THE FINAL-OVER-FINAL CONSTRAINT AS AN EFFECT OF COMPLEMENT STRANDING

MICHELLE SHEEHAN

Abstract

In this article I attempt to derive the Final-over-Final-Constraint (FOFC- Holmberg 2000, Biberauer, Holmberg and Roberts 2007, 2008 and Biberauer, Newton and Sheehan to appear) from fundamental facts about hierarchical structure and linearization. FOFC is a generalization over a striking asymmetry in the attested/grammatical word orders of the World's languages: a head-initial phrase cannot be (immediately) dominated by a head-final phrase. Diverging from the analysis of Biberauer, Holmberg & Roberts (2007, 2008), I propose that FOFC is a direct effect of linearization, derived from a version of Kayne's (1994) Linear Correspondence Axiom (LCA) without its 'induction step' (Uriagereka 1999) coupled with a variant of Chomsky's (1995) Bare Phrase Structure (BPS).

It is proposed that if head-finality is derived via movement, as Holmberg (2000) Biberauer, Holmberg & Roberts (2007, 2008) propose, following Kayne (1994), then FOFC is really a restriction on specifiers, and as such it might result from the familiar linearization challenge which the latter pose. It is claimed that there are two distinct ways to linearize a complex specifier with respect to the clausal spine: (i) Multiple Spell-Out (Uriagereka 1999), (ii) Complement Stranding. The second possibility only applies where a non-linearized, head-initial phrase moves from a complement position. In such cases the complement of this complex phrase gets stranded in its base-generated position to enable linearization. (i) derives CED effects and (ii) finds support from discontinuous constituents of various kinds as well as from the gap described by FOFC: if head-finality is derived by comp-to-spec movement then FOFC can actually be seen as a further instance of the same effect.

The crucial claims are: (i) that terminals define their own c-command domains via BPS, with the phrasal label counting as an instance of the projecting head; (ii) that the equivalent to X-bar levels are inert in terms of c-command; (iii) that there is no induction step stating that if X c-commands Y then everything dominated by X precedes Y; and (iv) that copy-deletion is subject to last resort, so that the need for a total linear order of terminals affects which copies are 'targeted' by PF.

1. Introduction

Since its inception, Kayne's (1994) Linear Correspondence Axiom (LCA), which claims that hierarchical position and linear order are correlated, has remained the subject of much interest and controversy. The abundance of head-final word orders proves challenging for the antisymmetry hypothesis, as deriving the former from base-generated spec-head-complement orders seems to involve unmotivated movement and, in some accounts, the presence of null heads, invisible at both the PF and LF interfaces. As such, many have proposed that what was traditionally been termed the 'head parameter' belongs in the PF component (cf. Richards 2004 *inter alia*). While this is an

attractive option, especially given the lack of asymmetry between head and complement in Bare Phrase Structure (Chomsky 1995), it is not without its own problems. In this paper, I will defend a movement account for head-final orders based on empirical facts about extraposition, discontinuous constituents, head movement and typological word order trends. It will be claimed that A-, A-bar and comp-to-spec (linearization) movement, despite their obvious differences, actually share certain properties and that these similarities cannot be captured by a non-syntactic theory of head-directionality.

The structure of the paper is as follows. I begin by outlining my theoretical assumptions regarding Bare Phrase Structure and the LCA in section 2, making it clear where they diverge from standard interpretations. I then discuss the linearization of specifiers in section 3, introducing Uriagereka's (1999) Multiple Spell-Out, which facilitates a highly Minimalist simplification of the LCA. Following this, in section 4, I point out a potential look-ahead problem with Multiple Spell-Out as it is presently conceived of, and I propose a solution to this problem based on Complement Stranding which makes clear empirical predictions. In section 5 I go on to show empirical support for Complement Stranding in English. Section 6 shows how Complement Stranding also derives the Final-over-Final Constraint essentially for free, if it is assumed that head-finality is movement-derived. Section 7 concludes and outlines potential implications and extensions of the approach.

2. The Linear Correspondence Axiom meets Bare Phrase Structure

2.1. Kayne's (1994) Linear Correspondence Axiom

Kayne's pioneering work on the relationship between hierarchical structure and linear order proposes that there should be a direct correlation between the two of the following kind:

(1) *A reworded version of Kayne's (1994) LCA*
a) Basic step: If α asymmetrically c-commands ß, then α precedes ß.
b) Induction step: If α precedes ß and α dominates γ then γ precedes ß.[1]

The basic insight of his proposal is that asymmetric c-command in syntax equates to precedence at PF, a hypothesis which allows us to infer hierarchical structure indirectly from linear order. A few stipulations are required to make this algorithm work, though. Firstly, the c-command is said to hold of categories, not terminals or segments. Secondly, the LCA includes what Uriagereka (1999) calls an 'induction step': if *: if α precedes ß and α dominates γ then γ precedes ß* (this is called the Nontangling Condition by Partee, Meulen & Wall (1990). Thirdly, Kayne relies crucially on X-bar theory (Chomsky 1970, Jackendoff 1977), claiming that the latter can actually be derived from the LCA.

Despite subsequent changes in the theory, the LCA has remained an appealing hypothesis, as it is in some respects highly restrictive. By hypothesis, all phrases are

[1] The fact that the relevant relation is 'precedence' rather than 'subsequence' does not seem to follow from anything deeper. Uriagereka (1999) claims that where two possibilities are equally optimal then either one can be chosen. Given the arbitrary nature of this choice it might be expected that this would be an obvious 'parameter' of UG (cf. Biberauer, Holmberg and Roberts to appear). However, the evidence presented in this paper will suggest that, for some poorly understood reason, this cannot be the case. Rather the LCA appears to be a deep principle of UG.

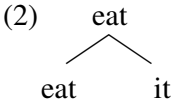
base-generated with the same spec-head-comp format, which will be mapped in the same way to a linear structure, in the absence of movement. The implication is that, if we are to adopt Baker's Uniformity of Theta Assignment Hypothesis (UTAH, Baker 1988: 46) then in all languages, a theme/experiencer object will always be merged as the complement of the verb and will therefore follow the verb in linear order if no further movement takes place. The subject of a transitive verb, likewise, will always be merged as the specifier of v, and so will precede the VP in the absence of any movement. In actual fact, unmarked SVO order in many languages is actually considered to be derived via movement. In English, for instance the subject and perhaps the object and verb move to a higher position (cf. Lasnik 2001 for discussion). Nonetheless, the LCA provides a formal way to link hierarchical structure to linear order, some version of which is presumably required by any theory of grammar.

In other respects, however, the LCA has proven to be highly unrestrictive, as many attested word orders necessitate massive remnant movement triggered by null heads present only to derive the correct surface order (cf. Koopman & Szabolcsi 2000 for a remnant account of complex word orders). In this paper, I will argue that an adjustment in our understanding of the LCA may help to eliminate some of this apparent redundancy from the grammar by providing a much simpler derivation for some cases which would otherwise be analysed using remnant movement.[2] Ultimately though, as Cinque (2005), Abels & Neeleman (2006) and Abels (2007) point out, a full account of the restrictions on movement is required if the LCA is to remain at all tenable. This paper makes a small contribution to this debate, showing that many apparently problematic cases are actually evidence in favour of the LCA, though it by no means solves all of the problems. Crucially it must be noted that the asymmetry discussed in this paper, noted by Holmberg (2000) and Biberauer, Holmberg and Roberts (2007, 2008), is very difficult to account for in a non-LCA-based approach to the relationship between hierarchical structure and linear order.[3]

2.2. Chomsky's (1995) Bare Phrase Structure

In an attempt to simplify the theory of phrase structure based on Minimalist principles, Chomsky (1995) advocates a rejection of X-bar theory and a Bare Phrase Structure (BSP) whereby the head itself provides the label for the phrase. Given BPS, '[t]here is no category-terminal distinction, so either may c-command' (Chomsky 1995: 416). Chomsky thus immediately removes two of the stipulations of Kayne's LCA: (i) the vacuous projections required by X-bar theory and (ii) (partially) the restriction that only categories can c-command. However, this reduction comes at a cost: under BPS, a non-complex complement and selecting head stand in mutual c-command and so cannot be ordered by the LCA:

(2)



[2] Of course I am by no means claiming that ideas put forth in this paper can straightforwardly derive all of the orders discussed by Koopman and Szabolcsi (2000).

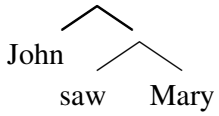
[3] One possible approach, though, is to remove this problem from the syntax proper and to attribute it to parsing or processing preferences. See Hawkins (1994, 2004) and Cecchetto (2007) for approaches along these lines, and Sheehan (2008b) for a critical appraisal of processing-based accounts.

Chomsky himself acknowledges this and proposes two potential solutions to the problem. Either a PF rule can regulate the order of head and complement (as taken up by Epstein, Groat, Kawashima & Kitahara 1998 and Richards 2004, or as Chomsky himself suggests, it can be resolved by movement (a line also pursued by Moro 2000). I will propose that it is resolved either via null structure or movement, but that this movement can be either of the complement or of the rest of the phrase.[4]

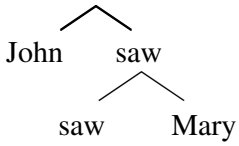
2.3. Eliminating X-bar levels

Departing from Chomsky (1995), I propose that intermediate projections, which are neither minimal not maximal (formerly X-bar levels) should not be labelled. As such, phrase structure should take the form in (2a) rather than (2b):

(3) a.



b.



This small difference in assumptions has wide-reaching ramifications and so I will take some time to justify it (cf. also Chomsky 2001: 40 for relevant discussion).

The empirical motivation for this step is clear: it is widely accepted that X-bar levels are not visible to the narrow syntax (Chomsky 1995: 242, Epstein 1999). As Epstein & Seely (2006) summarize:

> …single-bar projections seem not to participate in operations in C_{HL}. Thus single-bar projections do not assign or receive Case; they do not participate in agreement relations; they do not bind or control and are not bound or controlled; they do not move, delete or insert, nor do they undergo merge.[5]
>
> (Epstein and Seely 2006: 21)

It is also the case, according to Chomsky (1995: 416) that X-bar levels being 'neither minimal nor maximal […] do not c-command'. One very simple way to capture this is by not giving them a label. In many ways this is the null hypothesis and it is the

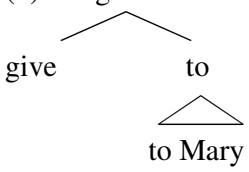
[4] While the proposal made here depends crucially on the assumption that at least disharmonic orders are derived via movement, it does not have much to say about this 'point of symmetry', where no asymmetric c-command obtains. It is not totally incompatible with the proposal that totally harmonic head-final languages are actually the result of a PF-parameter resulting from the mutual c-command between head and complement (as suggested by Epstein et al. 1998, M Richards 2004). If this were the case but disharmonic languages have the same PF-setting as head-initial languages then the proposal made here would still hold. That said, there are reasons to doubt such an analysis at least for many well-studied largely head-final languages like Turkish, Hindi and Lezgian, where there seems to be good evidence for roll-up movement (cf. Zwart 2002, Julien 2002).

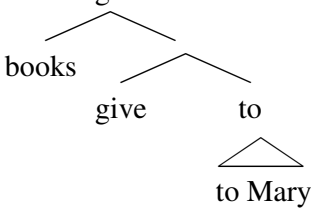
[5] Anders Holmberg (p.c.) points out that in one sense X-bar levels do undergo merge, namely with specifiers. However, in derivational terms, before the merger of a specifier, the X-bar level is actually a full phrase, able to undergo merge. One possible counterexample, however, comes from instances of 'tucking in' where a wh-phrase appears to merge with a genuine X-bar level (cf. Richards 2001). I leave these exceptions to one side here. Theresa Biberauer (p.c.) also reminds me that certain pro-forms are traditionally thought to replace X-bar levels, such as do so (V-bar) and one (N-bar). Given subsequent developments in the theory it is not so clear what the syntactic status of these elements is after all.

opposite stance that should require justification (plus the stipulation that they behave differently from other projections in several respects) (but cf. Chomsky 2001: 40 for a slightly different take).

If we think of projection in truly derivational terms then whatever represents a phrase at any stage of the derivation requires a label (by hypothesis, to render it a visible syntactic object). A phrase without a specifier is labelled until that phrase subsequently merges with a specifier. If a specifier is then merged, the label will project up to label the whole phrase, which clearly has phrasal status, and the intermediate bar level, which is no longer visible to the narrow syntax, will cease to need a label (b).

(4) a. give ⇒ b. give





As such, intermediate, unlabelled levels also fail to c-command and hence do not affect linearization via the LCA This is welcome result as if X-bar levels were visible to the LCA then much of syntax's inherent hierarchy would be removed (as has been noted by Abels 2003 *inter alia*).[6] In such a system everything would be required to move in order for order to be established according to the LCA, which would move all explanatory value from the latter.

This version of BPS allows us to adopt a very simple, highly Minimalist version of c-command:

(5) *Definition of c-command*
A label α c-commands its sister and anything dominated by its sister. A sister of α is any node which shares an immediate mother with α.

This means that, within a phrase, the specifier asymmetrically c-commands both the head and the complement.

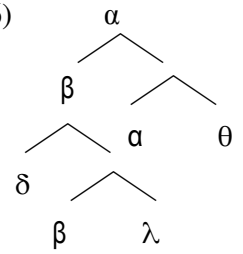
2.4. Really Bare Phrase Structure

One very literal interpretation of BPS is that the label of the whole phrase is merely a graphical representation of the fact that the head of the phrase projects. As such, the label of a phrase is simply another instance of the head. In Chomsky's (1995: 417) words 'a category can have two heads, one a bare head that projects, the other an X^{max}'. Given these facts, and the desire to minimize assumptions, an obvious

[6] Abels (2003) argues that this is exactly the desired result as it enables us to unify the type of relation in which theta-roles can be assigned. For him, specifiers which are the target of movement differ from specifiers which are the target of first-merge. Under his multidominance approach, specifiers which are the target of movement totally c-command the heads they are associated with, whereas the head does not totally c-command its moved specifier. While this is a highly interesting proposal, it implies that all specifiers must move in order for order to be imposed (or have their head and comp moved). This assumption is not compatible with the ideas presented here.

consequence is that, if a head provides the label for a maximal projection, then both instances of that head count for the purposes of c-command. The minimal/maximal distinction is now structurally defined and so the distinction is still there where relevant (cf. Roberts 2008 for discussion), but it is effectively true that heads c-command out of their phrase. Thus in (6), β will c-command both α and θ, but will be c-commanded by neither:

(6)



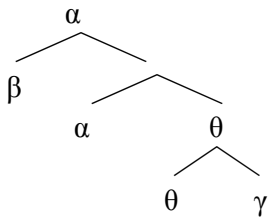
Thus for the purposes of c-command and subsequently the LCA, β_{max} counts as an instance of $\beta_{min.}$

The idea that, under BPS, heads themselves can c-command out of the phrase they project is not a standard one.[7] Nonetheless, under BPS, it is a logical step to take if the label of a phrase in really just another instance of the projecting head. Moreover, as we shall see below, it makes some interesting predictions once we remove the last remaining stipulation from the LCA.

2.5. Antisymmetry

Note that under these assumptions it is not only non-branching complements that are problematic for an LCA based on asymmetric c-command. As the label of a phrase counts as an instance of the projecting head for c-command purposes, every head-complement relationship is actually one of mutual c-command, which might potentially present a problem for linearization. Imagine that θ is a branching constituent, in this case with no specifier (though this makes no difference):

(7)



Even though α asymmetrically c-commands an instance of θ, another copy of θ mutually c-commands α. Thus, whereas Chomsky's version of BPS creates a point of symmetry at the bottom of each tree, this variant re-creates this problem at each point of merge. However, this is not necessarily a problem as long as the LCA can simply ignore instances of mutual c-command, and capitalize on inherent asymmetries:

[7] I do not claim, however, unlike Kayne that specifiers can c-command out of their phrase. This is a welcome result as the empirical evidence Kayne cites in favour of specifier c-command has proven not to be robust and (cf. Ruys 1992 for a discussion).

(8) *Revised LCA*
If α asymmetrically c-commands an instance of ß then α precedes ß.

As such, as long as an instance of a head α asymmetrically c-commands an instance of ß then α will precede ß according to the LCA. Thus this version of BPS is no more problematic than the standard version with respect to mutual c-command, though nor does it solve the base-pair problem in any satisfactory way. I must assume, therefore, as Chomsky does, that in the case of a non-projecting complement the mutual c–command problem will be resolved via null structure or movement (either of the complement itself or of the rest of the phrase, resulting in Complement Stranding).[8]

3. Linearizing specifiers

3.1. The complement/specifier asymmetry

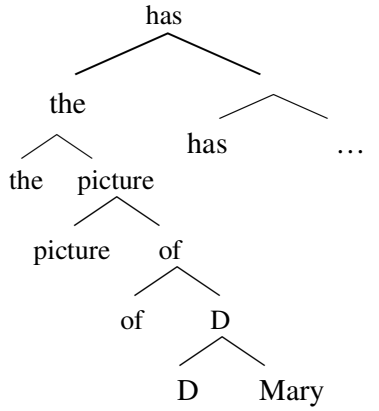
Given the assumptions outlined above, let us reconsider the linearization of complex specifiers such as subjects. Under Kayne's LCA, the complex subject in (10) is linearized with respect to the head *has* via the induction step (identical to the Non-tangling Condition of Partee, Meulen and Wall 1990):

(9) A *reworded (BPS-friendly) version of Kayne's (1994) LCA*
(a)Basic step: If α asymmetrically c-commands an instance of ß, then α precedes ß.
(b)Induction step: If α precedes ß and α dominates γ then γ precedes ß

This ensures that the whole specifier 'the picture of Mary' in (10) precedes the head 'has'.

[8] Of course this does not really resolve the issue of last-pair mutual c-command. Ideally, we would want to show either (i) systematic differences between the placement of branching vs. non-branching complements, or (ii) empirical evidence for the presence of null structure/string-vacuous movement. Chomsky (1995) mentions as evidence languages where (clitic) pronouns occupy a higher position than complex phrasal complements.

(10)



(a) *the>has*
the>picture>of>Mary

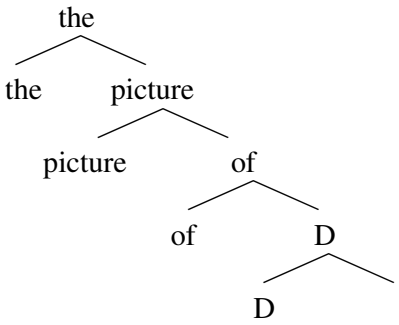
(b) *the* dominates *the>picture>of>Mary*
So *the>picture>of>Mary>has*

According to the induction step, because the head *the* asymmetrically c-commands *has,* everything dominated by *the* must precede *has.*

According to Minimalist principles, this stipulative induction step should ideally be derived from somewhere or else justified conceptually. Uriagereka (1999) proposes an account of it based on Multiple Spell-Out (MSO), whereby the number of applications of Spell-Out is subject to general economy principles. Where a complex phrase is first-merged in a specifier position, it must be spelled out prior to merge, to enable linearization of terminals (null categories like D are ignored by the LCA):

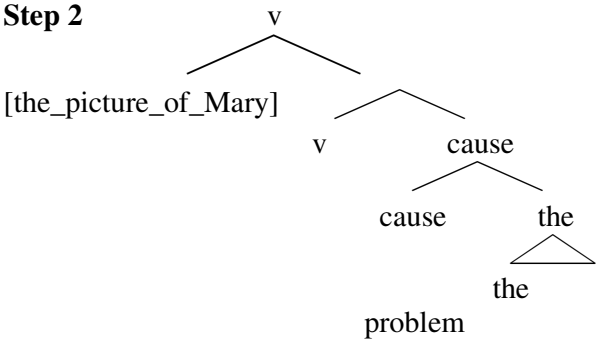
Figure 1: Multiple Spell-Out

Step 1



Mary
Spell-Out
⇒[the_picture_of_Mary],

Step 2



The complex lexical item can then be inserted in a specifier position.

The_picture_of_Mary now behaves like a word and so sub-extraction is impossible. This captures the fact that (first-merged) specifiers, including subjects and adjuncts, are islands:

(11) *__Who__ did [**a picture of** t] cause the problem?
(12) *__Who__ did Mary wash the car [**after she saw** t]?

Uriagereka takes the claim one step further, claiming that MSO accounts for Huang's (1982) Constraint on Extraction Domains, which states that sub-extraction is only

possible from complement positions, all specifiers being islands (underlying subjects and adjuncts *plus* derived subjects), hence the ungrammaticality, also, of (13):

(13) *__Who__ was [__the picture of__ t] taken?

As complex phrases which remain in a complement position can be linearized in one fell swoop with respect to the clausal spine, it follows that specifiers but not complements will display CED effects.

In order for MSO to account for the ungrammaticality of (13), one of the following must hold: either (i) any complex phrase which will end up in a specifier position is spelled out prior to merge, requiring look-ahead, or (ii) spellout is a subcomponent of move, so that a complex phrase which undergoes movement from complement to specifier position is first spelled out. The latter option is unappealing as it requires a vast complication of the operation 'move' which is standardly taken to be comprised only of the independently justified operations copy+merge. The former is counter-derivational and hence highly suspect given Minimalist assumptions. Crucially, it cannot be the case that (13) is ruled out by general economy considerations, contra Uriagereka (1999) and Nunes & Uriagereka (2000), because linearization of such a structure without MSO is in fact possible, as will be shown in the remainder of the section. Rather than being forced by general economy principles, it will be proposed that spell-out of certain nominals actually occurs prior to merge as an arbitrary property of certain functional heads. As such, the interface between the narrow syntax and the PF component is not as sophisticated as Uriagereka proposes. Arbitrary Spell-Out points interact with movement in a fully predictable way.

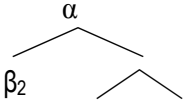
4. An alternative to multiple spellout

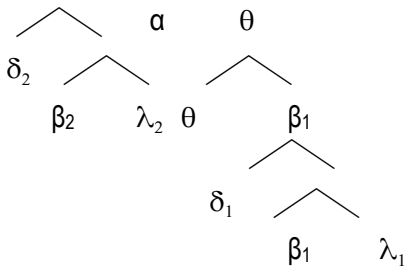
4.1. Copy deletion

Since Chomsky (1993), movement has been recast as copy+remerge in Minimalist syntax. Chomsky himself merely observes that PF always targets (i.e. pronounces) the 'highest' copy in a chain, deleting the lower copy/copies, but remains noncommittal as to why this should be the case. Nunes (1995, 1999), on the other hand, attempts to derive this chain reduction/copy deletion from the LCA, claiming (i) that the LCA cannot apply where $\alpha>\beta$ and $\beta>\alpha$ or where $\alpha>\alpha$ and both copies of α are 'non-distinct' (Chomsky's 1995 irreflexivity condition), and (ii) that it is the lower copy which is deleted because of economy.

Nunes' fundamental insight is that deletion is motivated by the linearizing contradictions present at PF, and exists only to resolve these paradoxes. The non-distinctness of copies means that only one copy of each terminal node can remain at PF. Let us assume, as is standard, that there is a preference to spell-out higher copies for economy reasons. Given our model in which terminal nodes define their own c-command, only *heads* will actually move 'higher' where a complex phrase undergoes movement. The *specifier(s)* and *complement* contained in a complex phase do not move higher in any real sense, in fact their c-command domain remains identical in both positions. Consider the following example:

(14)





In (14), the complex phrase β has moved from a complement to a specifier position. In its derived position, shown as β_2, it asymmetrically c-commands α and θ, whereas in its base position, shown as β_1, it is asymmetrically c-commanded by α and θ. As such, the preference for higher copies means that β_2 is targeted rather than β_1 at the transfer to PF, where copy deletion occurs.

Now consider the specifier δ. Both copies of δ asymmetrically c-command, and so must precede, both β and λ. This means that whichever copy of β/λ is targeted by PF, a higher copy of δ must be chosen. Given the fact that β_2 must be spelled out, it follows that δ_2 will also be targeted. Crucially, there are no direct linearization instructions regarding the ordering of δ with respect to α and θ, but this ordering is possible transitively: δ> β & β> α> θ therefore δ> β> α> θ.

The case of the complement λ is different. Like the specifier δ, its c-command domain is unaltered by movement. However, unlike δ it carries no instruction to precede anything. In both of its positions λ asymmetrically c-commands nothing. Moreover, in its derived position, shown as λ_2, the complement cannot actually be ordered with respect to α and θ. It neither asymmetrically c-commands nor is asymmetrically c-commanded by either terminal node. Because it does not precede β, this ordering cannot be resolved transitively (even if we assume that the mutual c-command problem is resolved).

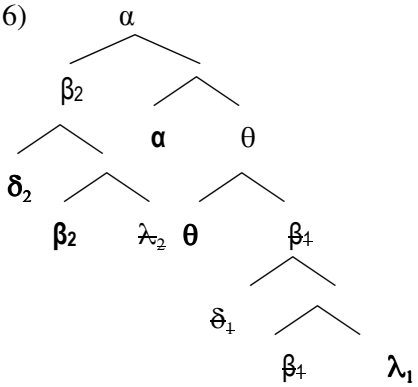
4.2. Complement Stranding

This situation can, however, be resolved in the following way, assuming, following Bošković (2001), that scattered deletion is available as a last resort, where full copy deletion is blocked:

(15) Delete lower copy of δ_1, β_1 & moved copy of λ_2 = δ >β> α>θ>λ (total order)

The surface effect of this copy deletion is one of Complement Stranding:

(16)



Spells out ⇒ **δ β α θ λ**

By targeting the base-generated copy of λ PF is able to achieve a total linear ordering or terminal without the need for the inductive step. The deletion of copies is therefore purely a matter for PF, asymmetric c-command plays a role indirectly via the LCA, but chain reduction itself is not regulated by c-command, as many moved terminals will not actually c-command their traces.

4.3. Summary and predictions

In short, it has been claimed that there are two distinct ways to solve the problem of the linearization of specifiers at PF, as summarized here:

(17) To enable 'absolute order' at PF, either (i) or (ii) must hold, where a complex constituent undergoes movement to a specifier position:
(i) a complex phrase is spelled out prior to movement and moved as a lexical item, giving CED effects (following Uriagereka 1999).
(ii) the base-generated copy of the complement contained in the moved phrase must be targeted by PF (henceforth Complement Stranding).

This makes the robust prediction that all pied-piped complements contained in specifiers will be islands, but stranded complements will not be. Initial investigations suggest that this appears to hold:

(18) a. [Many books about dieting] have been released [~~many books about dieting~~]
b. [Many books ~~about~~ ~~dieting~~] have been released [~~many books~~ about dieting][9]

(19) a. *Which topic have **many books about** been released?
b. Which topic have **many books** been released **about**?

Note that a further interesting prediction arises from this proposal. According to Uriagereka, complex phrases which are first-merged as specifiers should always be

[9] It is crucial to control for cases where a PP complement of N is re-analysed as a complements of V (cf. Davies and Dubinsky 2003, citing Bach and Horn (1976). While this is potentially the case with 'write', it is not true of 'read' or 'release':
(1) They wrote [NP it] [PP about Nixon]

spelled out prior to merge. This is because the complement contained in a hierarchically structured first-merged specifier will be impossible to linearize with respect to the other terminal nodes in the sentence even with Complement Stranding (in the absence of 'the induction step' of the LCA). As such, all first-merged specifiers should be islands.

In our terms, many discontinuous constituents are derived via stranding (including extraposition, discontinuous DPs and whPs etc.). It follows than that stranding will only be possible where a phrase is not spelled out prior to merge, hence. first-merged as specifiers will not give rise to discontinuity of this kind. Again, the data seem to support this prediction, as will be shown below. In the following section I detail several instances of Complement Stranding and show that initial research suggests that they display the predicted properties.

5. Empirical support for Complement Stranding

5.1. Discontinuous indefinites[10]

Roughly, the generalisation in English is that indefinite, but not definite or specific, nominals and complex wh-constituents allow stranding of their complements upon A/A-bar movement:

(20) a. **Many pictures** have been taken **of John**. [11]
b. **Many pictures of John** have been taken.
(21) a. **A film** has come out **about Bush**.
b. **A film about Bush** has come out.
(22) a. **Which book** did you read **about Chomsky**?
b. **Which book about Chomsky** did you read?
(23) a. **How many pictures** were taken **of him**?
b. **How many pictures of him** were taken?
(24) a. **How fond of Sally** are you?
b. **How fond** are you **of Sally**? [Baltin (1981: 262)]
(25) a. **How certain that the Mets will win** are you?
b. **How certain** are you **that the Mets will win**? [Baltin (1981: 262)]

I take 'a', 'many', 'which', 'how many' and 'how' to be specifiers of lexical heads (N/A) rather than D elements. This is not a naïve assumption and one which requires justification. One piece of support for this approach to 'a' comes from the fact that in English, indefinite articles appear in contexts of predication:

(26) He is a doctor.

---

(2) *They read [NP it] [PP about Nixon]
(3) *They released [NP it] [PP about Nixon]

[10] Note that extraposition of relative clauses (RCs) arguably results from the same kind of movement, if relative clauses are complements, as Kayne proposes. I do not discuss them here because (i) RCs are inherently islands, and so the islandhood test is rendered obsolete. I leave this matter to future research.

[11] Ian Roberts points out to me that it is always the whole complement of N which is stranded, it is not possible to pied-pipe part of N's complement and strand the rest:

(4) A film was made about a book about Chomsky.
(5) *A film about a book was made about Chomsky.

It is well-known that in many other languages, for instance the Romance languages, an indefinite article is impossible here:

(27)	Il	est	(*un)	médecin	[French]
	He	is	a	doctor	

Bošković (2008) argues convincingly, on a broad syntactic empirical basis, that in some, but not all, languages NPs can serve as arguments, advocating a Chierchia-style type-shifting semantics (cf. Chierchia 1998).[12] As such, it would not be problematic for English indefinite nominals to be NPs rather than DPs and still be arguments. Moreover, it is commonly held that indefinite articles occupy a lower position than definite articles (cf. Heycock & Zamparelli 2005 for a discussion) and so to take this position to be spec N, rather than the specifier of a lower functional head is not such a large step (cf. the papers in Craenenbroeck to appear). Pending a more thorough assessment of its validity, I therefore take these elements to be specifiers of the head N/A and for the examples above to be examples Complement Stranding.

Crucially, as predicted, the stranded complements of moved NPs are not islands:

(28) a. *Who was **a picture of** seen?
b. ?Who was **a picture** seen **of**?
(29) a. *Who did **a book about** appear?
b . Who did **a book** appear **about**?
(30) a. *Who did a **book about** appear to have been written?
b. Who did **a book** appear to have been written **about**?

In the case of the stranded complements of wh-constituents, there is a definite contrast between extraction from pied-piped and stranded complements, though it is perhaps not as sharp:

(31) a. This is the team that you said **how certain** you were that **the METS would beat**.
b. *This is the team that you said **how certain that the METs would beat** you were.
(32) a. ?This is the girl that you told me **how fond** you were **of**.
b. *This is the girl that you told me **how fond of** you were.
(33) a. ?That's the theory that you asked me **which book** I had read **on**.
b. *That's the theory that you asked me **which book on** I had read.
(34) a. This is the vase that you told me **how likely** John is **to break t**
b. *This is the vase that you told me **how likely to break t** John is

Finally, as predicted, it also seems to hold that this kind of stranding is not possible with complex indefinites first-merged as specifiers of vP:

(35) a. **Pictures of Diana** always provoke a scandal.

[12] Although he does not actually propose this analysis of English, the spirit of his analysis is in line with the present claim.

b. *__Pictures__ always provoke a scandal __of Diana__.
c. *__Pictures__ always __of Diana__ provoke a scandal

(36) a. *__Which book__ has provoked an outcry __on Chomsky__?
b. *__Which book__ has __on Chomsky__ provoked an outcry?

(37) a. *__A film__ has provoked a scandal __about childhood__?
b. *__A film__ has __about childhood__ provoked a scandal?

(38) a. *__Which book__ will depict his childhood __about Dickens__?
b. *__Which book__ will __about Dickens__ depict his childhood?

As such, these constructions appear to provide empirical evidence for Complement Stranding of the type predicted, suggesting that the latter is available as an alternative means of linearizing complex specifiers in English.[13]

Optionality of the sort displayed above is likely to indicate a difference in underlying structure, if the basic tenets of Minimalism hold. As it is approximately true that definite DPs are always islands in English, even in object position, let us assume that it is actually the presence of D which triggers Spell-Out of the type proposed by Uriagereka, before a DP is first-merged either in a specifier or complement position.[14] As such definite DPs pattern with definite DP subjects in being strong islands regardless of movement. Crucially, because it is already spelled, where a definite DP moves, it will not strand its complement.

NPs, on the other hand, are not inherently islands. Again it is approximately true that indefinite NPs are not islands in English in complement position:

(39) **Who** did you see **a picture of**?

This implies that NPs are not spelled out separately to the predicate which they combine with, at least where they are merged in a complement position. The apparent optionality

[13] Theresa Biberauer (personal communication) points out that ditransitives may be a problem for this approach, as it is standardly assumed that themes are merged as specifiers of V (cf. Harley 2000) but indefinites of this kind seem to allow Complement Stranding:

(6) a. Which picture did you send to Mary of me?
b. ?Which picture did you send of me to Mary?
(7) a. A picture was sent to Mary of me.
b. ?A picture was sent of me to Mary.
(8) a. ?Which picture did you send Mary of me?
b. *Which picture did you send of me Mary?
(9) a.*A picture was sent Mary of me.
b.*A picture was sent of me Mary.

The structure of ditransitives is a highly complex and moot point and it is by no means clear that they are a robust counterexample. In fact, a more obvious extension would be to ditransitive nominals (as pointed out by David Pesetsky). Nominals such as transfer clearly select two complements, but can strand only the second:

(10) A transfer of funds to my account will take place tomorrow.
(11) A transfer of funds will take place to my account tomorrow.
(12) Where will a transfer of funds take place to tomorrow?
(13) *A transfer will take place of funds to my account tomorrow.

This kind of evidence could eventually provide us with useful insights as to the structure of ditransitives. I leave both of these matters for future research.

[14] Actually the facts are rather more complex than this, as Davies and Dubinsky (2003) have shown. I leave the implications of these complexities for future research.

of Complement Stranding is therefore due to the fact that NPs can be concealed DPs. Consider the ambiguity of the following sentence:

(40) Two men are in the garden.

It has long proved problematic that sentences such as (40) are ambiguous between a strong and weak reading of the subject indefinite (cf. Diesing 1992 *inter alia*). One way of resolving this is to say that the strong reading of the subject involves a covert DP, whereas the weak reading involves a true NP:

(41) [$_{DP}$ Two men] are in the garden. = Two of the men are in the garden.
(42) [$_{NP}$ Two men] are in the garden. = There are two men in the garden.

The prediction would be that a weak reading of an NP would lead to obligatory Complement Stranding, whereas the strong reading would prevent it.

One context which favours a true weak indefinite reading is expletive passive constructions. For some reason, in this construction in English, the object is required to move to an intermediate, preverbal position, unlike in the other Germanic languages (cf. Vikner 1995 *inter alia*). In such cases, Complement Stranding is much preferred:

(43) a. There were many pictures taken of trees
b. *There were many pictures of trees taken.
(44) a. There was an examination carried out of the patient.
b. *There was an examination of the patient carried out.

There is, therefore, evidence for the existence of NP and DP subjects in English, only the former of which give rise to Complement Stranding.

6. Deriving the Final-over-Final Constraint

In this section I will argue that Complement Stranding, for which we have already seen independent theoretical and empirical justification actually derives the Final-over-Final Constraint for free. Holmberg (2000), Biberauer, Holmberg and Roberts (2007, 2008) and Biberauer, Newton & Sheehan (to appear-a, -b) have shown broad empirical support for what is descriptively termed the Final-over-Final Constraint (FOFC), characterized as follows:

(45) *The Final over Final Constraint* (Holmberg 2000)
A head-initial XP cannot be (immediately) dominated by a head-final phrase.

FOFC is supported by typological word order trends as well as patterns of borrowing and diachronic change. Details of the supporting data can be found in the papers cited above, but a summary is included here for reference. Note crucially that in all cases of a FOFC-related gap, the inverse disharmonic order *is* attested:

Language-internal gaps in languages with variable word orders:
a) *[N Complement] P in Finnish (Holmberg 2000)

Typological cross-linguistic lack of:

b) *[V Object] Aux (Julien 2002, Biberauer, Holmberg and Roberts 2007)
c) *[V Object] C (Hawkins 1990, Dryer 1992)
d) *[C TP] V (Hawkins 1994, Dryer in press, Sheehan 2008)
e) *[Q TP] C (Biberauer, Newton and Sheehan, to appear-a, Davison 2007)
f) *[Asp V] T (Julien 2007)
g) *[P DP] V (modulo Germanic, Dryer 1992)

Evidence from diachronic change (Biberauer, Newton and Sheehan to appear a)

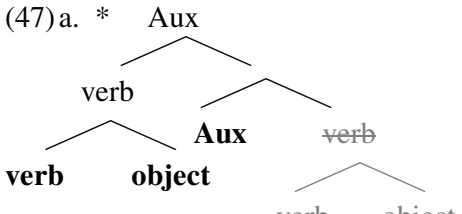
i. Change from head-final to head-initial must proceed top-down – true of the history of English, Afrikaans and French.
ii. Change from head-initial to head-final must proceed bottom-up – true of the history of Ethiopian Semitic.

If the LCA holds, it follows that head-final languages must be derived via movement from a universal spec-head-comp order. If we take a roll-up approach to head-finality, in the spirit of Kayne (1994), Zwart (1994), Holmberg (2000), Julien (2002) and Biberauer, Holmberg & Roberts (2007, 2008), triggered by Holmberg's (2000) c-selection, then in simple terms FOFC equates to the following observation:[15]

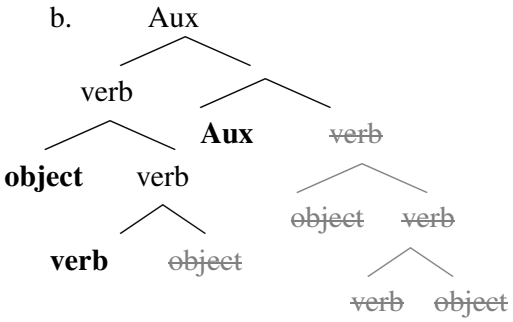
(46) *FOFC in roll-up terms*⇒ roll-up movement must begin at the bottom of the tree.

Let us consider the example of *[V Object] Auxiliary:

(47) a. *



b.



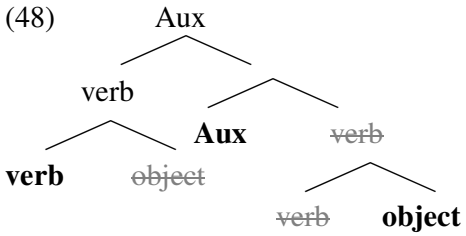
Descriptively, in these terms, (47) is out because roll-up movement does not begin with the complement of the verb. In (b) on the other hand, movement has begun

[15] It is important to note that while all of these authors assume some kind of roll-up movement, their assumptions differ somewhat as to the position of the object and subject. For Zwart, the object moves to spec AgrO at least in Dutch, for Roberts the position targeted by the object varies depending on the construction/language, and only for Julien is the object either moved to or first-merged in spec VP. Moreover, for Julien (2002) the presence of a specifier at any point blocks roll-up movement and so VP never moves to the specifier of v, even in a head-final language. This is incompatible with the proposal made here, as will become apparent below. In fact, under the version of c-command adopted here, the presence of more than one specifier is not a problem, so no such complementarity obtains between comp-to-spec movement and the merging of a specifier. In this much, the system proposed here is less restrictive than that proposed by Julien. Note that it would not be problematic for the approach outlined here if the object were to target a higher position than spec VP, as long as the head which attracts the object also triggers comp-to-spec.

at the bottom of the tree with comp-to-spec movement of the lowest complement, the object to the specifier of V under c-selection. In (b), on the other hand, the V phrase has moved to spec Aux under C-selection, without the object having moved, resulting in a FOFC-violating order. In descriptive terms, it seems that comp-to-spec movement can stop at any point, but cannot start except at the bottom of the spine.

Now recall the kind of movement which results in Complement Stranding: the movement of a non-spelled-out complex phrase from complement to specifier position. If head-finality is derived by comp-to-spec movement then FOFC can be seen merely as a further example of this effect. While there is nothing banning the movement of a head-initial phrase under C-selection, at PF the moved complement contained in the phrase will be deleted to enable linearization and the lower copy targeted, resulting in what we have called Complement Stranding. In these terms the derivation in (47) is not ungrammatical *per se* (contra Biberauer, Holmberg and Roberts 2007, 2008), but its PF order is not V-Obj-Aux, but rather V-Aux-Obj, as in (48).

(48)



spells out ⇒verb aux object

According to Biberauer et al. (2008) the order V-Obj-Aux "is required for CP-complements in German, Dutch, Afrikaans and their dialects; possible with PP-complements in Dutch and Afrikaans and, to a lesser extent, German; also possible with DPs in Old English and Old Norse". The only way to derive the FOFC-violating PF order V-Obj-Aux would be for the V+Obj to be spelled out prior to movement.[16] It is thus (correctly) predicted that the surface order V-Obj-Aux will occur only where the V+Obj is fronted under A/A-bar movement to a position higher than the specifier of Aux after it has been spelled out via cyclic transfer.

Under these assumptions, FOFC, a gap in attested word orders in disharmonic languages results directly from the fact that movement of a complex phrase prior to spellout strands its complement. The only exception to this will be where a phrase is spelled out prior to merger for independent reasons (triggered by a set of functionsal heads which trigger strong Islandhood). This is arguably the case with definite DPs in English.[17] Note that this makes the prediction that all apparent FOFC violations with either involve strong islands. Ultimately, the empirical validity of these predictions remains to be adequately tested.[18]

[16] This is arguably what happens in at least some VOS languages (cf. Aldridge 2006)

[17] Apparent exceptions to this would have to be examined more closely and the status of the 'determiner' assessed. This is a matter of ongoing research.

[18] I do not discuss the other exceptions to FOFC discussed by Biberauer et al. (2007, 2008) for reasons of space. Ultimately if this approach is at all valid it should also provide a non-stipulative account of these exceptions. I leave this matter to future work.

7. Conclusions and implications

It has been argued that a Minimalist interpretation of BPS coupled with a simplified version of Kayne's LCA predicts that a complex non-linearised phrase will strand its complement upon movement from a complement to a specifier position. We have seen evidence for this effect from various types of discontinuous structures in English. Crucially, in all cases a stranded complement is not a strong island, whereas a pied-piped complement is. It had then been proposed that if head-finality is also movement-derived then the same effect is expected to hold with c-selection. Movement of a non-linearised complex XP is expected to display Complement Stranding upon movement under c-selection. This essentially captures Holmberg's (2000) Final-over-Final Constraint. Unless a complement is first moved to the specifier of a non-linearised phrase, that complement will be stranded when the phrase containing it moves.

The implications of this analysis are far-reaching. Perhaps the most interesting possibility is that it might ultimately allow a unification of head and phrasal movement, the former occurring where a head-initial phrase with no specifier undergoes local comp-to-spec movement stranding its complement. Travis' (1982) observation that head-movement is always very local would then be due to the fact that c-selection is necessarily local (cf. Pesetsky and Torrego 2001, 2004).

The most direct implication of the analysis is that it provides empirical evidence in favour of movement-derived head-finality. The fact that we see the same stranding effect with A & A-bar movement and c-selection is strong support for comp-to-sepc movement under c-selection. It seems impossible to unify all these effects without something akin to the LCA.

References

Abels, K. (2007). Towards a restrictive theory of (remnant) movement, *Linguistic Variation Yearbook*, 7 (1): 53-120.

Abels, K. & Neeleman, A. (2006). 'Universal 20 without the LCA': UCL.

Baker, M. (1988). Theta theory and the syntax of applicatives in Chichewa, *Natural Language & Linguistic Theory*, 6 (3): 353-89.

Baltin, M. R. (1981). Strict Bounding, Baker, C.L. and McCarthy, J. (eds.), In *The Logical Problem of Language Acquisition*, Cambridge, Mass: MIT Press.

Biberauer, T., Holmberg, A. & Roberts, I. (2007). Disharmonic word-order systems and the Final-over-Final-Constraint (FOFC), Bisetto, A. and Barbieri, F. (eds.), In the proceedings of *XXXIII Incontro di Grammatica Generativa.*

Biberauer, T., Holmberg, A. & Roberts, I. (2008). Structure and linearization in disharmonic word orders, Chang, C.B. and Haynie, H.J. (eds.), In the proceedings of *The proceedings of 26th Western Coast Conference on Formal Linguistics*, Somerville, MA: Cascadilla Proceedings Project.

Biberauer, T., Newton, G. & Sheehan, M. (to appear-a). Impossible changes and impossible borrowings: the Final-over-Final Constraint, Willis, D. (ed.) In the proceedings of *Continuity and Change in Grammar*, Cambridge: John Benjamins.

Biberauer, T., Newton, G. & Sheehan, M. (to appear-b). Limiting synchronic and diachronic variation and change: the Final-Over-Final Constraint, *Language and Linguistics*.

Bošković, Ž. (2008). What will you have, DP or NP?, In the proceedings of *NELS 37*.

Cecchetto, C. (2007). Some preliminary remarks on a 'weak' theory of linearization, *Annali Online di Ferrara - Lettere*, 1: 1-13.

Chierchia, G. (1998). Reference to Kinds across Languages, *Natural Language Semantics*, 6: 339-405.

Chomsky, N. (1970). Remarks on Nominalization, Jacobs, R.A. and Rosenbaum, P.S. (eds.), In *Readings in English Transformational Grammar*, Waltham, Mass: Ginn.

Chomsky, N. (1995). Bare phrase structure, Webelhuth, G. (ed.) In *Government and binding theory and the minimalist program : principles and parameters in syntactic theory*, Cambridge, Mass.: Blackwell.

Chomsky, N. (2001). Derivation by phase, Kenstowicz, M. (ed.) In *Ken Hale: a life in language*, Cambridge, Mass.: MIT Press.

Cinque, G. (2005). Deriving Greenberg's Universal 20 and Its Exceptions, *Linguistic Inquiry*, 36 (3): 315–32.

Craenenbroeck, J. v. (to appear). *Alternatives to Cartography*, Berlin; New York: Mouton de Gruyter.

Davies, W., D. & Dubinsky, S. (2003). On extraction from NPs, *Natural Language and Linguistic Theory*, 21: 1-37.

Diesing, M. (1992). *Indefinites*, Cambridge, Mass.: MIT Press.

Dryer, M. (In press). The branching direction theory revisited, Scalise, S., Magni, E. and Bisetto, A. (eds.), In *Universals of Language Today*, Berlin: Springer.

Dryer, M. S. (1992). The Greenbergian word order correlations, *Language*, 68 (1): 81-138.

Epstein, S. D. (1999). Un-Principled Syntax and the Derivation of Syntactic Relations., Epstein, S.D. and Hornstein, N. (eds.), In *Working Minimalism*, Cambridge, Mass.: MIT Press.

Epstein, S. D., Groat, E. M., Kawashima, R. & Kitahara, H. (1998). *A Derivational Approach to Syntactic Relations* Oxford: OUP.

Epstein, S. D. & Seely, T. D. (2006). *Derivations in minimalism*, Cambridge: Cambridge University Press.

Hawkins, J. A. (1994). *A performance theory of order and constituency*, Cambridge ; New York: Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*, Oxford ; New York: Oxford University Press.

Heycock, C. & Zamparelli, R. (2005). Friends and Colleagues: Plurality, Coordination, and the Structure of DP, *Natural Language Semantics*, 13 (3): 201-70.

Holmberg, A. (2000). Deriving OV Order in Finnish, Svenonius, P. (ed.) In *The Derivation of VO and OV*, New York: Oxford University Press.

Jackendoff, R. S. (1977). *X-bar syntax : a study of phrase structure*, Cambridge, Mass.: MIT Press.

Kayne, R. S. (1994). *The antisymmetry of syntax*, Cambridge, Mass.: MIT Press.

Koopman, H. J. & Szabolcsi, A. (2000). *Verbal complexes*, Cambridge, Mass.: MIT Press.

Lasnik, H. (2001). Subjects, objects and the EPP, Davies, W., D. and Dubinsky, S. (eds.), In *Objects and other subjects*, Amsterdam: Kluwer.
Moro, A. (2000). *Dynamic antisymmetry*, Cambridge, Mass.: MIT Press.
Nunes, J. & Uriagereka, J. (2000). Cyclicity and extraction domains, *Syntax*, 3 (1): 20-43.
Partee, B. H., Meulen, A. G. B. t. & Wall, R. E. (1990). *Mathematical methods in linguistics*, Dordrecht ; Boston: Kluwer Academic.
Pesetsky, D. & Torrego, E. (2001). T-to-C movement: causes and consequences, Kenstowicz, M. (ed.) In *Ken Hale: a life in language*, Cambridge, Mass: MIT Press.
Richards, M. (2004). Object Shift and Scrambling in North and West Germanic: A Case Study in Symmetrical Syntax, Cambridge University.
Ruys, E. (1992). The Scope of Indefinites, Utrecht University.
Sheehan, M. (2008a). 'The Final-over-final constraint and phase theory', Newcastle: Newcastle University.
Sheehan, M. (2008b). 'The Final-over-final constraint: trend or principle': Newcastle University.
Travis, L. (1984). Parameters and Effects of Word Order Variation, Ph.D thesis, MIT.
Uriagereka, J. (1999). Multiple Spell-out, Hornstein, N. and Epstein, S.D. (eds.), In *Working minimalism*, Cambridge, MA: MIT Press.
Vikner, S. (1995). *Verb movement and expletive subjects in the Germanic languages*, New York ; Oxford: Oxford University Press.
Zwart, C. J.-W. (1994). Dutch is Head Initial, *The Linguistic Review*, 11 (3/4): 377-406.
Zwart, C. J.-W. (2002). The antisymmetry of Turkish, In the proceedings of *Generative Grammar @ Geneva 3*.

Michelle Sheehan
School of School of English Literature, Language and Linguistics
Percy Building
University of Newcastle upon Tyne
NE1 7RU
United Kingdom

michelle.sheehan@ncl.ac.uk