

# Last Resort with Move and Agree in Derivations and Representations

Željko Bošković

University of Connecticut

## 1. Introduction

Minimalism assumes language consists of a lexicon and a computational system, with the latter embedded in two performance systems: articulatory-perceptual and conceptual-intentional. Two linguistic levels, PF and LF, interface with the performance systems. A computation converges at the interface levels if it contains only legitimate PF and LF objects. However, we cannot define linguistic expressions simply as PF/LF pairs formed by a convergent derivation and satisfying interface conditions. The operations of the computational system that produce linguistic expressions must be optimal, in the sense that they must satisfy some general considerations of simplicity, often referred to as Economy Principles. One of them, the Last Resort Condition (LR), prohibits superfluous steps in a derivation. It requires that every operation apply for a reason. It has often been argued that a similar condition constrains representations, prohibiting superfluous symbols. These conditions require that derivations and representations in some sense be minimal. The goal of this chapter is to discuss the working of LR, as it applies to both derivations and representations. I will start with the former, examining how LR applies to both Move and Agree.

Before doing that, let me point out that there are various ways of implementing LR formally—it can be stated as an independent condition or built into the definition of Move and Agree. However, regardless of how it is formalized as soon as the issue of driving force for a syntactic operation is raised LR considerations become relevant, since the issue of driving force really makes sense only given LR. Much of what follows will in fact be more concerned with the driving force of syntactic operations than the exact implementation of LR.

## 2. The ban on superfluous steps in derivations

### 2.1 Last Resort and Move

Minimalism has insisted on the last resort nature of movement from its inception: In line with the leading idea of economy, movement must happen for a reason, in particular, a formal reason. Case provides one such driving force. Consider (1).

- (1) Mary is certain *t* to leave

*Mary* cannot be case-licensed in the position of *t*. Raising to matrix SpecIP rectifies its case inadequacy, since the raised position licenses nominative. Once *Mary* has been case-licensed, it is no longer available for A-movement, to a case or a non-case position. This follows from LR, if A-movement is driven by case considerations. Since *Mary* is case-licensed in the position of *t* in (2), LR blocks further movement of *Mary*.<sup>1</sup>

- (2) a. \*Mary is certain *t* will leave  
b. \*The belief Mary to be likely *t* will leave

---

<sup>1</sup> (2) involves A-movement from a CP, which is often assumed to be disallowed. See, however, Bošković (2007) and references therein for evidence that such movement is in principle possible, even in English.

One fruitful line of research regarding LR concerns the issue where the formal inadequacy driving movement lies. The options are: (a) always in the target (pure Attract); (b) always in the moving element (Greed); (c) in the target or in the moving element (Lasnik's 1995 Enlightened Self Interest). Greed was the earliest approach (Chomsky 1993), revived recently in Bošković (2007). Under this approach X can move only if X has a formal inadequacy, and if the movement will help rectify the inadequacy. Under pure Attract, the target head always triggers movement (Chomsky 1995), which means the target must always have a formal inadequacy to be rectified by the movement. Under this approach, movement of *Mary* in (1) is driven by T/I: Tense has a property (e.g. the EPP or case feature) that must be checked against an NP which triggers the movement of *Mary* (*Mary*'s case-checking is merely a beneficial side effect of the satisfaction of the attractor's requirement). If the trigger is T's case (i.e. Bošković's 1997 Inverse Case Filter, which requires traditional case assigners to check their case), (2a-b) can still be accounted for: the problem with (2a) is that *Mary* is case-checked in the embedded SpecIP so that the matrix T's case remains unchecked, and the problem with (2b) is that nothing triggers the movement. A question, then, arises why (3) is unacceptable (the question also arises under the Greed approach).

(3) \*the belief to be likely *Mary* will fail the exam.

In a framework that adopts the EPP, where the EPP drives movement, (3) is easy. However, (2) is problematic: the LR account cannot be maintained since there is reason for movement of *Mary*, namely the EPP. If the EPP/Inverse Case Filter accounts are combined, (3) and (2a) can be handled (as EPP/Inverse Case Filter violations respectively; note that A-movement in (1) is now redundantly driven by the EPP/Inverse Case Filter), but (2b) is still problematic. In other words, something additional needs to be said under the pure Attract account. Recall that under the Inverse Case Filter version of this account, (3) is at issue: for accounts of (3) that conform with this account, and which can also be extended to the Greed approach to LR, see Epstein, Pires, and Seely 2004, Bošković 2002a, among others. As for the EPP account (or the combined EPP/Inverse Case Filter), the additional assumption Lasnik (1995) makes to bring (2) in line is that once the case feature of an NP (like *Mary* in (2)) has been checked, the NP is no longer available for A-movement. Note that the assumption is also necessary under the option (c) from above. As long as we allow the target to drive movement, we have to deal with the question of why (2) is unacceptable. The conceptually unfortunate consequence of this account is that it basically brings back Greed into the system that was intended to eliminate it.

To capture the facts in question within a target-driven system Chomsky (2000) posits the Activation Condition, which says X can move only if X has an uninterpretable feature, i.e. a formal inadequacy. The approach is still sneaking in Greed into a system where movement is supposed to be target-driven. In fact, under this approach something essentially has to be wrong both with the target and the moving element in order for movement to take place.

Let us now compare Chomsky (2000) and Bošković (2007) in more detail, as representatives of target-driven and moving element driven approaches. Both of these works adopt the *Agree* account of traditional covert dependencies. Under *Agree*, two elements, a probe, which initiates an *Agree* operation, and its goal establish a feature checking operation at a distance without actual movement. This is all that happens in traditional covert dependencies. Chomsky assumes that *Agree* is a prerequisite for *Move*. Before *Move* takes place X and Y

establish an Agree relation, which is followed by movement if X is specified with an EPP property. This property of the target is what drives movement for Chomsky.

Bošković, on the other hand, places the trigger for movement on the moving element. In a phase-driven multiple spell-out system, where phases are heads whose complements are sent to spell-out (Chomsky 2001), element X undergoing movement moves from phase edge to phase edge until its final position, the underlying assumption being that if X is ever going to move, it cannot be contained in a unit that is shipped to Spell-Out. X then has to move to SpecYP, where YP is a phase, in order not to get caught in a spell-out unit. The analysis implies that there is some kind of marking on X indicating its need to move. So, how do we know that X will need to move? The question is not innocent, since in many cases what is assumed to trigger movement of X may not even be present in the structure at the point when X needs to start moving. To deal with such cases, Bošković (2007) argues the marking indicating the need for movement, which is standardly taken to be a property of the target (the EPP property of Chomsky 2000, 2001), should be placed on the moving element, not on the target.

To illustrate, consider (4).<sup>2</sup>

(4) What<sub>i</sub> do you think [<sub>CP</sub> t<sub>i</sub> [<sub>C'</sub> that Mary bought t<sub>i</sub>]]?

Chomsky's (2000) account of (4) is based on the PIC, which says only the edge (Spec/head positions) of a phase is accessible for movement outside of the phase. Given the PIC, since CP is a phase, *what* can only move out of the CP if it first moves to SpecCP. This movement is implemented by giving *that* the EPP property (Chomsky assumes complementizer *that* may, but does not have to, have the EPP property), which is satisfied by filling its Spec position. The EPP then drives movement to SpecCP, after which *what* is accessible for movement outside the CP.

(5) raises a serious problem for this analysis, given the derivation on which we have chosen the EPP option for *that*, just as in (4).

(5) \*Who thinks what that Mary bought?

To deal with this, Chomsky (2000, 2001) makes the assignment of an EPP property to heads that do not always require a Spec conditioned on it being required to permit successive cyclic movement. The embedded clause head in (4) can then be assigned the EPP property, since this is necessary to allow successive cyclic movement. However, this is disallowed in (5) since the assignment is not necessary to permit successive cyclic movement. The obvious problem for this analysis is look-ahead. Both (4) and (5) at one point have the structure in (6).

(6) [<sub>CP</sub> what<sub>i</sub> [<sub>C'</sub> that Mary bought t<sub>i</sub>]]

To drive movement to SpecCP, complementizer *that* must be given the EPP property at the point when the embedded clause is built. But at that point we do not know whether the assignment of the EPP property will be needed to make successive cyclic movement possible. We will know this only after further expansion of the structure. If the structure is expanded as in (5), it won't be needed, hence disallowed, and if it is expanded as in (4), it will be needed, hence allowed. So, at the point structure building has reached in (6) we need to know what is going to happen in the matrix clause, an obvious look-ahead problem.

---

<sup>2</sup> In what follows, I ignore vP as a phase for ease of exposition.

The problem is quite general. To appreciate this, consider (7), where X is a cyclic head (and XP a phase) and Y needs to undergo movement to W. In accordance with the Activation Condition Y has an uninterpretable feature (*uK*), which makes it visible for movement.<sup>3</sup> (8) represents the same scenario before W enters the structure.<sup>4</sup>

- (7)    W [<sub>XP</sub> ...X... Y]  
          *uF*                    *iF*  
          K                      *uK*  
          EPP
- (8)    [<sub>XP</sub>     ...X..... Y]  
                                  *iF*  
                                  *uK*

Since XP is a phase, given the PIC, if Y is to move outside of XP it first must move to SpecXP. In Chomsky's system this is implemented by giving X the EPP property to drive movement to SpecXP, with the further proviso that X can be given the EPP property only if this is needed to make successive cyclic movement possible. We then need to know at point (8) that W will enter the structure later, as in (7). Let us see how the look-ahead problem can be resolved. The problem here is that the EPP diacritic indicating Y has to move to SpecWP is placed on W, given that we need to know that Y will be moving before W enters the structure. The problem is quite general under the EPP-driven movement approach. The gist of the look-ahead problem that arises under this approach is that the EPP diacritic indicating Y moves is placed on an element (W) other than the one that is undergoing the movement in question, but Y often needs to move (i.e. start moving) before W enters the structure. The conclusion to be drawn from this state of affairs is obvious: we have been wrong in placing the diacritic indicating the need for movement on the target (W)—the diacritic should be placed on the moving element (Y). Bošković (2007) implements this as follows: It is standardly assumed that a probe must c-command the goal, and that the probe must have a *uK*; otherwise, there would be no need for it to function as a probe. Following an insight of Epstein and Seely (1999), Bošković (2007) assumes the correlation between functioning as a probe and having a *uK* is a two-way correlation: just like a probe must have a *uK*, a *uK* must function as a probe.<sup>5</sup> In other words, checking of a *uK* on X requires X to function as a probe (i.e. c-command the checker). This means Y in (7)-(8) will need to undergo movement outside of XP to license *uK*. In fact, Agree would not suffice for that even if Y is located in SpecXP. Most importantly, we now know Y will need to undergo movement outside of XP before W enters the structure: already at point (8) we know the structure will crash due to *uK* unless Y moves outside of XP. In other words, Y will have to move to a position c-commanding the *uK* licenser to check the feature. Since the *uK* licenser is not present within XP, this means Y will have to move outside of XP, hence has to move to SpecXP. Notice also that Bošković (2007) argues for the following formulation of LR: X can undergo movement iff

<sup>3</sup> Bošković (2007) shows the Activation Condition follows for Move as a theorem. As for Agree, I argue it does not hold for it.

<sup>4</sup> K is either checked as a reflex of F-feature checking between W and Y (see below for discussion of reflex feature checking) or W has a K feature that can check the K feature of Y. For ease of exposition, I adopt the latter option, returning to the issue below.

<sup>5</sup> The assumption has many empirical consequences; see Epstein and Seely (2006) and Bošković (2007).

without the movement, the structure will crash. Movement to SpecXP then conforms with LR although it does not involve feature checking between Y and X, a desirable result in light of arguments against feature checking in intermediate positions discussed below. Eventually, Y will have to move to a position c-commanding W. Given the Shortest Move requirement, it will move to the closest position c-commanding W, which means SpecWP.

The analysis also deduces generalized EPP effects. We have already seen that there is no need to mark intermediate heads (X in (7)), with the EPP property to drive movement to their Specifiers. The movement takes place so that the element undergoing movement escapes being sent to Spell-Out, which would freeze it for the possibility of movement, leaving its  $uK$  unchecked. Now, the generalized EPP effect is deduced in its entirety. Thus, Y in (7) now has to move to SpecWP even if W does not have the EPP property, which is then dispensable.<sup>6</sup>

Since the beginning of Minimalism there have been various ways of implementing the generalized EPP effect: in early Minimalism this was done via strength, and in Chomsky (2000, 2001) via the EPP diacritic, which indicates certain heads need Specifiers. In the above approach, generalized EPP effects follow from the  $uK$  of the moving element, which is independently needed even in Chomsky's EPP system. The interesting twist of the analysis is that the effect is stated as a property of the moving element, not the target.

It is also worth noting the restrictiveness of the above system. Thus, marking the K feature in (9) uninterpretable on Y will always lead to movement of Y to XP, i.e. it will result in Move, while marking it uninterpretable only on X will always lead to Agree.

(9) X...Y  
K K

This, however, brings us to a difference between the Chomsky/Bošković systems. Under both approaches a probe X, which initiates an Agree operation, must have a  $uK$ . Without a  $uK$ , there would be no reason for X to probe. Since there is no need for it, X cannot probe, given LR. In Chomsky's system, (10) then invariably crashes, since  $uK$  of Y cannot get checked.

(10) X .....Y  
iK uK

This is not the case in Bošković's system, where Y would move to SpecXP and probe X from there, checking  $uK$ . (10) thus yields different results in Chomsky's and Bošković's systems. Below, I discuss one case of this type, which favors Bošković's system.

However, the most important difference between Chomsky (2000) and Bošković (2007) for our purposes concerns the driving force of movement: while for Chomsky movement is target-driven, for Bošković it is moving-element driven. We have already seen one argument for the latter: it concerns the case where X must start moving before its target enters the structure. The target cannot drive movement in such cases for a very simple reason: it is not there.

Multiple wh-fronting (MWF), an example of multiple movement to the same position, provides us with another relevant test.

---

<sup>6</sup> See Epstein and Seely (2006) for discussion of the traditional EPP in this context, which is generalized in Bošković (2007), with an exploration of a number of additional consequences and an extension to successive cyclic and wh-movement. (Bošković's analysis of the traditional EPP is actually quite different from Epstein and Seely's. The latter crucially appeal to the Inverse Case Filter, which is dispensable in the former.)

- (11) a. \*Koj vižda kogo?  
           who watches whom  
       b. Koj kogo vižda? (Bulgarian)

Bošković (1999) discusses how MWF constructions can be handled within a target-driven (TD) and a moving-element driven system (MD). In TD, we need to adopt the Attract 1-F/Attract all-F distinction, where Attract 1-F heads attract only one element bearing feature F, while Attract all-F heads attract all elements bearing feature F. Interrogative C in English is an Attract 1-F head, and in Bulgarian an Attract all-F head, attracting all wh-phrases. In MD, all wh-phrases in Bulgarian are obligatorily specified with the *uK* feature that drives wh-fronting.<sup>7</sup> MWF is then implemented as follows within the two systems.

- (12) a. TD : Attract all-F C head  
       b. MD: each wh-phrase has *uK*

Suppose we try to implement optional MWF. In TD, the C head would optionally have the relevant Attract all-F property, while in MD, wh-phrases would optionally have the relevant *uK*. This provides us with a very interesting tool to tease apart the two systems. Consider, e.g., a sentence with four wh-phrases. In MD, we could give a *uK* to only two wh-phrases, which would result in two wh-phrases undergoing fronting, and two remaining in situ. I will refer to this pattern as partial MWF. The pattern is impossible in TD: the C head either has the Attract all-F property, in which case all wh-phrases front, or it doesn't, in which case they all stay in situ. Crucially, partial MWF is impossible in this system.<sup>8</sup> The question is then, whether there are languages that allow partial MWF. Surányi (2006) claims Hungarian is exactly such a language.

- (13) (Mondd el) mikor ki tévesztett össze kit kivel.  
       tell-imp prt when who-nom confused-3sg prt who-acc who-with (Balász Surányi, p.c.)  
       ‘(Tell me) who confused who with who when.’

Partial MWF thus provides an argument for MD.

Another argument is provided by QR. QR is somewhat controversial, but if it does exist it provides a strong argument for MD. Suppose QR involves IP adjunction. QR must be driven by the moving element. Clearly, there is nothing about I that would require adjunction of a quantifier. On the other hand, under the QR analysis quantifiers are supposed to be uninterpretable in situ. It must then be that an inadequacy of the quantifier, i.e. the moving element, rather than the target, i.e. I, drives QR.

There is a bit of a complication in the technical implementation of QR though. The standard assumption that quantifiers are uninterpretable in situ can be interpreted as indicating they have a *uK* which makes them uninterpretable in situ, requiring movement. The movement can either eliminate the *uK*, in a way releasing quantifiers for interpretation,<sup>9</sup> or we could

<sup>7</sup> See Bošković (2002b) regarding the nature of this feature, which is not important for current purposes.

<sup>8</sup> As discussed in Bošković (1999), independently of the Attract all-F property that is responsible for MWF, in some MWF languages (e.g. Bulgarian) the C also has an Attract 1-F property that requires one wh-phrase to move. If the C here could have this Attract 1-F property it would be possible to force one wh-phrase move, but not two.

<sup>9</sup> This could be done through valuation if the *uK* feature is lexically unvalued, given the discussion in section 2.3, where it is argued that only valued *uK*s can be deleted.

complicate the feature checking system by assuming that as a result of movement, the uninterpretable feature becomes interpretable (i.e. it is interpretable only in certain positions).

Another argument for MD is provided by Fitzgibbons (2007), who discusses negative concord in Russian, where she argues all negative concord items (NCIs) must move to the negative head. What could be driving this movement? In TD, it would be a property of negation, and in MD a property of the moving elements, i.e. NCIs. Fitzgibbons argues that it cannot be the former, since negation does not require NCIs; after all negation can occur without NCIs. Therefore, it must be a property of the NCIs themselves; in fact, in contrast to negation, which can occur without NCIs, the NCIs cannot occur without negation. The NCI movement then must be driven by the moving elements. This provides us with another argument for MD.

## 2.2 Freezing effects and Last Resort

I now turn to a discussion of freezing effects, where LR considerations are crucially involved. Consider again the issue of what drives successive cyclic movement. Since in Chomsky's (2000) system movement is driven by the EPP property, and Agree is a pre-requisite for movement, successive cyclic movement always involves feature checking. This means *what* must undergo feature checking with *that* in (4). On the other hand, in Bošković's (2007) system the reason why *what* in (4) moves to the embedded SpecCP is to avoid being sent to spell-out when the embedded IP, a phase complement, is sent to spell-out. This would freeze *what* for further movement, as a result of which the *uK* driving wh-movement would remain unchecked, given that a *uK* can only be checked if it serves as a probe. Although under this analysis successive cyclic movement is still in a sense feature checking driven since without it the *uK* of the moving element would remain unchecked, there is no feature checking in intermediate positions—*what* and *that* do not undergo feature checking.<sup>10</sup>

Bošković (2002a, 2007) and Boeckx (2003) provide a number of arguments that there is indeed no feature checking under successive cyclic (A and A') movement. I summarize here one argument from Bošković (2002a).

Lobeck (1990) and Saito and Murasugi (1990) note functional heads can license ellipsis of their complement only when they undergo Spec-Head agreement (SHA), i.e. feature checking. (14) shows tensed I, 's, and +*wh*-C, which undergo SHA, license ellipsis, while the non-agreeing

---

<sup>10</sup>In this respect, Bošković (2007) represents a return to early minimalism, where successive cyclic movement was not a result of feature checking. Rather, it was a consequence of the Minimal Link Condition (MLC) (Chomsky and Lasnik 1993 and Takahashi 1994; revived in Bošković 2002a; Boeckx, 2003; Chomsky, in press.) The MLC forces X undergoing movement of type Y to stop at every position of type Y on the way to its final landing site independently of feature checking. *What* in (1) then must pass through the embedded SpecCP on its way to the matrix SpecCP.

This analysis crucially relies on the Form Chain operation, where all relevant syntactic conditions, including LR and the Cycle, are stated with respect to the formation of chains, not chain links. Under this analysis, *what* in (1) starts moving only after *wh*-C enters the structure. The MLC forces formation of intermediate chain links. LR is satisfied since the formation of the whole chain, whose head is located in matrix SpecCP, has a feature-checking motivation. Since the whole chain extends the tree, the Cycle is also satisfied.

Chomsky (1995) eliminates Form Chain, which has led to the abandonment of this analysis of successive cyclic movement because with the elimination of Form Chain, formation of each chain link must satisfy LR and the Cycle. This means *what* in (1) must move to the Spec of *that* before higher structure is built, and the movement must have independent motivation. As discussed above, Bošković (2007) and Chomsky (2000) do not differ regarding the former, but do differ regarding how the latter requirement is satisfied, due to fundamentally different treatments of the issue where the formal inadequacy driving movement is located.

heads *the* and *that* do not.

- (14) a. John left and [<sub>IP</sub> Peter<sub>i</sub> [<sub>T</sub> did ~~*t<sub>i</sub>*~~ *leave*]] too.  
b. John's talk was interesting but [<sub>DP</sub> Bill [<sub>D'</sub> ~~*s talk*~~]] was boring  
c. \*A single student came because [<sub>DP</sub> [<sub>D'</sub> the ~~*student*~~]] thought it was important.  
d. John met someone but I don't know [<sub>CP</sub> who<sub>i</sub> [<sub>C'</sub> C ~~*John met t<sub>i</sub>*~~]].  
e. \*John believes that Peter met someone but I don't think [<sub>CP</sub> [<sub>C'</sub> that ~~*Peter met someone*~~]].

Significantly, intermediate C cannot license ellipsis of its IP complement.

- (15) \*John met someone but I don't know who<sub>i</sub> Peter said [<sub>CP</sub> t<sub>i</sub> [<sub>C'</sub> that ~~*John met t<sub>i</sub>*~~]].

This can be easily accounted for if passing through an intermediate SpecCP does not imply feature checking (SHA) with the C. (15) then provides evidence against the feature-checking view of successive cyclic movement, where *that* would undergo SHA in (15), just as in (14d) and in contrast to (14e).<sup>11</sup>

What about languages with overt reflexes of agreement with intermediate heads under wh-movement? As noted in Boeckx (2004), it is not clear there are languages with true intermediate wh-agreement. In many languages of this type, wh-agreement is only indirect: instead of a wh-phrase directly agreeing with an intermediate head, wh-movement induces special agreement between intermediate verbs and intermediate complementizers. I refer the reader to Bošković (2008a) for an analysis of this pattern that does not involve intermediate feature checking,<sup>12</sup> and turn to a much better candidate for such successive cyclic movement, namely Kinande. In Kinande the morphology of the C covaries with the morphology of the wh-phrase.

- (16) a. IyondI y0/ ABahI Bo Kambale alangIra  
who.1 that.1 who.2 that.2 Kambale saw  
'Who did Kambale see?'  
b. EkIhI ky0/ EBhI By0 Kambale alangIra  
what.7 that.7 what.8 that.8 Kambale saw  
'What did Kambale see?' (Rizzi 1990)

The agreement occurs with displaced wh/focus phrases and can be found in every clause on the path of movement.<sup>13</sup>

- (17) [**ekihi kyo** Kambale a.si [nga.kyo Yosefu a.kalengekanaya  
what wh-agr(eement) Kambale agr.know C.wh-agr Joseph agr.thinks  
[nga.kyo Mary' a.kahuka \_\_ ]]]  
C.wh-agr Mary agr.cooks  
'What did Kambale know that Joseph thinks that Mary is cooking?'

<sup>11</sup>The argument extends to Chomsky's (2000) system, where the SHA requirement would be restated as an EPP requirement.

<sup>12</sup>As discussed in Bošković (2008a), the analysis may be extendable to Irish.

<sup>13</sup>*Nga* occurs in the embedded clause because monosyllabic Cs are second position clitics.



(Schneider-Zioga 2005)

However, Boeckx (2004) suggests an iterative prolepsis account of Kinande long-distance wh-dependencies, analyzing traditional long-distance wh-movement from (18a) as in (18b), where the apparent argument of the lower V is generated as a matrix clause dependent that undergoes local wh-movement, binding a null element that also undergoes local wh-movement. Instead of a single, successive cyclic wh-movement, where the wh-phrase agrees with two Cs, we then have two local wh-movements, with different elements agreeing with the two Cs. In each case the C agrees with the head of a chain; (18b) doesn't involve true intermediate feature-checking.

- (18) a.  $[_{CP} Op_i [_{CP} t_i [t_i$   
b.  $[_{CP} Op_i t_i [_{CP} Op_i [ t_i$

Schneider-Zioga (2005) conclusively shows Kinande does not have true long-distance A' movement. Consider (19).

- (19) a. **ekitabu kiwe**<sub>j/k</sub> ky' **obuli mukolo**<sub>j</sub> a.kasoma \_ kangikangi.  
book his wh-agr each student agr.reads regularly  
'(It is) His<sub>j</sub> book that [every student<sub>j/k</sub>] reads regularly.'  
b. **ekitabu kiwe**<sub>k/\*j</sub> kyo ngalengekanaya  $[_{CP}$  nga.kyo [**obuli mukolo**]<sub>j</sub> akasoma \_  
book his wh-agr I.think C.wh-agr every student read  
kangikangi.  
regularly  
'(It is) His<sub>k/\*j</sub> book that I think [every student]<sub>j</sub> reads regularly.'  
c. **ekitabu kiwe**<sub>k/\*j</sub> kyo [**obuli mukolo**]<sub>j</sub> alengekanaya  $[_{CP}$  nga.kyo nganasoma \_  
book his wh-agr every student agr.think C.wh-agr I.read  
kangikangi]  
regularly  
'(It is) His<sub>k/\*j</sub> book that [every student]<sub>j</sub> thinks I read regularly.'

(19a) shows local A'-extraction allows reconstructed interpretation. However, reconstruction is impossible with a long-distance dependency. Under the standard view of reconstruction that ties reconstruction to movement, we are led to conclude the focused element undergoes movement from its  $\theta$ -position to SpecCP in (19a), but not (19b-c). (19b-c) then indicate Kinande doesn't have true long-distance A'-movement.

Consider also Schneider-Zioga's (20)-(21).

- (20) \*omukali ndi yo wasiga [<sub>island</sub> embere \_ wabuga]  
woman who wh-agr you.left before spoke  
'Which woman did you leave before (she) spoke?'  
(21) omukali ndi yo wasiga [<sub>island</sub> embere Kambale anasi  $[_{CP}$  ko.yo \_ wabuga]]  
woman who wh-agr you.left before Kambale knew C.wh-agr spoke  
'Which woman did you leave before Kambale knew that (she) spoke?'

(20) is unacceptable due to extraction from an adjunct. Significantly, (21), where the extraction site is embedded within an agreeing complementizer clause, is acceptable. This conclusively

shows the wh-phrase in (21) does not undergo wh-movement to the matrix clause from the gap site.

I therefore conclude Kinande agreeing long-distance A'-movement constructions do not involve a wh/focus phrase moving clause-to-clause, with a single wh/focus phrase undergoing agreement with more than one C.<sup>14</sup>

We have seen a wh-phrase undergoing successive cyclic movement does not undergo feature checking with intermediate heads. Kinande shows intermediate Cs actually can undergo agreement. What is, however, not possible is that after undergoing agreement with an intermediate C, a wh-phrase moves and establishes an agreement relation with another C. Those intermediate Cs in Kinande are really final Cs, since once a wh-phrase moves to SpecCP undergoing agreement with the C it is frozen in this position. The most straightforward way of interpreting this is that feature checking for the *uK* involved in wh-movement is possible only once; once a wh-phrase undergoes agreement for this feature it is frozen. A natural step to take is to generalize this to every feature, which in turn provides strong evidence against Chomsky's (2000) view of successive cyclic movement. Consider again Chomsky's system, where Y must have a *uK* to be visible for movement. X and Y in (22) undergo F feature checking, and as a reflex of this, the *uK* of Y is checked, which happens after movement of Y to SpecXP.

(22)	[ <sub>XP</sub> Y <sub>i</sub> (goal)	X(probe)	t <sub>i</sub>
	<i>iF</i>	<del><i>uF</i></del>	
	<del><i>uK</i></del>	<del>EPP</del>	

Successive cyclic movement has forced Chomsky to complicate this system by adopting the concept of defective heads, which are defective in that they are unable to check off the feature of the goal that has made the goal visible for movement to the head in question.

As an illustration, consider wh-movement: for Chomsky, the embedded C in both (23) and (24) undergoes feature checking with *what*. The difference is that the embedded C in (23) isn't, and the embedded C in (24) is, a defective head. Consequently, only the C in (23) checks off the *uK* of *what*, freezing it for further wh-movement. Since the embedded C in (24) is defective, it doesn't check the *uK* of *what*, which can then move to another SpecCP.<sup>15</sup>

(23) I wonder	what <sub>i</sub>	C	Mary bought t <sub>i</sub> .
	<i>iF</i>	<del><i>uF</i></del>	
	<del><i>uK</i></del>	<del>EPP</del>	

(24) a.	What <sub>i</sub> do you think	[ <sub>CP</sub> t <sub>i</sub> that	Mary bought t <sub>i</sub> ]
	b. You think	[ <sub>CP</sub> what <sub>i</sub> that	Mary bought t <sub>i</sub> ]
		<i>iF</i>	<del><i>uF</i></del>

<sup>14</sup> A modification of Boeckx's analysis is necessary to account for (19c): the focused NP should not even undergo local wh-movement, or the reconstruction would be possible. The contrast in (19a,c) indicates local A'-movement is possible only from the  $\theta$ -position, i.e. we are not dealing here with a proleptic object undergoing A'-movement. (17) then has a structure like (i), where only the lowest null element undergoes movement.

(i) [<sub>CP</sub> Op<sub>i</sub> [<sub>CP</sub> Op<sub>i</sub> [<sub>CP</sub> Op<sub>i</sub> t<sub>i</sub>

<sup>15</sup>For actual features involved in feature checking under wh-movement, which are not important for our purposes, see Bošković (in press a) (*C/that* may also have the *K* feature).

As noted in Bošković (2008a), under non-feature checking approaches to successive cyclic movement, it is not necessary to stipulate the defectiveness of intermediate heads with respect to feature checking since such heads are not involved in feature checking in the first place. In other words, if there is no feature checking with intermediate heads, we do not need to assume some heads are defective regarding how they participate in feature checking. We can then make the process of feature checking completely uniform in that all feature checking inactivates the moving element, deleting the *uK* that has made it active for movement. This immediately captures the freezing effect of agreement with C. If a wh-phrase moves to SpecCP and undergoes agreement even with an intermediate C like *that*, the *uK* that makes it active for wh-movement will be erased, freezing it in SpecCP. There are then two options to get legitimate long-distance wh-structures: (a) reanalyzing long-distance wh-dependencies as a series of local wh-dependencies, where the declarative C undergoes agreement with X in its Spec, freezing it in place; (b) a wh-phrase moves to the Spec of *that* without undergoing agreement with *that*, the movement being driven by the considerations from section 1. Kinande takes option (a) and English (b).

The above discussion should be generalized. It is not only that wh-movement (i.e. feature-checking movement to SpecCP) cannot feed another wh-movement. As shown in Bošković (in press a) and references therein, no instance of A'-movement can feed another instance of A'-movement. Thus, Lasnik and Uriagereka (1988) observe that although it is standardly assumed QR is clause bounded, many speakers allow *every problem* to have wide scope in (25a). Significantly, even for them *every problem* cannot have wide scope in (25b).

- (25) a. Someone thinks that Mary solved every problem.  
 b. Someone thinks that every problem, Mary solved.

Assuming *every problem* scopes over *someone* in (25a) as a result of QR into the matrix clause, (25b) indicates topicalization cannot feed QR.

Grohmann (2003) notes wh-movement cannot feed topicalization based on (26), where *who* undergoes topicalization after wh-movement to SpecCP, with the comma intonation indicating a pause associated with topicalization.<sup>16</sup>

- (26) \*Who, does Mary detest?

Bošković (in press a) shows focus movement cannot feed wh-movement based on MWF. I simply refer the reader to Bošković (in press a) and references therein for additional arguments that A'-movements like wh-movement, focus movement, topicalization, QR and the NPI movement cannot feed each other. Why is this the case? Bošković (in press a) argues there is a general, operator (Op)-type feature that is shared by elements undergoing A'-movements. It is the Op-feature that makes a phrase visible for an operator-style (A'-) movement (a topic/focus/wh-phrase then has *iTop/iFoc/iWH* and *uOp*). Given that there are no defective heads, once a phrase undergoes feature-checking A'-movement, its Op-feature is deleted, as a result of which the phrase cannot undergo another A'-movement.

Boskovic (2008a) (see also Rizzi 2006) argues the freezing effect is even more general.

<sup>16</sup> To Peter, *what should Mary give* indicates the landing site of topicalization precedes SpecCP in matrix clauses.

Above, we have come close to saying no instance of feature-checking movement can feed another instance of feature-checking movement. The only feeding relation still allowed involves feature checking A-movement feeding feature checking A'-movement. There is evidence that even this is disallowed, which gives us (27) within Bošković's (2007) system (more theory neutral, X undergoes feature-checking movement only once).<sup>17</sup>

(27) X probes only once (i.e. X undergoes feature checking as a probe only once).

Consider Q-float under wh-movement in West Ulster English (WUE).

(28) a. Who<sub>i</sub> was arrested all t<sub>i</sub> in Duke Street?

b. \*They<sub>i</sub> were arrested all t<sub>i</sub> last night. (McCloskey 2000)

WUE allows (28a) but disallows (28b). McCloskey argues *who* in (28a) must move to SpecCP without moving to SpecIP, the reasoning being that if *who* were to move to SpecIP in (28a), it would be impossible to account for the contrast in (28). Whatever rules out movement to SpecIP in (28b) should also rule it out in (28a). (28) shows that what is standardly assumed to happen in subject wh-questions, the subject moves to SpecIP and then to SpecCP, actually does not happen: the subject moves directly to SpecCP. Notice now that on the 'standard' derivation (28) would involve feature-checking A-movement feeding feature-checking A'-movement, violating (27) (see also (30)).<sup>18</sup>

There are two questions to answer now. How is the requirement that SpecIP be filled in English satisfied in (28a), given that *who* never moves to SpecIP. Does (27) follow from anything?

Consider the first question within Bošković's (2007) system. Recall Bošković (2007) dispenses with the EPP: EPP effects follow from case considerations. *John* in (29) has *uCase*.

(29) [<sub>VP</sub> John left]

Since *uK* must be a probe, *John* must move to a position c-commanding the case-checker (I). Given Shortest Move, *John* moves to the closest position c-commanding I, SpecIP.<sup>19</sup> In

---

<sup>17</sup>A similar claim is made in Rizzi (2006). However, his treatment of the claim is quite different from the one developed below. For another approach that bans A-A' feeding in a system quite different from the one adopted here, see Chomsky's (in press), who argues for the existence of parallel movement. (For ways of teasing apart the analyses in question, see Bošković 2008a, in press b. The Kinande case in (30) actually cannot be captured within Chomsky's system. Moreover, the system does not ban A'-A' feeding.)

<sup>18</sup>In Bošković's (2007) system, *who* would be a probe in both cases, probing C and I from SpecCP/SpecIP respectively.

<sup>19</sup>One argument for this system concerns (i).

(i) \*I know what John conjectured.

*Conjecture* is not a case assigner (cf. \**John conjectured it*). A question, however, arises why *know* cannot case-license *what*, given that *know* has the ability to do that. (i) presents a serious problem for Chomsky's (2000) system. It appears nothing prevents establishment of a probe-goal relation between the matrix *v* and *what*, which should case-license *what*. From Bošković's (2007) perspective, (i) is straightforward: the derivation in question is blocked because *what* with its *uCase* must function as a probe. The only way this can be accomplished is if *what* moves to

principle, *John* could move to SpecCP instead of SpecIP. Both movements would result in the checking of all relevant features. However, movement to SpecIP is preferred by Shortest Move, which favors the shortest movement possible. Consider now (28a). If *who* moves to SpecIP, its *uCase* will be checked. However, given (27), its *uOp*-feature will never get checked since *who* will be prevented from moving to a position c-commanding C. This derivation therefore crashes. The derivation is then irrelevant for economy comparison, including Shortest Move, which compares only possible derivations. Notice now that if, instead of SpecIP, *who* moves to SpecCP, from this position *who* can probe both C and I, checking both its Case and Op-feature. Movement to SpecCP is then the only possibility.

Consider now Kinande, where canonical subject/object agreement are impossible when the subject/object undergo wh-movement (Schneider-Zioga 1995). This can be straightforwardly captured in the above system if we make the natural assumption that canonical subject/object agreement are triggered in Kinande when the subject/object probe I/v from SpecIP/SpecvP respectively.

- (30) a. [<sub>IP</sub> Subject I-agreement]]  
 b. [<sub>vP</sub> Object v-agreement]]  
 c. [<sub>CP</sub> Wh-Subject [<sub>IP</sub> I-(\*agreement)]]  
 d. [<sub>CP</sub> Wh-Object [<sub>IP</sub> [<sub>vP</sub> v-(\*agreement)]]]

Since under the current analysis subject and object undergoing wh-movement probe both C and I/v from SpecCP, it follows canonical agreement cannot co-occur with wh-movement.<sup>20</sup>

I now turn to a deduction of (27). Consider the line of reasoning employed in the discussion of freezing effects above. Suppose X must have a *uK* to make it active for movement Y. Once X undergoes feature-checking movement to a Y, the *uK* will get checked off so that X cannot undergo another Y-movement. As discussed above, the freezing effect can be generalized to all A' feature checking by generalizing the *uK* that is involved in A'-movement checking. If the same feature of the moving element is checked under all instances of A'-movement, once X undergoes feature-checking A'-movement, the relevant feature will get checked off, freezing X for further A'-movement. (27) then suggests a further generalization: It is the same feature of the moving element that is checked in all instances of movement, A or A'. As noted in Bošković

---

the matrix SpecvP. However, this derivation is blocked because *who* is located outside of its scope (embedded CP), which is disallowed (see Saito 1992). (i) thus provides evidence that case cannot be licensed in situ without movement to the case licenser (see Bošković 2007 for discussion of cases where case movement was previously assumed not to occur).

<sup>20</sup>The object would pass through SpecvP in (30d) due to considerations from section 1, but it would not probe v from there because of (27).

The above analysis has many consequences, discussed in Bošković (2007). To mention one here, it captures the behavior of verbs like *wager*, which ECM wh-traces, but not lexical NPs (Postal 1974).

- (i) a. \*John wagered Mary to be smart.  
 b. Who did John wager to be smart?

Assuming overt object shift in English, Bošković (1997) argues that due to the presence of an additional VP shell with a filled Spec that occurs with this class of verbs, *Mary* cannot reach matrix SpecvP in (ia) without a locality violation. (ib) is then straightforward in the above system, where *who* probes *wager* from the matrix SpecCP. Since there is no A feature-checking movement to the matrix SpecvP in (ib), the locality problem that arose in (ia) does not arise in (ib).

(2008a), this means that once X undergoes any feature-checking movement it will no longer be able to undergo another feature-checking movement. This requires changing the way we have been treating movement. We can no longer consider the specific features like *uCase* or *uOp* to be the driving force of movement since the driving force needs to be generalized. What we need is a general property X which can be given to any element when it enters the structure. This general property is tied to probing: it indicates a need to function as a probe and is satisfied under successful probing.<sup>21</sup> An element A marked with X (which cannot probe in situ) would move to the edge of a phase to attempt a probing operation: if A successfully undergoes probing, X is deleted, freezing A in place. If A fails to probe due to the lack of a goal (so it still has X), it moves to the higher phase Spec to attempt probing again. The X property is then used to drive successive cyclic movement (*instead of uK*, as in Bošković 2007). Another way of looking at this is as follows: Suppose X is PF uninterpretable (after all, the property ‘I need to function as a probe’ is not a PF-related property). This means sending an element with the X property to Spell-out would cause a PF crash. Assuming what is sent to Spell-out is the complement of a phase head, A in (31) will have to move to the Spec of the phase head B to avoid being sent to Spell-out, which would cause a crash.

(31) W [BP B A  
           K uK  
               X

Successive cyclic movement works as before, without feature checking with intermediate heads. When A moves to SpecWP it successfully probes W, checking *uK* and deleting X, which is tied to feature checking under probing.

(32) A W [BP  
       ~~uK~~ K  
       ~~X~~

The result of this system is that A can move to probe only once. Once A undergoes feature-checking movement, X is deleted, freezing A in place. (27) is then deduced.<sup>22</sup>

Rodríguez-Mondoñedo (2007) observes a rather interesting prediction of this system. Consider (33).

(33) X Y Z  
       K F u.K  
       F u.F

<sup>21</sup>Giving X to an element Y without uninterpretable features, which then wouldn’t function as a probe, would lead to a crash. But there is always the derivation on which Y does not get X.

<sup>22</sup> Natasha Fitzgibbons suggests an alternative deduction of (27). She suggests maximizing feature checking under probing to the effect that if X probes, X must check all its *uK*s. The A-A’ feeding relations, where X would first move to probe for *uCase* and then move to SpecCP to probe for *uOp*, are also ruled out under this approach, which means (27) is deduced. The deduction is consistent with the derivations discussed above. Thus, *who* in (27) still moves directly to SpecCP, probing for both the *uCase* and the *uOp* feature from there.

Z in (33) has two uninterpretable features, which need to be checked, more precisely, receive a value.<sup>23</sup> The closest valuator for F is Y. In the absence of freezing effects, we would expect Y to value the F feature of Z, which in Bošković's (2007) system would happen after Z moves to SpecYP. However, given the freezing effect, probing for F from SpecYP would freeze Z, leaving its K feature unvalued. In the above system, we would therefore expect Z to move to SpecXP, probing for all its features from there. Locality then requires that X rather than Y values the F feature of Z. Rodríguez-Mondoñedo argues that this rather interesting prediction of the freezing system, where the closest valuator (Y) unexpectedly fails to do the job, is borne out based on some previously unexplained instances of obligatory *a*-object marking with Spanish inanimates.

### 2.3. Last Resort and Agree

I now turn to the effects of LR for Agree, focusing on the claim that X can only be a probe if it has a *u*K.

Chomsky (2001) argues that in addition to the interpretable/uninterpretable distinction, we need a valued/unvalued distinction, where some features are fully valued lexically, while others receive their value during the derivation. Consider Serbo-Croatian (SC) (34) (*kola* is a pluralia tantum).

- |      |    |                             |           |                |
|------|----|-----------------------------|-----------|----------------|
| (34) | a. | Zelena                      | kola      | su kupljena.   |
|      |    | green.fem                   | car.fem   | are bought.fem |
|      |    | 'The green car was bought.' |           |                |
|      | b. | Zeleno                      | auto      | je kupljeno.   |
|      |    | green.neut                  | car.neut  | is bought.neut |
|      | c. | Zeleni                      | automobil | je kupljen.    |
|      |    | green.masc                  | car.masc  | is bought.masc |

The gender of the adjective and the participle depends on the gender of the noun. *Green* can be feminine, neuter, or masculine; which gender it has depends on the noun it modifies. As noted by Pesetsky and Torrego (2007) (PT), the dependence of the gender specification of adjectives and participles on the syntactic context in which they occur can be easily captured if they are lexically unvalued for gender: they receive their gender value after undergoing agreement with a noun that already has a valued gender specification. In contrast to the adjective/participle in (34), nouns like *kola*, *auto*, and *automobil* have a fixed gender specification: *kola* is always feminine, *auto* neuter, and *automobil* masculine. The most straightforward way of capturing this is to assume that nominal gender is lexically valued; in contrast to adjectives and participles, nouns do not receive their gender value during syntactic derivation, hence their gender value does not depend on their syntactic context.<sup>24</sup>

Since SC gender is quite clearly grammatical (it depends on the declension class a noun belongs to), we also have here evidence for the existence of valued uninterpretable features, a

<sup>23</sup> In the current system, checking is interpreted in terms of valuation of unvalued features, see section 2.3. The technical implementation of checking has not been important until now.

<sup>24</sup> Recall that *kola* in (34a) is a pluralia tantum, i.e. its number is plural although it is interpreted as singular. This kind of lexical quirks also call for full lexical specification of  $\phi$ -features of nouns. As pointed out by PT, there are no pluralia tantum verbs or adjectives, which is not surprising if the  $\phi$ -features of these elements are lexically unvalued: such treatment does not leave room for lexical quirks like the one exhibited by the number of the noun in (34a).

possibility that is disallowed in Chomsky's (2000, 2001) system essentially by a stipulation. Allowing for the existence of valued uninterpretable features also allows us to simplify the feature-checking process. Since in Chomsky's system uninterpretable features are always unvalued, the system does not allow feature checking between two uninterpretable features. Feature checking is supposed to result in valuation of unvalued features. If both the probe's and the goal's feature are unvalued, their feature checking cannot result in valuation. Disallowing the possibility of checking two uninterpretable features against one another forces Chomsky quite generally to tie checking of an uninterpretable feature *F* of a goal to checking of a different uninterpretable feature *K* of its probe (note that interpretable features, which are always valued for Chomsky, cannot serve as probes due to LR; since there is no need for them to initiate probing they are not allowed to do it), which makes feature checking rather cumbersome and leads to a proliferation of features involved in checking. Thus, since (35a-b) cannot result in the checking of the *K* feature of *Y* ((35a) because, being unvalued, the *uK* of *X* cannot value the *uK* of *Y*, and (35b) because *X* cannot function as a probe due to the lack of uninterpretable features), Chomsky is forced to posit (35c), where the *uK* of *Y* is checked as a reflex of *F* feature checking. This kind of reflex checking considerably complicates the feature-checking mechanism and leads to a proliferation of features involved in checking (we cannot simply have *K*-feature checking in (35); rather, we need to assume an additional feature *F* is involved in feature checking between *X* and *Y*).

- (35) a. X Y  
       *uK* *uK*  
       b. X Y  
           *iK* *uK*  
       c. X Y  
           *uF* *iF*  
               *uK*

Allowing valued uninterpretable features enables us to simplify the feature checking relations from (35c). In particular, (35a) is now allowed, if one of the *K* features is valued.<sup>25</sup>

Given this much background, let us reconsider the question of what drives Agree. It is standardly assumed semantics cannot deal with uninterpretable features, hence such features need to be eliminated before entering semantics. The elimination takes place through feature checking. A question, however, arises why such features simply could not be deleted, in which case they would not need to be checked. It is argued in Bošković (2008c) that such features indeed can be deleted without checking, but only if they are valued (see also Chomsky 2001). In other words, valuation is a prerequisite for deletion of uninterpretable features. But if a valued *uK* can simply be deleted, there is no need for it to undergo feature checking (see below for evidence to this effect). Then, given LR, a valued *uK* cannot function as a probe. On other hand, an unvalued *uK* can function as a probe, since such elements do induce a crash, hence there is a need for them to undergo Agree. PT argue that just like uninterpretable features can be either valued or unvalued, as we have seen above, interpretable features can also be either valued or unvalued. As an example of unvalued interpretable features they give the Tense feature of the Tense node, which for them is the locus of semantic tense interpretation but its value depends on

<sup>25</sup> See below and Bošković's (2008c) analysis of SC gender, where the gender feature of both the gender probing head, which is responsible for participial gender, and the noun is uninterpretable, but unvalued only on the former.



its syntactic context, i.e. the verb it co-occurs with. They also implement clausal typing in terms of an unvalued interpretable feature of C. It seems natural to assume that an unvalued *iK* would still be a problem for semantics; i.e. semantics would know what to do with an *iK* only if K has a value (see also PT). Unvalued *iK*s can then also function as probes. From this perspective what drives Agree is valuation: only unvalued features can function as probes. (36) then shows which contexts can yield a legitimate Agree relation, where X is a probe and Y its goal.<sup>26</sup>

- (36) a. X[*uval/uK*]...Y[*val/uK*]  
 b. X[*unval/iK*]...Y[*val/iK*]  
 c. \*X[*val/uK*]...Y[*val/uK*]  
 d. \*X[*val/iK*]...Y[*val/iK*]  
 e. \*X[*unval/uK*]...Y[*unval/uK*]  
 f. \*X[*unval/iK*]...Y[*unval/iK*]  
 g. \*X[*val/uK*]...Y[*unval/uK*]  
 h. \*X[*val/iK*]...Y[*unval/iK*]

Agree cannot take place between X and Y in (36c-d) due to LR (there is no reason for X to probe). The same holds for (36g-h) in Chomsky's system (2001), though the structures would be treated differently in Bošković's (2007) system, as discussed below. Finally, the problem with (36e-f) is that the unvalued features of X and Y cannot be valued. An innovation of this system is that it allows interpretable features to trigger feature checking (see also PT), which was not possible in Chomsky (1995), where uninterpretable features were the trigger for feature checking.<sup>27</sup> Also, in contrast to Chomsky (2000), two uninterpretable features can undergo feature checking, as long as the probe is unvalued and the goal valued. However, even uninterpretable features fail to trigger Agree if they are valued.

Another important property of the valuation-driven system is that valued uninterpretable features do not need to be checked, given that they can be deleted. This is a departure from Chomsky (1995), where all uninterpretable features have to undergo checking. (On the other hand, while in Chomsky 1995 interpretable features do not need to undergo checking, in the above system interpretable features do need to undergo checking if they are unvalued.) There is strong evidence that valued uninterpretable features indeed do not need to undergo checking. Consider first conjunct gender agreement in SC.

- (37) Uništena su sva sela i sve varošice.  
 destroyed.neut are all villages.neut and all towns.fem  
 'All villages and all towns were destroyed.'

The participle in (37) agrees in gender (i.e. undergoes feature checking for gender) with the first conjunct, which means the second conjunct is not involved in gender feature checking. Notice also that the conjunct does not have default gender, which is masculine in SC. Its non-default gender feature simply goes unchecked in (37). This is exactly what is expected given the above discussion: the gender feature of the noun is uninterpretable, but valued. As a result, it can be deleted (so that it does not enter semantics, where it would cause a Full Interpretation violation) without checking.

<sup>26</sup> *Val* indicates valued and *unval* unvalued features.

<sup>27</sup> I am putting aside here strength, which was used to drive overt movement.

Another relevant case concerns case checking. Case checking is rather cumbersome in Chomsky's (2000) system. Case is quite clearly uninterpretable on both the traditional case assigner (e.g. Tense) and the assignee (NP), i.e. on both the probe and the goal. Since, as discussed above, Chomsky disallows Agree between two uninterpretable features, he cannot have direct case feature checking between T and NP. Rather, Chomsky has to complicate the system by appealing to the notion of reflex feature checking, where case checking is tied to the checking of another feature. Thus, for Chomsky, phi-features of T in (38) probe the NP, and somehow as a reflex of this phi-feature checking the case feature of the NP gets checked. The "somehow" here is rather mysterious given that T does not even have a case feature for Chomsky.

(38) T            NP  
       *uPhi*        *iPhi*  
                   *uC*

The current system makes possible a much more natural approach to case licensing, where both Tense and the NP have a case feature, in line with the attempt to eliminate the undesirable concept of reflex feature checking. The case feature of both Tense and the NP is uninterpretable. Furthermore, since (finite) T always governs nominative, and the case of NPs depends on the syntactic context in which they occur, T's case is valued and NPs' case unvalued.

(39) T            NP  
       *val/uCase*    *unval/uCase*

Case licensing in (39) can proceed without any problems and without reflex feature checking, but crucially only in Bošković's (2007) system. In the target-driven system of Chomsky's (2000), even if the above assumptions regarding valuation are adopted so that valuation drives Agree, Agree would fail in (39) because T could not function as a probe due to LR. On the other hand, in Bošković's (2007) system, the NP would move to SpecTP and then probe T from this position. Since the NP has an unvalued case feature it can function as a probe.

The above account makes another prediction. Since the case feature of traditional case assigners is valued, which means it can be deleted even without checking, it does not have to undergo checking. This is in contrast to the case feature of NPs, which is unvalued, hence needs to be checked. This amounts to saying that the traditional Case Filter holds, but the Inverse Case Filter does not hold. There is strong empirical evidence that this is indeed correct. It is pretty clear that the Case Filter holds. As for the Inverse Case Filter, all attempts to enforce it (e.g. Bošković 2002a, Epstein and Seely 1999) have come up short against persistent empirical problems which pretty clearly indicate traditional case assigners do not have to check their case, which means the Inverse Case filter does not hold. E.g., the existence of verbs that assign case only optionally, as in (40), goes against the spirit of the Inverse Case Filter.

- (40) a. John laughed.  
       b. John laughed himself silly.  
       c. Mary is dressing (herself).  
       d. Peter is eating (apples).

Slavic genitive of quantification/negation also provides evidence against the Inverse Case Filter (see Franks 2002). In a number of Slavic languages verbs that assign structural accusative fail to assign it when their object is a higher numeral NP. (*Kola* in SC (41b), which must bear Genitive, receives its case from the numeral.) The same happens when a verb is negated, as illustrated by Polish (42b), where genitive of negation is obligatory. (There are similar arguments against obligatory assignment of nominative as well as some lexical cases; see Franks 2002).

- (41) a. On kupuje kola.  
       he buys car.acc  
       b. On kupuje pet kola.  
           he buys five cars.gen (SC)
- (42) a. Janek czytał książkę.  
       Janek read books.acc  
       b. Janek nie czytał książki.  
           Janek neg read books.gen (Polish)

I conclude, therefore, that the valuation-driven version of Bošković's (2007) system not only captures case licensing without additional assumptions that were required in Chomsky's (2000) system, but also accounts for the fact that the Case Filter, but not the Inverse Case Filter holds (i.e. only the former is enforced).<sup>28</sup>

### 3. Lexical insertion/pure Merge and Last Resort

I now turn to the question of whether lexical insertion, or more generally, pure Merge, should be subject to LR. Chomsky (1995) assumes no aspect of lexical insertion, including pure Merge, is subject to LR, the underlying assumption being that if cost is assigned to lexical insertion, the cheapest thing to do would always be nothing, which means no lexical insertion would ever take place, resulting in silence. On the other hand, Chomsky (2000) suggests pure Merge is subject to LR, and it is motivated by selectional requirements. The assumption leads to a considerable enrichment of the theory of selection, since all lexical insertion/pure Merge now has to be driven by selection. This is unfortunate, since selection was previously shown to be close to eliminable.<sup>29</sup> In Bošković (1997) I took the position that falls in between Chomsky's (1995) and (2000) positions: only pure Merge of functional elements is subject to LR.<sup>30</sup> As discussed below, the literature contains a number of appeals to economy-of-representation principles intended to ban unnecessary projections (see (44)). Interestingly, in actual practice they are all applied only to functional elements; they are used to ban only unnecessary functional structure. This "accident" can be made more principled by taking the position that only pure Merge of functional elements is subject to LR. The functional/lexical category distinction makes sense

<sup>28</sup> As for default case, which clearly does not need to be checked, the most appropriate way to handle it is to assume that default case involves valued case on the NP, which means it does not need to be checked. Since the value of default case is fixed for each language for all constructions (i.e. it does not depend on syntactic context), it is clear that it should be valued. Since valued uninterpretable features do not need to be checked, we then also capture the fact that default case does not need to be checked (nouns with default case occur in environments where there is no plausible case assigner).

<sup>29</sup> More precisely, it was shown to follow from the semantic properties of lexical items, which should not be driving syntactic computation; see Pesetsky (1982) and Bošković (1997).

<sup>30</sup> I am actually generalizing here the position I took regarding lexical insertion to pure Merge in general.

given that lexical elements determine what we want or choose to say, and functional elements merely help us build legitimate grammatical structures. Bošković (1997) appeals to the natural assumption that the latter (building legitimate grammatical structures), but not the former (what we want or choose to say), is subject to economy principles to justify subjecting only pure Merge of functional elements to LR. Functional elements are then inserted into the structure only to the extent that they are necessary to build legitimate structures. Another way to approach this issue would be to assume that only functional categories are selected, a natural consequence of which would be to require only pure Merge of functional elements to be motivated by selectional requirements. Bošković (2004a, 2008b) shows the assumption that only pure Merge of functional projections is subject to LR enables us to deduce a rather interesting generalization concerning scrambling.<sup>31</sup>

(43) Only languages without articles may allow scrambling

SC, Latin, Japanese, Korean, Turkish, Hindi, Chukchi, Chichewa, and Warlpiri all have scrambling and lack articles. Particularly interesting are Slavic and Romance. Bulgarian, e.g., has noticeably less freedom of word order than SC. Also, all modern Romance languages have articles and lack scrambling, while Latin lacked articles and had scrambling. I argued in Bošković (2008b) that article-less languages do not project DP; the traditional Noun Phrase in such languages is an NP. I also adopted Bošković and Takahashi's (1998) approach to scrambling (BT), on which scrambled elements are base-generated without feature checking in their surface position, and then undergo LF lowering to the position where receive case/theta role. The main goal of this approach was to make scrambling conform to LR. Scrambling is standardly treated as an optional overt movement operation that takes place for no reason at all, which should violate LR. Under BT's approach, the optional, LR violating overt movement is replaced by obligatory LF movement that conforms with LR. Now, given that the traditional NP is DP, a functional category, in nonscrambling languages, and NP in scrambling languages, inserting it into the structure must have independent motivation (i.e. involve feature checking/satisfaction of selectional properties) in nonscrambling languages, but not in scrambling languages. Since scrambling is pure Merge that does not involve feature checking/satisfaction of selectional requirements under BT's analysis, it is then possible only in NP languages.

#### **4. The ban on superfluous structure: Economy of Representation and Last Resort**

The above discussion of LR as it applies to pure Merge has bearing on Economy of Representation principles that ban superfluous structure, which can be re-stated in terms of LR if pure Merge is subject to LR.

A number of authors have proposed principles whose goal is to ban superfluous symbols from representations.<sup>32</sup>

---

<sup>31</sup>Scrambling here is taken to be the kind of movement referred to as scrambling in Japanese, not German, whose "scrambling" is a very different operation with very different semantic effects from Japanese scrambling. One of the defining properties of scrambling for the purpose of (43) is the existence of long-distance scrambling from finite clauses, which German lacks.

<sup>32</sup> For additional principles along these lines, see Grimshaw (1997), Speas (1994), Radford (1994).

- (44) a. The Minimal Structure Principle (MSP)  
 Provided that lexical requirements of relevant elements are satisfied, if two representations have the same lexical structure, and serve the same function, then the representation that has fewer projections is to be chosen as the syntactic representation serving that function (Law 1991, Bošković 1997)  
 b. At any point in a derivation, a structural description for a natural language string employs as few nodes as grammatical principles and lexical selection require (Safir 1993)  
 c.  $\alpha$  enters the numeration only if it has an effect on output (Chomsky 1995)

The basic idea behind (44) is that superfluous projections are disallowed. Thus, the MSP requires that every functional projection be motivated by the satisfaction of lexical requirements (such as selectional requirements and checking of features specified in lexical entries). Among other things, MSP has been argued to force the IP status on control infinitives and finite relatives and declarative complements not introduced by *that* (see Bošković 1997). Such clauses, which are potentially ambiguous in that they can be either CPs or IPs, are disambiguated by the MSP in favor of the IP option, the null operator being IP-adjoined in the relatives in question, essentially undergoing topicalization.<sup>33</sup>

- (45) a. John tried [<sub>IP</sub> PRO to leave]  
 b. the man [<sub>IP</sub> Op<sub>i</sub> [<sub>IP</sub> John left t<sub>i</sub>]]  
 c. We [<sub>VP</sub> think [<sub>IP</sub> John left]]

Bošković (1997) gives a number of arguments for the IP analysis. Thus, the analysis accounts for the ungrammaticality of short zero subject relatives, which under this analysis reduces to the impossibility of short subject topicalization (see Bošković 1997 for a uniform account of both of these).

- (46) \*the man [<sub>IP</sub> Op<sub>i</sub> [<sub>IP</sub> t<sub>i</sub> likes Mary]]

- (47) \*I think that [<sub>IP</sub> John<sub>i</sub>, [<sub>IP</sub> t<sub>i</sub> likes Mary]]<sup>34</sup>

The IP analysis also captures the contrast in (48), given Saito's (1985) claim that resumptive pronouns are not allowed under adjunction structures.

- (48) a. \*The book [<sub>IP</sub> Op [<sub>IP</sub> I was wondering whether I would get it in the mail]]  
 b. The book [<sub>CP</sub> Op [<sub>C'</sub> that I was wondering whether I would get it in the mail]]  
 (Kayne 1984)

<sup>33</sup> I assumed that complementizer that is nominal in nature, hence unaffected by the MSP.

<sup>34</sup> See Baltin (1982), Lasnik and Saito (1992), Rochemont (1989), and Saito (1985) for the IP adjunction analysis of topicalization and Lasnik and Saito (1992) for evidence that short subject topicalization is disallowed. Thus, they observe that if short subject topicalization were allowed we would expect that *John* and *himself* can be coindexed in (ib), just as in (ia).

- (i) a. John<sub>i</sub> thinks that himself<sub>i</sub> Mary likes.  
 b. \*John<sub>i</sub> thinks that himself<sub>i</sub> like Mary.

Turning to declarative complements, the IP analysis provides a straightforward account of the Comp-trace effect, which has been a recurring problem for the CP analysis. The reason why (49a) does not exhibit a Comp-trace effect under the IP analysis is trivial: there is no Comp.

- (49) a. Who do you believe left?  
b. \*Who do you believe that left?

The analysis also accounts for the obligatoriness of that with topicalization.

- (50) a. [<sub>IP</sub> Mary, [<sub>IP</sub> John likes]]  
b. Peter believes that [<sub>IP</sub> Mary, [<sub>IP</sub> John likes]]  
c. \*Peter believes [<sub>IP</sub> Mary, [<sub>IP</sub> John likes]]

Given that the embedded clause in (50c) is an IP and that topicalization involves IP adjunction, (50c) is ruled out because it involves adjunction to an argument, which is disallowed (Chomsky 1986, McCloskey 1992, Bošković 2004b). The problem does not arise in (50a-b).

A rather interesting question addressed in Bošković (1997) is how (44) interacts with Chomsky's (1995) numeration, which is defined as an array of lexical items that is mapped by the computational system into a linguistic expression. Chomsky's (44c) determines the numeration itself. This is rather problematic due to its globality. To determine the effects of (44c) we need to know PF/LF outputs. But the numeration, which is determined by (44c), must be present in the initial stage of the derivation. The problem can be solved if elements affected by (44) are not present in the numeration. Under (44a) all we need to do is define the numeration on lexical elements only. Under this view, only lexical elements are present in numerations.<sup>35</sup> Repeated access to the lexicon is then allowed to ensure we have all functional elements that are necessary to build legitimate structures. Instead of positing (44a), we can then simply require that lexicon be accessed only when needed, i.e., when a certain functional category becomes necessary in structure building. This amounts to assigning cost to merger of elements that are not taken from the numeration. Under this view, merger of such elements is subject to the ban on superfluous operations, i.e. LR. Moreover, we don't need to exempt lexical insertion from the numeration from LR: if derivations that do not exhaust numerations do not converge inserting an element from a numeration into the structure is a step toward a well-formed derivation (see Collins 1997), in accordance with LR. A tacit assumption here is that selection of lexical elements into numerations is costless. Assigning cost to numeration-formation, or trying to determine why one numeration is formed rather than another, would mean bringing the question of what we want to say into the domain of inquiry covered by the study of the working of the computational mechanism of human language. As Chomsky (1995) observes, requiring the computational mechanism of human language to deal with the issue of what we choose to say and why we choose it would not be different from requiring a theory of the mechanism of vision to explain what we choose to look at and why we do it.

Under the above approach, the MSP can be dispensed with. Its effects are derivable from the ban on superfluous steps in a derivation, i.e. LR. This is desirable, since while the MSP has an element of globality LR applies locally. The representations that the MSP rules out in favor of more economical representations cannot even be built under the derivational approach since they

---

<sup>35</sup> This seems natural if the contents of numerations are determined by what we want or choose to say, given that, as discussed above, this is determined by lexical elements.

violate LR.

## References

- Baltin, Mark R. 1982. A landing site theory of movement rules. *Linguistic Inquiry* 13: 1-38.
- Boeckx, Cedric. 2003. *Islands and chains: Resumption as stranding*. Amsterdam: John Benjamins.
- Boeckx, Cedric. 2004. Bare syntax. Ms., Harvard University.
- Bošković, Željko. 1997. *The syntax of nonfinite complementation: An economy approach*. Cambridge, Mass.: MIT Press.
- Bošković, Željko. 1999. On multiple feature-checking: Multiple wh-fronting and multiple head-movement. In *Working Minimalism*, ed. Samuel Epstein and Norbert Hornstein, 159-187. Cambridge, Mass: MIT Press.
- Bošković, Željko. 2002a. A-movement and the EPP. *Syntax* 5: 167-218.
- Bošković, Željko. 2002b. On multiple wh-fronting. *Linguistic Inquiry* 33: 351-383.
- Bošković, Željko. 2004a. Topicalization, focalization, lexical insertion, and scrambling. *Linguistic Inquiry* 35:613-638.
- Bošković, Željko. 2004b. Be careful where you float your quantifiers. *Natural Language and Linguistic Theory* 22: 681-742.
- Bošković, Željko. 2007. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry* 38: 589-644.
- Bošković, Željko. 2008a. On successive cyclic movement and the freezing effect of feature checking. In *Sounds of silence: Empty elements in syntax and phonology*, ed. Jutta M. Hartmann, Veronika Hegedüs, and Henk van Riemsdijk, 195-233. Amsterdam: Elsevier North Holland.
- Bošković, Željko. 2008b. What will you have, DP or NP? *Proceedings of NELS* 36.
- Bošković, Željko. 2008c. Unifying first and last conjunct agreement. Ms., University of Connecticut.
- Bošković, Željko. in press a. On the operator freezing effect. *Natural Language and Linguistic Theory*.
- Bošković, Željko. in press b. Review of *Wh-Movement: Moving On*, ed. Lisa Lai-Shen Cheng and Nobert Corver. *Language*.
- Bošković, Željko, and Daiko Takahashi. 1998. Scrambling and last resort. *Linguistic Inquiry* 29: 347-366.
- Chomsky, Noam. 1986. *Barriers*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from Building 20*, ed. Kenneth Hale and Samuel J. Keyser, 1-52. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 1995. Categories and transformations. In *The minimalist program*, 219-394. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89-155. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. Michael Kenstowicz, 1-52. Cambridge, Mass.: MIT Press.

- Chomsky, Noam. in press. On phases. In *Foundational issues in linguistic theory*, ed. Robert Freidin, Carlos P. Otero, and Maria Luisa Zubizarreta. MIT Press: Cambridge, Mass.
- Chomsky, Noam, and Howard Lasnik. 1993. The theory of principles and parameters. In *Syntax: An international handbook of contemporary research*, Vol. 1, ed. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann, 506-569. Berlin: Walter de Gruyter.
- Collins, Chris. 1997. *Local economy*. Cambridge, Mass.: MIT Press.
- Epstein, Samuel D, Acrisio Pires, and T. Daniel Seely. 2004. EPP in T: More controversial subjects. *Syntax* 8: 65-80.
- Epstein, Samuel D. and T. Daniel Seely. 1999. SPEC-ifying the GF “subject”; eliminating A-chains and the EPP within a derivational model. Ms., University of Michigan and Eastern Michigan University.
- Epstein, Samuel D. and T. Daniel Seely. 2006. *Derivations in Minimalism*. Cambridge: Cambridge University Press.
- Fitzgibbons, Natasha. 2007. A close-up of Russian *n*-words. Ms., University of Connecticut, Storrs.
- Franks, Steven. 2002. A Jakobsonian feature based analysis of the Slavic numeric quantifier genitive. *Journal of Slavic Linguistics* 10: 141-182.
- Grimshaw, Jane. 1997. Projection, heads and Optimality. *Linguistic Inquiry* 28: 373-422.
- Grohmann, Kleanthes. 2003. *Prolific domains: on the anti-locality of movement dependencies*. Amsterdam: John Benjamins.
- Kayne, Richard. 1984. *Connectedness and binary branching*. Dordrecht: Foris.
- Lasnik, Howard, and Juan Uriagereka. 1988. *A course in GB syntax: Lectures on binding and empty categories*. Cambridge, Mass.: MIT Press.
- Lasnik, Howard, and Mamoru Saito. 1992. *Move  $\alpha$ : Conditions on its application and output*. Cambridge, Mass.: MIT Press.
- Lasnik, Howard. 1995. Case and expletives revisited: On Greed and other human failings. *Linguistic Inquiry* 26: 615-633.
- Law, Paul. 1991. *Effects of head movement on theories of subjacency and proper government*. Doctoral dissertation, MIT, Cambridge, Mass.
- Lobeck, Anne. 1990. Functional heads as proper governors. In *Proceedings of the North East Linguistic Society* 20, 348-362. GLSA, University of Massachusetts, Amherst.
- McCloskey, James. 1992. Adjunction, selection, and embedded verb second. Santa Cruz, Calif.: Linguistics Research Center.
- McCloskey, James. 2000. Quantifier float and wh-movement in an Irish English. *Linguistic Inquiry* 31:57-84.
- Pesetsky, David. 1982. *Paths and categories*. Doctoral dissertation, MIT, Cambridge, Mass.
- Pesetsky, David, and Torrego, Esther. 2007. The syntax of valuation and the interpretability of features. In *Phrasal and Clausal Architecture: Syntactic derivation and interpretation. In honor of Joseph E. Emonds*, eds. Simin Karimi, Vida Samiian and Wendy Wilkins, 262-294. Amsterdam: John Benjamins.
- Postal, Paul M. 1974. *On raising: One rule of English grammar and its theoretical implications*. Cambridge, Mass.: MIT Press.
- Radford, Andrew. 1994. The nature of childrens initial clauses. Ms., Essex University, Essex.
- Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, Mass.: MIT Press.



- Rizzi, Luigi. 2006. On the form of chains: Criterial positions and ECP effects. In *Wh-movement: Moving on*, ed. Lisa Lai-Shen Cheng and Norbert Corver, 97-133. Cambridge, Mass.: MIT Press.
- Rochemont, Michael. 1989. Topic islands and the subjacency parameter. *Canadian Journal of Linguistics* 34: 145-170.
- Rodríguez-Mondoñedo, Miguel. 2007. The syntax of objects: Agree and differential object marking. Doctoral dissertation, University of Connecticut, Storrs.
- Safir, Ken. 1993. Perception, selection, and structural economy. *Natural Language Semantics* 2: 47-70.
- Saito, Mamoru. 1985. Some asymmetries in Japanese and their theoretical implications. Doctoral dissertation, MIT, Cambridge, Mass.
- Saito, Mamoru, and Keiko Murasugi. 1990. N'-deletion in Japanese. In *University of Connecticut Working Papers in Linguistics* 3, ed. Javier Ormazabal and Carol Tenny, 87-107. University of Connecticut, Storrs.
- Saito, Mamoru. 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1: 69-118.
- Schneider-Zioga, Patricia. 1995. Specifier/head agreement in Kinande. *Cahiers Linguistiques d'Ottawa* 23: 67-93.
- Schneider-Zioga, Patricia. 2005. Bounded unbounded movement. Ms. University of Southern California, Los Angeles.
- Speas, Margaret. 1994. Null arguments in a theory of economy of projection. In *University of Massachusetts Occasional Papers* 17, 179-208. Department of Linguistics, University of Massachusetts, Amherst.
- Surányi, Balázs. 2006. Mechanism of wh-saturation and interpretation in multiple wh-movement. In *Wh-movement: Moving on*, ed. Lisa Lai-Shen Cheng and Norbert Corver, 289-318. Cambridge, Mass.: MIT Press.
- Takahashi, Daiko. 1994. Minimality of movement. Doctoral dissertation, University of Connecticut, Storrs.