is some inner regularity in what might seem a completely free word order system. Recently, this investigation has been systematically carried out in a series of important papers by Theresa Biberauer, Anders Holmberg and Ian Roberts (but see Edmonds 1976 and Hawkins 1994 for earlier discussion). In this paper, I will base my discussion mainly on crosslinguistic data reported in Biberauer's et al. work. For this reason, I will summarize their main findings in this section.
Holmberg (2000), by focusing on Finnish, points out a specific configuration that does not seem attested. This configuration is excluded by FOFC in (4):

(4) Final-over-Final Constraint (FOFC)
If α is a head-initial phrase and β is a phrase immediately dominating α, then β must be head-initial.
If α is a head-final phrase, and β is a phrase immediately dominating α then β can be head-initial or head-final. (from Holmberg 2000, cf. also Julien 2002)

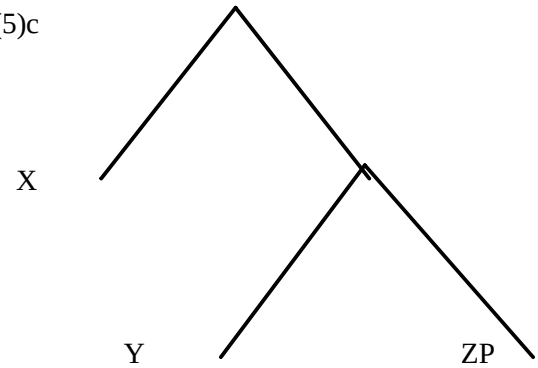
The Final-over-Final Constraint is schematized in (5). Out of the four logically possible arrangements, only three are attested. The illicit arrangement is (5a), in which YP is the head-initial complement of the head-final projection XP (for simplicity, specifiers are omitted in the diagrams in 5):

(5) a. * $[_{XP}\ [_{YP}$ Y ZP ] X ]
b. √ $[_{XP}\ [_{YP}$ ZP Y] X ]
c. √ $[_{XP}$ X $[_{YP}$ Y ZP ] ]
d. √ $[_{XP}$ X $[_{YP}$ ZP Y] ]

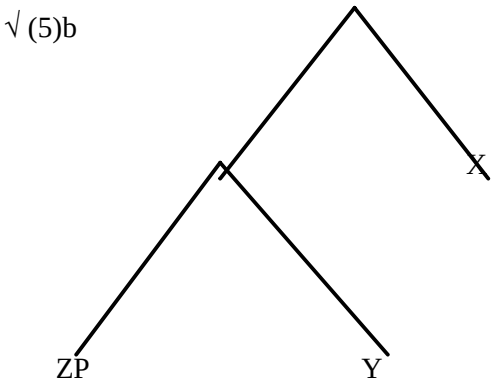
* (5)a



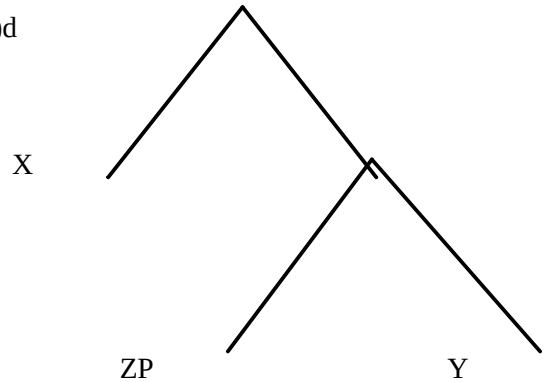

√ (5)c




√ (5)b




√ (5)d



Biberauer, Holmberg and Roberts (2008a and 2008b) mention several independent pieces of evidence supporting FOFC. In (6) below, I mention some of them, but I refer to their papers for data presentation and for a more complete discussion:

(6) i) Old and modern Germanic varieties exhibit a mix of head-initial and head-final orders in VP and TP, with all permutations of Aux, V and Object attested except one: V-Object-Aux. This configuration is the one that violates Final-over-Final Constraint.

ii) Similarly, in Finnish and Basque, the FOFC violating V-Object-Aux configuration is out while the three other permutations, namely Aux-V-Object, Aux-Object-V and Object-V-Aux are attested.

iii) Sentence-final complementizers are unattested in VO languages (cf. Dryer 2009). The underlying configurations that might generate the unattested order are:
$[_{CP} [_{TP} [_{VP}$ V Obj ] T ] COMP ]
and
$[_{CP} [_{TP}$ T $[_{VP}$ V Obj ]] COMP ]
Crucially, they both violate FOFC.

iv) In the nominal domain, Finnish has mixed projections too: it has both pre- and postpositions and N-Complement as well as Complement-N order. All permutations are found except the $[_{PP} [_{NP}$ N-Complement ] P] one, which violates FOFC.

v) FOFC makes two very strong diachronic predictions (cf. Biberauer, Newton and Sheehan, to appear a. and to appear b.). Firstly, the change from head-final to head-initial order should proceed "top-down", with C being the first category switching from head-final to head-initial. Otherwise the language should have gone through a stage in which FOFC is violated. Conversely, change from head-initial to head-final order should proceed "bottom-up", with V being the first category to become head-final. As far as diachronic data are available, these predictions seem to be borne out.

Biberauer et al. (2008a and 2008b) also discuss some (potential) counterexamples to FOFC, which I will examine in section 8.
I would like to add a further piece of evidence to the ones discussed by Biberauer et al. In Italian Sign Language (Lingua dei Segni Italiana, LIS), as in many other sign languages, a peculiar sociolinguistic situation occurs, since the majority of signers are non-native. This is so because less than 10% of signers have deaf parents or siblings. While this minority group of signers have usually been exposed to LIS from birth, the other signers had access to sign language at a later time, after they had some exposition to Italian (typically via lipreading). This sociolinguistic situation may have had an impact on word order in LIS in ways that are interesting for our problem.
Word order in LIS is relatively flexible, but with interesting limitations. In particular, native signers consistently prefer the SOV order in (7a) over the SVO order in (7b) (cf. Cecchetto, Geraci and Zucchi 2006 and 2009). However, non native signers use the SVO order illustrated in (7b) more productively, possibly due to the influence of the word order of Italian. For reader's convenience, glosses of LIS sentence are given in English (as is standard practice, the glosses for signs are in capital letters).

(7) a. GIANNI CONTRACT SIGN
b. GIANNI SIGN CONTRACT
"Gianni is signing the contract"

LIS has an aspectual marker (glossed as DONE), a negative sign (glossed as NOT) and a modal verb (glossed as CAN) that consistently occur post-verbally.

(8) GIANNI CONTRACT SIGN DONE ("Gianni signed the contract)
(9) GIANNI CONTRACT SIGN NOT ("Gianni does not sign the contract")
(10) GIANNI CONTRACT SIGN CAN ("Gianni can sign the contract")

Crucially, when the sentence contains these post-verbal signs, the SVO order becomes ungrammatical even for signers that would normally use it.

(11) * GIANNI SIGN CONTRACT DONE
(12) * GIANNI SIGN CONTRACT NOT
(13) * GIANNI SIGN CONTRACT CAN

The ungrammatical sentences in (11)-(13) are an instance of the configuration that is excluded by FOFC. For example, VP would be the head-initial complement of the head-final projection TP:

(12) * [$_{TP}$ GIANNI [$_{VP}$ SIGN CONTRACT] DONE ]
(13) * [$_{TP}$ GIANNI [$_{VP}$ SIGN CONTRACT] CAN ]

Data in (11)-(13) are very suggestive, since they indicate that what is wrong with the FOFC violating configuration is a problem powerful enough to apply to the grammar of non native speakers/signers.[3]
It should be clear that FOFC is a descriptive generalization that as such should follow from some more basic principles. Biberauer et al. discuss several ways to do this. Here I report Biberauer's et al. (2009) attempt. They assume that linearization proceeds in accordance with Kayne's (1994) Linear Correspondence Axiom (LCA). Since phrase-structure is uniformly head-initial in Kayne's system, some mechanism must be introduced to derive the head-final order. Biberauer et al. assume that any head may have a linearization feature which forces head-final order. They then derive FOFC from the condition in (14):

(14) a. As a parametrized property, any head may have a diacritic ^ indicating head-final order.
b. If a phase-head has the diacritic ^, then ^ must spread to all heads on the projection line associated with that head.

(14) straightforwardly explains what is wrong with the FOFC violating configuration. Take the V-Object-Aux order, which is unattested. For this order to obtain, the head T should have the diacritic ^ but the diacritic ^ should not spread on the lower verbal head, in violation of (14b). Similarly for the other FOFC violating configurations.
Several considerations prevent me from adopting Biberauer's et al. (2009) account. First, the insertion of the feature ^ in the derivation violates Chomsky's (1995) Inclusiveness Condition, according to which narrow syntax merely operates on lexical items and cannot "add" extra material (but see Biberauer et al. 2009 for some discussion of this point). Second, in Biberauer's et al. approach, head-final order is more costly than head-initial order, since the former requires the extra operation of inserting the diacritic ^ and spreading it to lower heads in the projection line. The available data do not seem to support this, though. In WALS Online database, the number of SOV languages exceed the number of SVO languages. Furthermore, Julien (2001:18) reports that nine out of ten OV languages in Dryer's (1992) sample are 'canonical', meaning that every clausal head follows its complement in the surface syntax. So, it does not seem that there is an inherent cost associated with head-finality.
Third, as I am going to argue extensively in the next section, FOFC is a condition that strongly resembles a well known constraint on rightward movement, namely the Right Roof Constraint. A condition like (14) cannot capture the striking similarity between FOFC and the Right Roof Constraint. For this reason, in section 6 I will take an alternative path and extend to FOFC an account that has been originally proposed for the Right Roof Constraint. This way, both conditions will be both derived from the same underlying principle, as it is desirable.

## 4. What the Right Roof Constraint and FOFC have in common

Starting from the seminal work of Ross (1967), it has been repeatedly observed that there is a systematic difference between leftward and rightward movement. While in principle a category can leftward move over any distance, provided that it respects familiar locality constraints, rightward movement obeys the Right Roof Constraint in (15).

(15) Right Roof Constraint
An element cannot move rightward out of the clause in which it originates.
(Preliminary formulation)

Sentences (16) and (17), in which Heavy NP shift applies, are among the motivation that have been offered for the Right Roof Constraint. Since an adverb like "dearly" modifies the main verb, the heavy NP has moved out of the embedded clause in (17), resulting in an ungrammatical output. (16), on the other hand, is grammatical because rightward movement is clause-bound (the adverb "carefully" modifies the embedded verb).

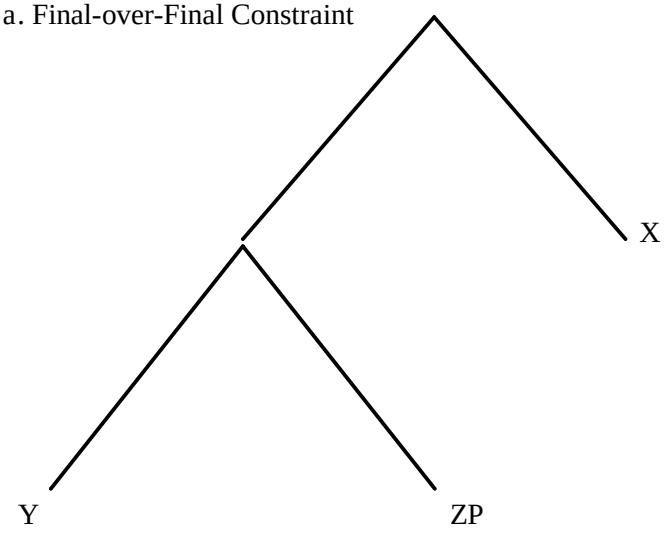
[3] An anonymous reviewer observes that, if markers like NOT, DONE and CAN were verbal affixes or clitics, the ungrammaticality of (11)-(13) could be explained with no need to refer to FOFC, since NOT, DONE and CAN are separated from the verb in these sentences. It is unlikely that NOT, DONE and CAN are verbal affixes, though. I refer to Geraci (2006) and to Zucchi (2009) for extensive discussion of the properties of NOT and DONE. Here I can offer a simple piece of evidence that shows that NOT is not an affix. In sentences like (i), a temporal adverb can occur between the verb and NOT (I thank Anna Folchi for providing this information).
(i) GIANNI ARRIVE ON-TIME NOT
"Gianni did not arrive on time"

(16) John wanted to [study t carefully [the entire book of Revelation]]
(17) *[John wanted to [study t ] dearly [the entire book of Revelation]]]
(from Rochemont e Culicover 1990)

We will discuss the evidence for the Right Roof Constraint and its status shortly. Before doing that, I would like to stress the analogy between this constraint and FOFC. The schematic representation in (18) can illustrate this analogy quite well.

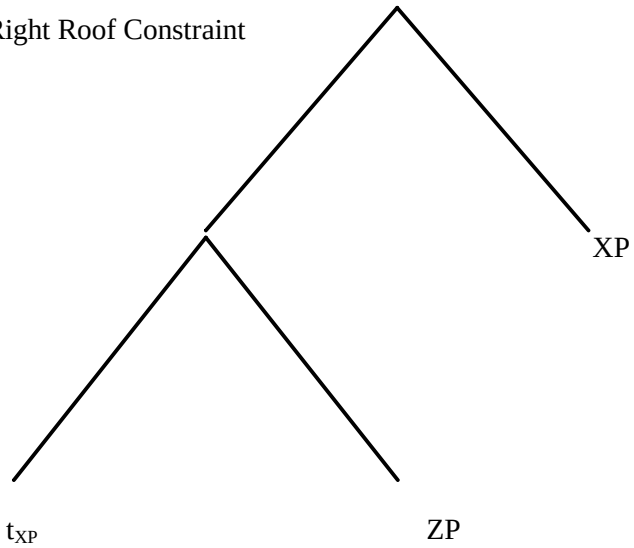
(18)

a. Final-over-Final Constraint




X selects Y but follows it

b. Right Roof Constraint




XP is the antecedent of t but follows it

What the configuration excluded by FOFC and the one excluded by the Right Roof Constraint have in common is that they both involve a dependency in which one link (the selected category in 18a and the trace/copy in 18b) depends on a category that linearly follows it. So, the problem with the configurations in (18) seems to be that there is an asymmetric dependency in which the dependent link linearly precedes the one it depends on.
Let us further elaborate on the parallelism between the dependency between a trace/copy and its antecedent and the dependency between a selecting and a selected head.
FOFC applies within the boundary of the clausal domain, the skeleton of which is the C-T-V heads. The heads inside this domain stay in a special relation, with the c-commanding head not only selecting the c-commanded one, but also sharing tense and agreement features with it. That COMP "agrees" with T is shown by the fact that the tense/untensed character of the clause is reflected at the COMP level. That T "agrees" with the V is clearly indicated by cases in which the verbal inflection is not realized by an independent auxiliary. In this case, T hosts a bound morpheme that V eventually attaches to.
It is reasonable to assimilate the link between COMP and T and the one between T and V to the link between a moved category and its trace. That this assimilation is legitimate is suggested by at least two considerations:
(i) COMP and T and T and V are actually linked by the occurrence of head movement in many languages (cf. German, English and other Germanic varieties for T-to-COMP and French, Italian and other Romance varieties for V-to-T)
(ii) There is an asymmetry between COMP and T and between T and V, as there is an asymmetry between a moved category and its trace.
While the point in (i) does not deserve much discussion, it is worth it is worth elaborating on (ii). It is obvious that there is no trace/copy without an antecedent. A similar asymmetry can be shown in the clausal domain. Each sentence displaying COMP will have a T layer as well. However, T does not warrant the existence of the CP area, as shown by is Exceptional Case Marking configurations like (19) in which the CP layer is absent (under standard assumptions):

(19) I believe John to be smart.

In the lowest section of the clause, the same asymmetry is found between T and V. The T layer implies the presence of the VP, but a VP does not warrant the existence of the (entire) TP area above it. This is shown by structures like absolute small clauses (cf. 20) and reduced relative clauses (cf. 21), illustrated here with Italian examples (cf. Belletti 1990):

(20) Mangiata la mela, Gianni morì.
Eaten the apple, Gianni died
'Having eaten the apple, Gianni died'
(21) La mela mangiata da Gianni era avvelenata.
'The apple eaten by Gianni was poisoned'

All in all, the dependency between a C and T and the one between T and V closely resemble the link between a moved category and its trace/copy. In all these cases, there is an asymmetric dependency in which one link depends on the other. If the dependent link (trace, selected head) follows the category it depends on, no problem arises. However, if the dependent link precedes the category it depends on, the configuration is more problematic (as will see, it is acceptable only if the dependency is very local). For the sake of clarity, I will refer to a structure in which the "dependent link" precedes the link it depends on as an "inverted dependency".
Having said this, we can see what FOFC and the Right Roof Constraint have in common: they impose tight limits on inverted dependencies. This is the intuition that I will try to make more precise in the rest of this paper. In order to do that, we need to discuss the Right Roof Constraint in some more detail.

## 5. How much can an inverted dependency extend?

I have assumed that inverted dependencies are problematic (in section 6, I will explain why they are, reviving a traditional parsing account). For the time being, we can observe that there seems to be a difference between FOFC and the Right Roof Constraint. FOFC bans inverted dependencies altogether, while the Right Roof Constraint allows them as long as they are short (clause bound, in the preliminary definition in 15). We need to closely examine this point to see if a substantial difference is at stake here. Therefore, in this section I will review the literature which investigates how far rightward movement can go, in order to double check if this difference is real.
Although Ross' initial version of the Right Roof Constraint, reported in (15), states that rightward movement is clause bound, later research, especially by Baltin (1981) and Rochemont and Culicover (1990), indicates that rightward movement might be more local than this, at least in languages like English. In fact, it has been proposed that rightward movement might be phrase bound (this was in a framework in which clausal structures was composed by CP, IP and VP[4]).
I will now review some of this evidence. In my presentation, I will use the term "rightward movement" to refer to dependencies in which the category that is displaced linearly follows its canonical position. I will gloss over an important question that has been addressed in the literature, namely whether the inverted dependency is created by an occurrence of movement or it is a base generated one (cf. Baltin 2005 for discussion on this). This question may be important but it is not crucial in the present context, which is the investigation of how far inverted dependencies can go. As we will see, inverted dependencies are tightly constrained, no matter how they are created.
The literature has mainly focused on relative clause extraposition in English, so I will start from them:

### 5.1 The first type of evidence: linear order between extraposed relative clauses

Rochemont and Culicover (1990) discuss the contrast between (22) and (23) and conclude that a relative extraposed from object position is attached closer to the direct object than is a relative extraposed from subject position:

(22) A man entered the room last night that I had just finished painting who had blond hair.
(23) *A man entered the room last night who had blond hair that I had just finished painting.

If linear order corresponds to relative height of attachment, the contrast between (22) and (23) suggests that a relative extraposed from object position sits somewhere in the VP periphery.

### 5.2 The second type of evidence: Condition C effects with extraposed relative clauses in English

Rochemont and Culicover (1990) discuss the pattern in (24)-(26) below (but see Büring and Hartmann 1997 for a critical discussion). The Condition C violation in (24) shows that "her" c-commands the unextraposed relative clause. However, Condition C is obeyed in (25), therefore "her" does not c-command the relative when it is extraposed.

(24) *I sent her$_i$ many gifts that Mary$_i$ didn't like last year.
(25) I sent her$_i$ many gifts last year that Mary$_i$ didn't like.

Crucially, (26) shows that the subject c-commands the extraposed relative. This indicates that extraposition is a local phenomenon (a relative extraposed from the object position sits in a position lower than Spec,T):

(26) *She$_i$ invited many people to the party that Mary$_i$ didn't know.

[4] Assuming the richer structure emerging from cartographic approaches (see Rizzi 1997 and Cinque 1999, among many others) in which the VP, IP and CP areas are segmented in many independent phrases , the generalization that rightward movement is phrase bound should be reformulated by saying that a category cannot rightward move out of the area of the clause in which it originates.

In traditional terms, the pattern in (24)-(26) suggests that the relative extraposed from the object position is adjoined to VP.

5.3 The third type of evidence: VP ellipsis and extraposition from the VP

Baltin (1981) observes that a relative clause that is extraposed from subject position can be stranded when the VP is elided (cf. 27), while a relative clause extraposed from object position cannot be stranded when the VP is elided (cf. 28):

(27) Although not many people would ride with Fred who knew just him, some would __ who knew his brother.
(28) *Although he didn't call people up who are from Boston, he did __ who are from New York.
(From Baltin 1981)

This pattern, as argued by Baltin, can be explained if (i) constituents that are extraposed from subject position are adjoined to IP, while constituents that are extraposed from within objects are adjoined to VP and (ii) when VP is elided, the material adjoined to it cannot be stranded.

5.4 The fourth type of evidence: Condition C effects in Italian clitic right dislocation

The evidence discussed up to now comes just from English. However, that "rightward movement" is tightly bound can be shown by looking at other languages as well. Cecchetto (1999a) discusses an asymmetry between Clitic Left Dislocation and Clitic Right Dislocation in Italian that confirms the special nature of inverted dependencies. (29) is a familiar case of Condition C violation. However, (30), in which the direct object modified by the relative clause had been left dislocated, is much better.

(29) *pro$_i$ smentisce sempre gli annunci che Berlusconi$_i$ dà alla stampa
(he) denies always the announcements that Berlusconi gives to the press
(30) Gli annunci che Berlusconi$_i$ dà alla stampa, pro$_i$ li smentisce sempre
The announcements that Berlusconi gives to the press (he) it denies always

The contrast between (29) and (30) is reminiscent of the contrast between (31) and (32), discussed by Lebeaux (1995) and Chomsky (1995) among many others:

(31) *He$_i$ was willing to discuss the claim that John$_i$ made.
(32) Which claim that John$_i$ made was he$_i$ willing to discuss?

The explanation proposed for English sentences easily extends to the Italian case: adjuncts can be inserted in the derivation later than arguments, so, in (30), the relative clause containing "Berlusconi" is inserted when the DP "gli annunci" has already reached the topic position. As a result, no Principle C violation results.
Cecchetto then considers the right dislocation sentence correspondent to the left dislocation sentence (30), namely (33):

(33) *pro$_i$ li smentisce sempre, gli annunci che Berlusconi$_i$ dà alla stampa.

Crucially, (33), unlike its left counterpart (30), is a Condition C violation. Cecchetto concludes that the minimal pair between (30) and (33) is evidence that the right dislocated DP occupies a position lower than the one of the (null) preverbal subject. So, (clitic) right dislocation involves a very local dependency (the right dislocated category sits somewhere in the right periphery of the VP).

5.5 The final version of the Right Roof Constraint

In this section I discussed four pieces of evidence show that rightward movement is very local. This has led various scholars to propose that a rightward moved phrase is always adjoined to the maximal projection in which it originates:

(34) Right Roof Constraint
An element cannot rightward move out of the phrase in which it originates.
(Second formulation)

In our terms, this means that the unit within which inverted dependencies are tolerated is the phrase. In this paper, I will assume the version of the Right Roof Constraint in (34), but not without introducing two important caveats. The first caveat concerns extraposition in OV languages. In fact, the literature discusses cases of relative clause extraposition that do

not seem to obey (34). One example, from de Vries (1999), is the Dutch sentence (35), in which the extraposed relative crosses at least two DP boundaries and one PP boundary, regardless of where in the V-projection it ends up.

(35) Ik heb [de papieren van de man] gecontroleerd die een rode koffer droeg.
I have [the papers of the man] checked who a red suitcase carried
'I checked the papers of the man who carried a red suitcase'

When confronted with these facts, one possibility is abandoning (34), because it is too restrictive, However, this would leave the English and Italian facts unaccounted. Another possibility, which I will adopt, is sticking to the restrictive formulation and denying that German and Dutch cases are genuine cases of extraposition. In fact, something like that has been proposed for Dutch extraposition. Based on the fact that (alleged) extraposition and coordination share eleven formal properties, de Vries (1999) proposes that the Dutch equivalent of (36a) would be derived from a coordinated structure like (36b) by applying ellipsis as in (36)c (I refer to de Vries for discussion of condition governing ellipsis).

(36) a. A man was noticed who carried a red suitcase.
b. A man was noticed, namely, a man who carried a red suitcase (was noticed).
c. A man was noticed, ~~namely, a man~~ who carried a red suitcase ~~was noticed~~.

A similar concern is raised by the observation that right dislocation is attested in harmonic SOV languages like Japanese (cf. Cecchetto 1999b) and Turkish (cf. Kural 1997) and in Indo-Aryan languages such as Hindi-Urdu and Bangla (cf. Bhatt and Dayal 2007 and references cited therein). Some instances of right dislocation in these languages suggest that rightward movement might not be phrase-bound. For example, a VP internal category can be right dislocated to the right of an auxiliary in (some of) these languages. This is schematized in (37). (37a) indicates the canonical word order in a consistent head-final language, while (37b) indicated the order following right dislocation of the direct object (however, right dislocation is not limited to direct objects). The order in (37b) is problematic for the version of the Right Roof Constraint in (34), since the direct object seems to cross at least the VP boundary.

(37) a. Subject Object V AUX (John Mary watched has)
b. Subject V AUX Object (John watched has Mary)

However, there are analyses that, at least in principle, might be compatible with the idea that right dislocation is phrase-bound, despite initial appearances to the contrary. For example, Bhatt and Dayal (2007), based on the correlation between linear order and scope and on the restricted scope for rightward scrambled wh-expressions, argue that apparent right dislocation of the VP internal material in Hindi should be analyzed as right dislocation of the entire VP, out of which the verb has previously moved (this type of analysis is roughly schematized in 38).

(38) $[_{TP}$ John $t_j$ watched$_i$ has$]$ $[_{VP}$ Mary $t_i]_j$

If Bhatt and Dayal's analysis is on the right track, it might be possible to analyze cases of apparent long right dislocation of a DP/PP as cases of short (phrase-bound) right dislocation of the category that contains the DP/PP. Future research will determine if this analysis is viable and can be extended to other SOV languages.
The second caveat concerns well attested cases of remnant movement configurations like (39).

(39) a. [Fired $t_i$ by the company]$_j$ John$_i$ indeed was $t_j$
b. [Andato $t_i$ a casa]$_j$ Gianni$_i$ non è $t_j$
Gone to house Gianni NEG is
c. [Gelesen $t_i$]$_j$ hat Hans das Buch$_i$ nicht $t_j$
read has Hans the book not

In all these sentences, the configuration typically resulting from rightward movement gets created, since the trace $t_i$ ends up being to the left of its antecedent. Note that $t_i$ and its antecedent are separated by several phrase boundaries, in apparent violation of the version of the Right Roof Constraint in (34). However, these cases are not ruled out by (34), if this is intended as a derivational condition constraining rightward movement. In (39), no rightward movement applies, so (34) does not exclude these sentences, a desirable result.
In this section, I have proposed that rightward movement is phrase bound. I claimed that apparent cases showing the contrary admit (at least in principle) a treatment that does not question the phrase bound character of rightward movement. If I am right, an important consequence follows for my attempt to unify FOFC and the Right Roof Constraint. Unlike what I said at the beginning of this section, FOFC and the Right Roof Constraint are not different. In fact, they both amount to saying that inverted dependencies cannot extend outside the phrase boundary. In the next section I will discuss why inverted dependencies are so tightly constrained.

## 6. An account for why inverted dependencies are phrase bound

In this section I will claim that the condition that forces inverted dependencies to be confined within the phrase can be explained by revamping a classical approach to the Right Roof Constraint in terms of parsing[5].
To the best of my knowledge, this approach was first proposed by Fodor (1978), but see Rochemont (1992) for further elaboration of this idea and Ackema and Neeleman (2002) for an elaborated defence. In this section I summarize Ackema and Neeleman's (2002) approach. They make four crucial assumptions to explain the bound character of rightward movement:
(i) the parser scans the input string from left-to-right
(ii) in analysing an input string, the parser closes off certain units of already parsed structure and removes them from short term memory
(iii) the parser can postulate a trace only after it reaches the antecedent of the trace
(iv) the parser cannot re-open a unit that has been closed and removed from short term memory
Since (i) and (ii) are fairly uncontroversial assumptions, I will not discuss them. The assumption in (iii) is uncontroversial for traces of adjuncts: since traces are phonologically empty, there is nothing in (40) which can tell the parser that the sentence contains a gap. Therefore, the postulation of the trace is dependent on the presence of the fronted wh phrase:

(40) How do you think Mary fixed the bike t?

Admittedly, (iii) is less obvious for traces of arguments. Experimental evidence (Frazier and Flores d'Arcais 1989) supports it, though. The assumption in (iv) needs to be discussed as well, since it might seem incompatible with garden path sentences like "The horse raced past the barn fell", in which the parser seems to be able to "go back" (with some cost). However, as Ackema and Neeleman (2002) point out, there is a fundamental difference between backward localization of a trace and garden path sentences. Garden path sentences are temporarily ambiguous, so the parser can build an alternative parse once it has made an "erroneous" first attempt. What is impossible is altering established relations within a single parse, but this is exactly what backward localization would require.
If conditions (i) to (iv) are assumed, it is possible to explain why inverted dependencies are problematic. Since the parser proceeds left-to-right, rightward movement requires backward localization of a trace. However, backward localization is possible only if trace and antecedent are close enough to belong to the same parsing unit (if they don't, the trace might be in a fully analyzed and, therefore closed, unit by the time the parser reaches the antecedent). Since in leftward movement the antecedent precedes the trace, backward localization is not required. Once an antecedent is identified as a moved element (on the basis of the grammar of the language in question, which the parser uses as a knowledge base), the position where to insert a gap is looked for and the parser will not close a unit until this position is identified. This can derive the virtual unboundness of leftward movement (modulo working memory limitations).
The approach that I have summarized is not the only parsing approach that can explain why (some) inverted dependencies are not admitted. Hawkins (1994: especially chapter 5) proposes a parsing explanation for why the FOFC violating configuration is out (Hawkins does not use the label FOCP). He discusses in much detail the four logically possible combinations between a head X and the immediately dominated category YP in (41).

(41)
a. [ X [$_{YP}$ Y ZP]]
b. [[$_{YP}$ ZP Y] X]
c. [ X [$_{YP}$ ZP Y]]
d. [[$_{YP}$ Y ZP] X]

Hawkins discusses typological evidence showing that the harmonic orders in a. and b. are favoured, c. is disfavoured but well attested, while d. (the FOFC violating order) is exceedingly rare. After discussing converging psycholinguistic evidence showing that the order in (41d) is most problematic for the parser, he proposes an account according to which linear orderings that result in faster and more efficient structural recognition in performance are grammaticalized. Space considerations do not allow me to summarize Hawkins's account. However, an extension to the Right Roof Constraint would not be straightforward, since in Hawkins's approach, (41d) is not ruled out by the fact that the dependent link (selected head) precedes the category it depends on (the selecting head).
I will assume that the parsing account to the Right Roof Constraint stemming from Fodor (1978) is more adequate, since it can be easily extended to FOFC if the notion of inverted dependency applies both to cases of movement and to cases of head selection (cf. section 4 above).
Let me conclude this section by introducing an important qualification. In principle the approach to backward dependencies that I am proposing admits two alternative implementations. According to a first one, the configurations that involve backward localization are grammatical but are filtered out by the parser, since they cannot be processed. According

[5] Other accounts for the Right Roof Constraint have been proposed in the literature. Kayne's (1994) approach is compatible with LCA. I refer to Baltin (2005) for a critical discussion.

to an alternative implementation, the grammar contains principles that block backward localization, in order to facilitate the work of the parser. I will assume the latter implementation, much in the spirit of Hawkins's (1994, 2004) work. Hawkins offers several pieces of evidence in defence of what he calls the Performance-Grammar-Correspondence Hypothesis, according to which grammars conventionalize syntactic structures in proportion to their degree of preference in performance.
The idea that the grammar contains principles blocking linear ordering difficult for the parser may seem at odds with the assumption, shared by much recent minimalist work, that hierarchical structures are built through recursive application of the operation Merge, which does not contain any information about order. In minimalist approaches, the fundamental structure building operations are linearly insensitive. Can this minimalist insight be reconciled with the approach that I am advocating?
The answer depends on the specifics of the linearization principles. Clearly, for articulatory reasons, the hierarchical structure must be mapped into a linear sequence at latest by PF. One possibility is that all syntactic operations apply before linearization takes place. In this scenario, syntax would be totally blind to word order considerations. Obviously, this scenario is not compatible with my approach to FOCF and the Right Roof Constraint.
Another possibility, however, is that linearization takes place cyclically after the relevant portion of the hierarchical structure has been built, but before other syntactic operations take place. It has been proposed that a phrase (or a phase, or any other relevant syntactic unit) is linearized as soon as it has been completed (an example of this approach is Fox and Pesetsky's 2005 cyclic linearization theory). In this alternative scenario, syntax is not necessarily blind to word order, because some syntactic operations may take place after the relevant portion of the structure has been linearized. I will assume that this is what happens with the operation Agree that links two discontinuous positions (and can trigger Internal Merge= Move). I propose that Agree is sensitive to liner order in the way described in (42).

(42) In a forward dependency, Agree can be cross-phrasal. In a backward dependency, Agree is phrase-bound.

So, the locality principle in (42), although ultimately motivated by the need to facilitate the parser, is a principle operating in the syntactic component. As it should be clear, (42) can derive both FOFC and the Right Roof Constraint. However, (42) has general and non trivial consequences for head-final languages. I will discuss them in section 9.

## 7. Towards an implementation: what is a phrase in Bare Phrase Structure Theory?

Up to now, I used X-bar theory as a useful descriptive tool and I have stuck to a simplified clausal structure. Now I will recast my approach by assuming a version of Bare Phrase Structure theory. This reformulation will make my approach more precise and will help us in section 8, when we will consider apparent counterexamples to FOFC.
Since Bare Phrase Structure does not assume the X-bar schema and the familiar distinction between terminal nodes, intermediate projections and maximal projections as a primitive, one can ask how a phrase can be defined in this impoverished system. I base my discussion in this section on Cecchetto and Donati's (2010) version of Bare Phrase Structure theory.
If the Inclusiveness Condition is taken seriously, when a syntactic object $\{\alpha, \beta\}$ is formed by merging $\alpha$ and $\beta$, its label cannot be a new object distinct from $\alpha$ and $\beta$ (as it was in standard X-bar theory or in Chomsky's 1995 version of Bare Phrase Structure). Clearly, a version of Bare Phrase Structure, in order to be satisfactory, must contain some algorithm that determines the label of $\{\alpha, \beta\}$ for each instance of Merge between $\alpha$ and $\beta$. Cecchetto and Donati, building on Chomsky's (2008), propose the following algorithm (they extend the notion of Probe to include cases of selection, for example a verb is taken to be a Probe, if it selects for a DP):

(43) Probing Algorithm:
The label of a syntactic object $\{\alpha, \beta\}$ is the feature(s) which act(s) as a Probe of the merging operation creating $\{\alpha, \beta\}$

For example, the following are the fundamental steps of the derivation of a simple transitive sentence like 'the boy ate the cake' (I refer to Cecchetto and Donati 2010 for a discussion of other types of sentence):

(44) Sketch of the derivation of "The boy ate the cake"

(i) The label of {ate, {the, cake} } is the categorial feature of V
(the transitive verb selects for a direct object)

(ii) The label of {v , {ate, {the, cake} }} is the categorial feature of v
(v selects for the VP)

(iii) The label of {{the, boy}, {v , {ate, {the, cake} }}} is the categorial feature of v

(when the external argument is merged in Spec,v the feature which triggers the merging operation is the categorial feature v, which requires that an external argument be merged)

(iv) The label of {T, {{the, boy}, {v , {ate, {the, cake} }}}} is the categorial feature of T
(T selects for vP)

(v) The label of {{the, boy}, {T, {{the, boy}, {v , {ate, {the, cake} }}}}} is the categorial feature of T
(when the subject is internally merged in Spec,T the feature which triggers this operation is the categorial feature of T -- I assume that the phi-features of T can be checked in situ via Agree, so they do not, at least directly, trigger merge of the external argument).

As Cecchetto and Donati note, when α provides the label, only a subset of its features "percolate" to (namely, become the label of) {α , β}. For example, at step (44i) the categorial feature of V becomes the label when V is merged with direct object but the phi features of V do not "percolate". As should be apparent, at each application of Merge there is some sort of "feature stripping", since the features that do not become the label are not projected further up. Assuming this approach, we can define a maximal projection/phrase as in (45):

(45) A maximal projection αP is the biggest syntactic object in a which a subset of the features of α are the label.

Interestingly, paradigmatic cases of maximal projections (as opposed to the corresponding cases of intermediate projections or heads) are never Probes: they are Goals. On the other hand, heads and intermediate projections are, or can be, Probes. For example, a head selects for, or probes, its complement and an intermediate projection probes a category, which can move to its Spec. The following is a representative list (X-bar terminology is used for expository convenience).

| Maximal Projections (Goals) | Intermediate Projections (Probes) | Heads[6] (Probes) |
|---|---|---|
| VP (selected by v) | V' (selects verb arguments) | V (selects external argument) |
| vP (selected by T) | v' (selects external argument) | v (selects VP) |
| TP (selected by C) | T' (probes external argument) | T (selects vP) |
| CP (selected by main V) | C' (probes wh phrases) | C (selects TP) |

This approach can derive the fact that maximal projections are Goals (and not Probes) as a corollary of the combined definitions of Probing Algorithm (cf. 43) and of maximal projection (cf. 45). If we define a maximal projection αP as the biggest syntactic object in which a subset of the features of α are the label, by definition α cannot provide the label when αP will be merged with β to form {αP, β}. But, if α does not provide the label to {αP, β}, by the Probing Algorithm, no feature of α is the Probe of the merging operation creating {αP, β}. Whence αP cannot be a Probe.
In this section, we have established that, within a phrase αP, a process of feature stripping determines that only a subset of the features of α "percolate up" at the αP level. This process of feature stripping turns out to be crucial in the next section.

8. Some apparent counterexamples to FOFC

After sharpening our understanding of the notion of maximal projection in Bare Phrase Structure theory, we are ready to deal with some counterexamples to FOFC, which are discussed by Biberauer et al. (2008a and 2008b). They come in at least two varieties:
(i) FOFC does not apply across extended projections (in the sense of Grimshaw 1991)
and
(ii) Final particles occurring in head-initial languages.
Let us consider them in turn.

8.2 FOFC does not apply across extended projections

FOFC applies well enough within the extended projection of verb and noun, as we have seen. However it is rather clear that it does not apply across extended projections, as the German sentence (46) clearly shows:

(46) Johann hat [$_{vP}$ [$_{VP}$ [$_{DP}$ den Mann] gesehen ] ]
Johann has the man seen

[6] Heads can be Goals as well (for example, V can be probed by T).

(46) should be a FOFC violating configuration, since the VP is head-final while the DP complement of V is head-initial. However this sentence is perfectly OK.
In order to explain why FOFC does not rule (46), we can capitalize on the fact that the V head selects for the entire category DP, instead of entering into a special relationship with the head D. This is consistent with the fact that crosslinguistically there is no systematic head movement of D to V (banning morphological processes like noun incorporation and clitic-like movement of intransitive Ds). Adopting the perspective on labeling summarized in section 7, we can say that (46) is not problematic since V looks just at the subset of features of D that percolate up at the DP level (categorial features and possibly others). So in (46) the verb "gesehen" does not form an inverted dependency with the determiner "den", because it selects the entire DP [$_{DP}$ den Mann].[7]
It is instructive to compare (46) with cases excluded by FOFC, say a configuration like VERB-OBJ-AUX, which is unattested in Germanic:

(47) * [$_{TP}$ [$_{vP}$ [$_{VP}$ V DP] T ]]

Assuming that tense and agreement features on the verb do not "percolate" at the VP level, T in (47) should look inside the VP to search V, but this would create an inverted dependency across a phrase. This dependency would stand in violation of the locality condition in (42).
In (46), as we said, V needs to see only the features (typically the categorial ones) that "percolate" at the DP level. No cross-phrasal inverted dependency is needed. The reader can easily verify that a similar configuration is created in the instances of FOFC violating configurations that I reported in (6) above.[8] So, my approach naturally explains some apparent counterexamples to FOFC as cases in which an inverted dependency between two heads is not created, because the relevant features percolate from the lower head at the phrasal level.

8.2 Final particles occurring in head-initial languages

Clause final particles in SVO languages like Chinese are a different type of challenge to FOFC, since there seems to be a violation inside the extended projection of the verb. Sentences with a structure like (48) are expected to be out, since CP is head-final but TP is head-initial. However, Chinese is an SVO language with clause final particles that plausibly sit in the C area.

(48) * [$_{CP}$ [$_{TP}$ T [$_{vP}$ [$_{VP}$ V Obj ]]] COMP ]

It would be desirable to extend the same approach applied to the German cases to Chinese cases displaying the configuration in (48), namely arguing that clause-final particles in Chinese might only look at the TP label with no need to access the features of the T head. This would set clause final particles of the Chinese type apart from complementizers in many Indo-European languages. Complementizers in these languages have a closer relationship with T, as witnessed by the fact that (i) they are sensitive to the inflectional features on the verb (the complementizer may take a different form depending on whether the verb is finite, non finite, subjective etc). and (ii) AUX-TO-COMP movement is well attested.
On the other hand, the available research seems to indicate that no clause particle is directly related to the impoverished inflectional system of Chinese. Up to now, mostly for simplicity, we have been working with a very simplified clausal structure, that does not consider the results emerging from cartographic studies. However, in a cartographic work devoted to the fine structure of the COMP area in Mandarin, Li (2006), building on Rizzi's (1997), analyzes four Mandarin clause final particles (a, ba, ma, ne) and concludes that the articulated structure of Mandarin CP is as follows:

(49) Discourse > Degree > Force > Evaluative > Mood > Fin
a ba, ma ne

Crucially, no particle occurs in the Fin layer of the COMP area in Mandarin. Assuming that only complementizers in the FIN layer need to look at the morphological features of the verb, we can tentatively conclude that Mandarin clause final

[7] A reviewer observes that, if the selection relationships hold between heads and their phrasal complements, rather than between heads and the heads of their complements, then selection-based probing is fundamentally different from Agree-based probing. I do not think that this difference should lead us to postulate the existence of two distinct types of Probing, though. The operation of Probing might be unique, with the difference due to the locus of the probed features. Selectional features are found in the phrasal complement of the relevant head, so the Goal relevant for selection happens to be the sister node of the Probe. However agreement features do not percolate (see section 7), so, in this case, the closest Goal is embedded within the sister node of the Probe.

[8] This is not one hundred percent true, due to the fact that in the nominal domain, Finnish allows all permutations except [$_{PP}$ [$_{NP}$ N-Complement ] P], the FOFC violating order. The absence of this order is not explained given the approach to FOFC that I assume, since P arguably selects the entire NP and does not need to establish a cross-phrasal inverted dependency with N. I have nothing to say on the Finnish case, expect for observing that Biberauer et al. report that that the [$_{PP}$ [$_{NP}$ N-Complement ] P] order may be attested in other languages (they mention Slovenian, German and Gungbe as possible candidates). Therefore it is possible that the impossibility of the [$_{PP}$ [$_{NP}$ N-Complement ] P] order in Finnish is not due to FOFC, but to language-internal reasons that must be better understood.

particles do no need to inspect the (very impoverished) inflectional morphology of the verb, whence they do not look inside the TP.
This suffices to explain why particles in Mandarin can occur in a clause final position, in apparent violation of FOFC: no inverted dependency is created between any head in the CP area and T. What remains to be seen is whether this approach can be generalized to other cases of final particles in head-initial languages. I leave this to future research.
In this section, we have seen that potential counterexamples to FOFC are easily explained if FOFC is reduced to a condition like (42), which rules out cross-phrasal inverted dependencies. In fact, potential counterexamples are cases in which no cross-phrasal inverted dependency is created at all, since the selecting head selects for the entire phrase that it immediately c-commands, not for the head of this phrase.

## 9. Typological consequences: morphological implications for head-final languages

My approach unifies conditions on backward selection (FOFC) and conditions on rightward movement (Right Roof Constraint) since, as I have argued, both FOFC and the Right Roof Constraint exclude cases in which an inverted dependency extends outside the phrase. Although this unification is desirable, it seems to cause a problem with (harmonic) head-final languages. For example, a condition like (42) seems to predict that COMP should not be able to select T in the following configuration, since the dependency between COMP and T would be a case of inverted dependency that extends outside the phrase (similarly for the other cases of selection by a head in the lower part of the clause):

(50) $[_{CP}\ [_{TP}\ \ldots\ VP\ \ T]\ COMP\ ]$

The problem arises since we have assimilated conditions on movement and conditions on selection. If a cross-phrasal inverted dependency involving (phrasal) rightward movement is not allowed, a cross-phrasal inverted dependency involving selection (or head movement) shouldn't either.
Clearly, the logic of my approach leads us to extend to cases like (50) the treatment proposed for Mandarin clause final particles. Namely, we might say that in head-final languages, COMP accesses only the subset of the features of T that percolate up to the TP level and does not form an inverted dependency with T (similarly for other cases of backward selection in the lower part of the clause).
Is there any evidence that supports this claim for head-final languages? One way to answer this question would be inspecting if rightward head movement occurs in these languages. My approach predicts that it shouldn't. However, rightward head movement is string vacuous in head-final languages, so any evidence for (or against) it is often theory internal and therefore not very compelling (but see Vikner 2005 for non-theory-internal evidence against postulating rightward verb movement in OV Germanic)[9].
A more indirect way to answer the question is asking what a language would look like in which an inflectional head cannot be linked to the closest c-commanded inflectional head either by head movement or by direct selection. In such a language the two pieces of morphological information that are expressed by the two heads should not conflate but should be independently expressed. Therefore, in a hypothetical language of this type affixes should not become fused with others, and should not change form conditioned by others. But this description applies to many languages that have been traditionally described as agglutinative. Stated differently, my approach predicts that developing an agglutinative-like morphology might be a strategy used by SOV languages to avoid illicit (namely, cross-phrasal) inverted dependencies. If I am right, there should be a strong correlation between the property of being head-final and the property of being agglutinative in the inflectional system. In fact, this correlation has been noted by several authors (cf. Julien 2002[10], Lehmann 1973, van Riemsdijk 1998[11] and Shukla, Nespor and Mehler 2007) and it is quite clear that typical agglutinative languages are head-final languages.

---

[9] My approach is not compatible with the occurrence of rightward head movement within the nominal domain either, since that would be another case of illicit backward dependency. In this paper I cannot explore this topic but I can mention the conclusions emerging from the recent literature on Greenberg's Universal 20. There are at least two distinct approaches to Universal 20, namely Cinque's (2005) approach, which is consistent with antisymmetry, and a more classical approach recently revamped by Abels and Neeleman (2007), which assumes that right branching structures can be base generated. Crucially, both Cinque and Abels and Neeleman converge on the fact that rightward head movement does not occur within the DP.

[10] Julien (2001) claims that agglutinative morphology is a consequence of the fact that verb movement does not occur in consistent head-final languages. In her theory, in these languages the verb and its affixes do not correspond to a syntactic unit but are perceived as a single word because they regularly appear linearly adjacent to each other (cf. the theory of wordhood developed in Julien 2002). In this respect, my approach is similar to Julien's. However, our approaches sharply diverge in other respects. Following Kayne (1994), Julien (2001) assumes that all languages are underlyingly SVO and that the order of morphemes that we see in head-final languages is derived by repeated movement of the complement of X to the specifier of X. Assuming this framework, Julien has to stipulate that, in head-final languages, heads in the clausal domain lack the special 'strong head feature' that triggers head movement in non agglutinating languages. In my approach, this stipulation is not needed, since verb movement is blocked in head-final languages by the locality condition in (42), which is independently motivated.

[11] In order to explain the correlation between head-finality and agglutinative morphology, van Riemsdijk (1998) claims that agglutinative structure results from head adjunction and assumes that head adjunction (as opposed to substitution) can take place only

However, the notion that a language as a whole is agglutinative is today rejected as an over-simplification by most authors, so I tried to be more precise when checking the prediction of my approach for the syntax/morphology interface. In order to do so, I made a search by using the tools made available by WALS Online. I selected languages that share two revealing morphological properties.

The first property has to do with fusion. Fusion, as defined in chapter 20 of WALS (Bickel and Nichols 2008a), refers to the degree to which grammatical markers are phonologically connected to a host word or stem. Three basic values are identified: isolating, concatenative, and nonlinear. Isolating formatives are full-fledged phonological words of their own. Strings of concatenative formatives can be segmented into clear-cut morphemes. Nonlinear formatives are realized not in linear sequence, but by direct modification of their host (for example, they may involve tonal modification).

The second property has to do with exponence. Exponence, as defined in chapter 21 of WALS (Bickel and Nichols 2008b), refers to the number of categories that cumulate into a single formative. Each category may be expressed by a dedicated formative. When this happens the formative is monoexponential. However formatives can simultaneously code more than one category. A well-known example is number and case in many Indo-European languages. Marking of case in these languages involves the same formative as marking of number, and it is impossible to identify separate markers of case and number. Formatives coding more than one category are called polyexponential.

Having identified fusion and exponence as the relevant morphological properties, I selected all languages in WALS Online database that are exclusively concatenative and in which the case morpheme expresses only information about case (monoexponential case). I assume that languages having both these features are good candidate for being languages in which different heads not to conflate, namely for being agglutinative in the sense relevant for this paper[12].

WALS Online lists 55 such languages and for 45 of them it also contains information on word order. I checked if a certain word order prevails among these 45 languages. The expectation is that, if head-finality induces heads not to conflate, the SOV order should be overrepresented.

The result of this search are illustrated in (51). It is instructive to compare the two major groups: SVO languages and SOV languages. The entire WALS Online database contains 1228 languages. The total number of SVO languages is not dramatically different from the total number of SOV languages: 497 languages (40,47 %) have an unmarked SOV order while 436 (35,51 %) have an unmarked SVO order (the remaining languages either have another word order or their word order could not be identified). However, out of 45 languages that share the two relevant features, there are 31 SOV languages (68.89%), while there are only 3 SVO languages (6.66%).

---

(51)

Word order in 45 languages that are exclusively concatenative and have a monoexponential case

SOV LANGUAGES: 31 (68,89 %)
Arrernte (Mparntwe), Awa Pit, Aymara, Basque, Burmese, Burushaski, Cahuilla, Digaro, Epena Pedee, Georgian, Hindi, Hunzib, Ingush, Japanese, Kannada, Kewa, Khalkha, Koasati, Korean, Lezgian, Maricopa, Mundari, Nubian (Dongolese), Oromo (Harar), Persian, Quechua (Imbabura), Rama, Sanuma, Tlingit, Turkish, Yidiny

SVO LANGUAGES: 3 (6,66 %)
Purépecha, Spanish, Martuthunira

VSO LANGUAGES: 1 (2,22 %)
Squamish

VOS LANGUAGES: 3 (6,66 %)
Malagasy, Wembawemba, Zoque (Copainalá)

OVS LANGUAGES: 1 (2,22 %)
Mangarrayi

LANGUAGES WITH NO DOMINANT WORD ORDER: 6 (13,33 %)
Barasano, Hungarian, Imonda, Miwok (Southern Sierra), Wardaman, Wichita

(Source: WALS Online, accessed on December, 28, 2008)

---

between linearly adjacent heads. van Riemsdijk's account is hardly rephrasable in a minimalist setting, since, assuming Bare Phrase Structure theory, there is no easy way to distinguish between head adjunction and substitution.

[12] I also chose these two features because they allow an automatic search. At the moment of writing, WALS does not allow automatic searches that would be equally (or possibly more) interesting. For example, I could not look at the word order of languages that are exclusively concatenative and monoexponential in the Tense-Aspect-Mood domain. My approach predicts that the SOV order should be overrepresented in this group of languages, as well.

It seems to me that the results in table (51) clearly confirm the existence of a correlation between word order (head-finality) and morphological features that, at least in my interpretation, originate when heads do not conflate.
However, agglutination-like morphology may not be the only strategy head-final languages adopt to deal with the problem raised by backward selection. Another device these languages may use is phrasal inflectional morphology. Jo (2003) discusses a very clear case of phrasal morphology, namely nominalization of embedded clause as in Korean, Turkish and many other head-final languages. Jo shows that, in languages like Korean, affixal nominalizers attach to the clausal constituent and not to the verb, as revealed by the fact that a conjunct clause can be affixed by a nominalizer (in 52 the nominalizer is "um" and the coordinating particle is "ko"):

(52)
Na-nun [[John-i pap-ul mek-ess-ko Mary-ka ice cream-ul mek-ess]-um]-ul al-ko-iss-0-ta.
I-TOP John-NOM rice-ACC eat-PAST-and Mary-NOM ice cream-ACC eat-Past-NOMINALIZER-ACC know-INF-IS-PRES-DEC
`I know that John ate rice and Mary ate ice cream.'
(from Jo 2003)

The contrast with English type nominalization is sharp, since nominalizers in English form a morphological unit with the verb, not with the clause:

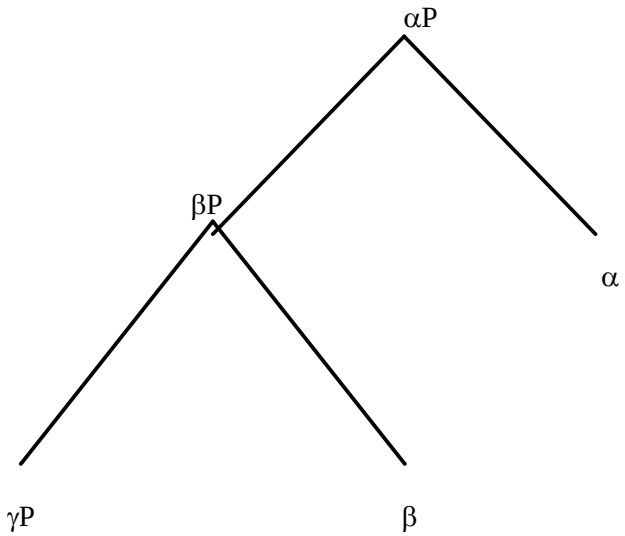
(53) * [John sing and Bill read]-ing bothered me.

The fact that typical head-final languages show instances of phrasal morphology (as opposed to word level morphology) is expected, given my assumption that a selecting head in these languages does not directly access the selected head but only the phrase headed by the latter[13].
I conclude by reporting an intriguing remark made by an anonymous reviewer, who notes that the English possessive 's construction exemplifies phrasal morphology and is a FOFC violating configuration. This is illustrated in (54), in which the clitic head 's is head-final but DP is head-initial.

(54) [$_{PossP}$ [$_{DP}$ That [$_{NP}$ old man]]'s] bycicle

It is very tempting to assume that the same device that operates in the clausal domain in head-final languages operates in the nominal domain in English. If this suggestion is on the right track, phrasal morphology can be seen as a general repair mechanism for fixing FOFC violating configurations.
In this section, I have argued that, given the approach to FOFC that I developed, an inflectional head α should not be able to establish a direct link with the inflectional head β in a configuration like (55), although α can access the subset of the features of β that percolate at the βP level:

(55)




[13] The fact that phrasal morphology (including clause nominalization) may be forced in head-final languages might be related to the well known observation that internally headed relatives clauses are mostly found in these languages, although I cannot explore this fact in this paper.

I proposed that this has important consequences for the syntax/morphology interface. First, no head movement should occur between α and β. Second, if α does not directly access β, α and β should not conflate and get fused in complex words. This might explain the correlation between head-finality and agglutinative-like morphology in the inflectional system, which has been noted in the literature and seems to be confirmed by a search in the WALS Online data base. Third, phrasal morphology (as opposed to word level morphology) is expected to occur in the configuration (53) as a way to avoid backward inspection by α inside the βP phrase.
Since the absence of rightward head movement and the presence of phrasal or agglutinative-like morphology all derive from the fact that consistently head-final languages select their complements differently from head-initial languages, I propose that these properties should be viewed as part of the parametric cluster associated with head-finality.

## 10. Conclusion

In this paper, I highlight a close resemblance between the configuration excluded by FOFC and the one excluded by the Right Roof Constraint. Both require backward localization, either of a trace or of a selected head. I argued that this resemblance is not coincidental. On the contrary, FOFC and the Right Roof Constraint have the same underlying motivation: since backward localization is costly for the syntactic parser, which proceeds left-to-right, backward localization is possible only if it applies very locally, namely phrase-internally. Therefore, grammar internalizes a principle that blocks long inverted dependencies. This principle has been formally stated in (42).
After defining a phrase in Bare Phrase Structure theory as the biggest syntactic object in a which a subset of the features of the head can percolate, I could more closely explore the consequences of my approach. On the one hand, I could straightforwardly explain certain apparent counterexamples to FOFC, like the fact that this condition does not apply across extended projections.
On the other hand, I observed that head-final languages introduce a potential challenge for my approach, since in these languages inverted dependencies may be more difficult to avoid. I proposed that, while rightward head movement might be altogether banned, selection may take place in head-final languages only because the relevant features of the selected head percolate up to the phrasal level, where they become visible for the selecting head. The impossibility of directly linking the selecting head and the selected one is not without consequences, though. I suggested that this might explain why typical head-final languages have an agglutinative-like inflectional system and display forms of phrasal morphology. This indicates a close link between morphology and syntax, with the latter forcing the former to take a specific form.
I conclude with a rather speculative observation about the general issue of linearization. I proposed that the grammar contains a principle that blocks linear orderings that are difficult for the parser. So, at least in the case of "pathological" (disharmonic) word orders, Hawkins's (2004) Performance-Grammar-Correspondence Hypothesis seems to be on the right track. This suggests to me that the explanation for the occurrence of "proper" (harmonic) word orders should also be searched by looking at the interaction between grammar and performance. In this view, linearization is still done in syntax but is governed by interface-oriented principles that optimalize the syntax/PF interface. Future work will further evaluate the soundness of this research program.

## References


Abels, Klaus and Ad Neeleman
2007 Linear Asymmetries and the LCA, manuscript, University of Tromsø and University College London.
Ackema, Peter and Ad Neeleman
2002 Effects of Short-Term Storage in Processing Rightward Movement. In: Sieb Nooteboom, Fred Weerman and Frank Wijnen (eds.), Storage and Computation in the Language Faculty, 219-256. Dordrecht: Kluwer.
Baltin, Mark R.
1981 Strict Bounding. In: C. L. Baker and John McCarthy (eds.) The Logical Problem of Language Acquisition, 257-295. Cambridge, MA: MIT Press.
Baltin, Mark
2005 Extraposition. In: Martin Everaert, Henk Van Riemsdijk, Rob Goedemans and Bart Hollebrandse (eds.), The Blackwell Companion to Syntax, 237-271. Oxford: Blackwell.
Belletti, Adriana
1990 Generalized Verb Movement. Turin: Rosenberg and Sellier.
Bhatt. Rajesh and Veneeta Dayal
2007 Rightward Scrambling as Rightward Remnant Movement. Linguistic Inquiry, 38(2): 287–301.
Bianchi, Valentina
2002 Headed relative Clauses in Generative Syntax. Part I. Glot International 6.7: 197-204.
Biberauer, Theresa, Anders Holmberg and Ian Roberts
2008a Disharmonic word-order systems and the Final-over-Final-Constraint (FOFC). In: Antonietta Bisetto e Francesco E. Barbieri (eds.), Proceedings of the XXXIII Incontro di Grammatica Generativa, 86-105. Bologna: AMS Acta.
Biberauer, Theresa, Anders Holmberg and Ian Roberts
2008b Structure and Linearization in Disharmonic Word Orders. In Charles B. Chang and Hannah J. Haynie (eds.) Proceedings of the 26th West Coast Conference on Formal Linguistics, 96-104. Somerville, MA: Cascadilla Proceedings Project.
Biberauer, Theresa, Anders Holmberg and Ian Roberts
2009 A hierarchical universal and its consequences for the LCA and word-order parameters. talk given at GLOW 32, Nantes.
Biberauer, Theresa, Glenda Newton & Michelle Sheehan
to appear a. On Impossible Changes and Impossible Borrowings. In Breitbarth, A., C. Lucas, S. Watts and D. Willis (eds.). Continuity and Change in Grammar. Amsterdam: Benjamins.
Biberauer, Theresa, Glenda Newton & Michelle Sheehan
to appear b. Limiting synchronic and diachronic variation and change: the Final-Over-Final Constraint. To appear in a Language and Linguistics.
Balthasar Bickel and Johanna Nichols
2008a Fusion of Selected Inflectional Formatives. In: Haspelmath, Martin, Matthew S. Dryer, David Gil and Bernard Comrie, (eds.) The World Atlas of Language Structures Online. Chapter 20. Munich: Max Planck Digital Library. Available online at http://wals.info/feature/20. Accessed on December, 28, 2008.
Balthasar Bickel and Johanna Nichols
2008a Exponence of Selected Inflectional Formatives. In: Haspelmath, Martin, Matthew S. Dryer, David Gil and Bernard Comrie, (eds.) The World Atlas of Language Structures Online. Chapter 21. Munich: Max Planck Digital Library. Available online at http://wals.info/feature/21. Accessed on December, 28, 2008.
Büring, Daniel and Katharina Hartmann
1997 The Kayne Mutiny. In: Dorothee Beermann, David LeBlanc and Henk van Riemsdijk (eds.) Rightward Movement, 59- 79. Amsterdam: Benjamins.
Cecchetto, Carlo
1999a A Comparative Analysis of Left and Right Dislocation in Romance. Studia Linguistica: 53 (1): 40-67.
Cecchetto, Carlo
1999b Optionality and Directionality. A View from Leftward and Rightward Scrambling in Japanese. In: Grant-in-Aid for COE Research Report No. 08CE1001, Kanda University of International Study
Cecchetto Carlo and Caterina Donati
2010 On labeling: Principle C and head movement. Syntax, 13:2.
Cecchetto Carlo, Carlo Geraci, and Sandro Zucchi
2006 Strategies of Relativization in Italian Sign Language. Natural Language and Linguist Theory 24: 945–975.
Cecchetto Carlo, Carlo Geraci, and Sandro Zucchi
2009 Another way to mark syntactic dependencies. The case for right peripheral specifiers in sign languages. Language 85 (2): 1-43.
Chomsky, Noam
1995 The Minimalist Program. Cambridge, MA: MIT Press.
Chomsky, N. (2008). On Phases. In: Robert Freidin, Carlos Otero, and Maria Luisa Zubizarreta (eds.) Foundational Issues in Linguistic Theory, 133-166. Cambridge, MA: MIT Press
Cinque, Guglielmo
1999 Adverbs and functional heads: A cross-linguistic perspective. Oxford: Oxford University Press.
Cinque, Guglielmo
2005 Deriving Greenberg's Universal 20 and its Exceptions. Linguistic Inquiry 36: 315–332.
Dryer, Matthew S.
1992 The Greenbergian Word Order Correlations. Language, 68 (1): 81-138.
Dryer, Matthew S.

2008 Relationship between the Order of Object and Verb and the Order of Adposition and Noun Phrase. In: Haspelmath, Martin, Matthew S. Dryer, David Gil and Bernard Comrie, (eds.) The World Atlas of Language Structures Online. Chapter 95. Munich: Max Planck Digital Library. Available online at http://wals.info/feature/95. Accessed on December, 28, 2008.
Dryer, Matthew S.
2009 The branching direction theory revisited. In: Sergio Scalise, Elisabetta Magni, and Antonietta Bisetto (eds.). Universals of Language Today, 185-207. Berlin: Springer.
Emonds, Joseph E.
1976 A Transformational Approach to English Syntax. New York: Academic Press.
Fodor, Janet Dean
1978 'Parsing strategies and constraints on transformations', Linguistic Inquiry 9 (3): 427-473.
Fox, Danny and David Pesetsky
2005 Cyclic linearization and syntactic structure. Theoretical Linguistics 31: 1-46.
Frazier, Lyn and Giovanni Flores d'Arcais
1989 Filler driven parsing: a study of gap filling in Dutch. Journal of Memory and Language 28: 331- 344.
Geraci, Carlo
2006 Negation in LIS. In: Bateman, Leah and Cherlon Ussery (eds.), Proceedings of the Thirty-Fifth Annual Meeting of the North East Linguistic Society, Vol. 2, 217-230. Amherst, MA: GLSA
Greenberg, Joseph (ed.)
1966 Universals of Language. Cambridge, MA: MIT Press.
Grimshaw, Jane
1991 Extended projection. Unpublished paper, Brandeis, University.
Hawkins, John A.
1994 A performance theory of order and constituency. Cambridge: Cambridge University Press.
Hawkins, John A.
2004 Efficiency and Complexity in Grammars. Oxford: Oxford University Press.
Holmberg, Anders
2000 Deriving OV order in Finnish. In: Svenonius, P. (ed.) The derivation of VO and O, 123-152.. Amsterdam: Benjamins.
Jo, Mi-Jeung
2003 The correlation between syntactic nominalization and the internally headed relative constructions in Korean. Studies in Generative Grammar, 13 (4): 535-564.
Kayne, Richards
1994 The Antisymmetry of the Syntax. Cambridge, MA: MIT Press.
Kiss, Tibor
2005 Semantic Constraints On Relative Clause Extraposition?. Natural Language and Linguistic Theory 23: 281–334.
Kural, Murat
1997 Postverbal Constituents in Turkish and the Linear Correspondence Axiom. Linguistic Inquiry, 28 (3): 498-519.
Julien, Marit
2001 Word Order Type and Syntactic Structure. In: Rooryck, Johan and Pierre Pica (eds.) Yearbook of language variation 1, 17-59. Amsterdam: John Benjamins.
Julien, Marit
2002 Syntactic heads and word formation: a study of verbal inflection Oxford: Oxford University Press.
Lebeaux, David
1995 Where does the Binding Theory Apply? University of Maryland Working Papers in Linguistics, 3.
Lehmann, Winfred P.
1973 A Structural Principle of Language and Its Implications. Language 49: 47-66.
Li, Boya
2006 Chinese Final particles and the Syntax of the periphery. Utrecht: LOT.
Nespor, Marina, Maria Teresa Guasti and Anne Christophe
1996 Selecting word order: the Rhythmic Activation Principle. In Ursula Kleinhenz (ed.) Interfaces in Phonology, 1-26. Berlin. Akademie Verlag.
Riemsdijk, H. van.
1998 Head movement and adjacency. Natural Language and Linguistic Theory 16 (3): 633-678.
Rizzi, Luigi
1997 The Fine Structure of the Left Periphery. In: Liliane Haegeman (ed.), Elements of Grammar, 281-337. Kluwer: Dordrecht.
Rochemont, Michael S.
1992 Bounding rightward A'-dependencies. In: Goodluck, Helen and Michael S. Rochemont (eds.), Island constraints, 373-397. Dordrecht: Kluwer.
Rochemont, Michael S. and Peter W. Culicover
1990 English Focus Constructions and the Theory of Grammar. Cambridge: Cambridge University Press.
Ross John R.
1967 Constraints on Variables in Syntax, Ph.D. dissertation, MIT
Shukla, Mohinish, Marina Nespor and Jacques Mehler
2007 Grammar On A Language Map, manuscript, SISSA, Trieste.
Vikner, Sten
2005 Immobile complex verbs in Germanic. Journal of Comparative Germanic Linguistics 8: 83-115.
Vries, Mark de
1999 Extraposition of relative clauses as specifying coordination. In: Tina Cambier-Langeveld, Anikó Lipták, M. Redford and E.J. van

der Torre (eds.) Proceedings of ConSole VII, 293-309. Leiden: SOLE.
Zucchi, Sandro
2009 Along the time line. Natural Language Semantics 17: 99–139.