

# Feature-Splitting Internal Merge: Improper Movement, Intervention, and the A/A' Distinction

*Miki Obata and Samuel David Epstein*

*Abstract.* A central goal of this paper is to present a new account of improper-movement phenomena based on Chomsky's (2007, 2008) phase-based derivational approach. We claim that improper movement is excluded by virtue of Agree failure between a moving element and a finite T as a consequence of "feature-splitting" Internal Merge, which we argue is the most (or at least a very) natural implementation of Chomsky's  $\phi$ -feature-inheritance system and Richards's (2007) value-transfer-simultaneity analysis. This analysis has a number of empirical and theoretical consequences: (i) regarding the explanation of A'-opacity/transparency intervention effects (Rezac 2003, Carstens 2005); (ii) the possible elimination, or reduction in scope, of the Activity Condition; and (iii) the possible characterization of A/A'-position types solely in terms of categorial features. Moreover, we propose that (iv) the ban on improper movement is, in fact, not universal but is morphologically parameterized (at least) between English and the Bantu language Kilega.

## 1. Introduction

A central goal of this paper is to present a new generalized, parameterized, and agreement-based account of improper-movement phenomena (first discussed in Chomsky 1973), as in (1c)/(2).

- (1) a. John seems to be intelligent.  
 b. It seems that John is intelligent.  
 c. \*John seems (that) is intelligent.
- (2) [TP John<sub>3</sub> seems [CP <John<sub>2</sub>> [TP <John<sub>1</sub>> [VP is intelligent]]].  
 A ← A' ← A (disallowed)

Although May's (1979) Condition C analysis and Fukui's (1993) Chain Uniformity analyses are both insightful and intriguing, the account we present here is based only on derivational and local computations as exemplified particularly in Chomsky's (2007, 2008) phase-based derivational approach.<sup>1</sup>

This paper is a revised and extended version of Obata & Epstein 2008a. Aspects of this work were presented at the Crash-Proof Grammar conference (February 2008), GLOW 31 (March 2008), and WCCFL 27 (May 2008). We would like to thank the audiences at those conferences for insightful comments and suggestions. Also, we are very grateful to two anonymous *Syntax* reviewers, Klaus Abels, Christopher Becker, Željko Bošković, Cedric Boeckx, Noam Chomsky, Betty McNulty Epstein, Gerardo Fernández-Salgueiro, Catherine Fortin, Samuel Gutmann, Mark Hale, Jason Merchant, Chizuru Nakao, Acrisio Pires, Mike Putnam, Koji Sugisaki, Takashi Toyoshima, Jan-Wouter Zwart, and especially to Vicki Carstens, Hisatsugu Kitahara, Milan Rezac, and Daniel Seely for extremely valuable and helpful discussion, suggestions, and comments.

<sup>1</sup> For other universal prohibitions against "improper movement," see Abels 2007 and Williams 2002. In this paper, we argue that the legitimacy of improper movement is parameterized, and we suggest that accounts based on universals, including the fixed architecture of UG, are too restrictive.

We claim that improper movement is excluded by virtue of Agree failure between a moving element and a finite T as a consequence of “feature-splitting” Internal Merge. We propose feature splitting as the most (or at least a very) natural implementation of Chomsky’s  $\phi$ -feature-inheritance system and Richards’s (2007) value–transfer simultaneity. Additionally, this analysis is empirically supported by and seeks to explain, without stipulation, A’-opacity intervention effects as discussed in Rezac 2003. Furthermore, the proposed account enables us to rule out improper movement without appealing to the arguably stipulative Activity Condition (see Nevins 2005 and Bošković 2007 for empirical arguments against the Activity Condition).<sup>2</sup> One of the potentially intriguing consequences of our analysis is that it might enable us to recapture A/A’-position types in terms of “features” on categories—specifically, the presence or absence of  $\phi$ -features. We also claim that improper movement is, in fact, parameterized (not universally excluded), in contrast to previous approaches. Data from Bantu languages such as Kilega and Lusaamia (Carstens 2005, 2008) suggest to us that there are certain relations between C-agreement phenomena and “proper” improper movement.

The paper is organized as follows: Section 2 reviews some theoretical background and presents the feature-split analysis. Section 3 illustrates how feature split works, particularly with regard to bans on improper movement, and also articulates several empirical and theoretical consequences of our proposal including the A/A’ distinction. Section 4 extends the data to Bantu languages and discusses the parameterization of (the morphosyntactic phenomena of) feature split. Section 5 concludes our study.

## 2. The Logic of Feature Splitting

### 2.1 Chomsky’s (2007, 2008) Feature-Inheritance System

In this section, we consider Chomsky’s (2007, 2008) feature-inheritance system and reveal new aspects of its predictive content. Since *On phases* (2008), the treatment of T (and also V) has radically changed: T does not bear  $\phi$ -features inherently, but rather inherits them from C.<sup>3</sup> Consequently, T cannot operate as a probe until C is introduced into the derivation. This feature-inheritance analysis from C to T makes it possible to nonlexically distinguish a finite/control T from the one appearing with raising predicates. T in raising predicates is not selected by C, so T does not inherit  $\phi$ -features. Lacking  $\phi$ -features, the T appearing with raising predicates does not have the ability to value Case. That is, the availability of C determines the potential of T. Therefore, we no longer need to stipulate that the lexicon contains two different Ts. Rather, the features of the sole lexical T are determined “functionally” by the

<sup>2</sup> In this paper, we pursue an account without appeal to the Activity Condition. (See Chomsky 2007:fn. 31 for an attempt to eliminate or reduce appeal to the Activity Condition). Although there are still cases that the Activity Condition explains well, the attempt to eliminate the condition, we believe, contributes to constructing a theory with fewer UG-specific properties and perhaps stipulations.

<sup>3</sup> We assume that the same feature-inheritance system applies to v/V, following Chomsky 2007, 2008: V inherits its features from v. For the limited scope of this paper, however, we mainly focus on C-to-T inheritance.

(selectional) context in which the T appears. Crucially, under this system, movement to Spec,CP and to Spec,TP takes place simultaneously and independently. (See Chomsky 2007 and 2008 for more details.) This analysis, whereby certain movements to Spec,CP do not proceed through Spec,TP (more generally, A-movement does not feed A'-movement), is empirically motivated to account for asymmetries in the suppression of subject condition effects as presented in Chomsky 2008. This system clarifies the possible status of T as a phase head: phases are only CP and vP, but not T in that it does not have the ability to serve as a probe by itself.

We adopt the feature-inheritance system in the following discussion and further consider its implications. A fundamental question concerning this system still remains unanswered: How/why do features on C get inherited by T? Richards (2007) claims that all of the uninterpretable features (henceforth [uF]) on C are sent to T as a necessary precondition for convergence. He deduces this from Chomsky's system of feature-valuation and the timing of Transfer, whereby representations generated in the narrow syntax are "sent to" the interfaces. According to Richards's argument, Transfer (i.e., transfer to the interfaces) and feature valuation must occur simultaneously for convergence. That is, one operation can neither precede nor follow the other. If Transfer applies *before* feature valuation, [uF] is sent to the interface unvalued, which causes crash of the derivation. If Transfer applies *after* feature valuation, on the other hand, the distinction between syntactically valued [uF] and [iF] disappears in the eyes of Transfer (See Chomsky 2001, 2007 and Epstein & Seely 2002 for analysis and possible problems). This leads to crash of the derivation because the operation Transfer fails to remove now valued [uF] from syntactic objects that are sent to the semantic component. With respect to this issue, Chomsky (2007:18–19) writes:

If transferred to the interface unvalued, uninterpretable features will cause the derivation to crash. Hence both interface conditions require that they cannot be valued after Transfer... Furthermore, [Transfer] cannot take place after the phase level at which they are valued, because once valued, they are indistinguishable at the next phase level from interpretable features, hence will not be deleted before reaching the Conceptual-Intentional (CI) interface. It follows that they must be valued at the phase level where they are transferred, that is, at the point where all operations within the phase take place and the Transfer operation therefore "knows" that the feature that has just been valued is uninterpretable and has to be erased at (or before) CI.

That is, valued [uF] has to be deleted by Transfer early enough for the computational system to distinguish it from inherently interpretable features [iF]. This entails that valuation and Transfer must occur simultaneously for a derivation to converge. Given this argument, Richards (2007) suggests that [uF] cannot remain on C but has to be discharged to T. This is because Transfer of a phase edge (including C) is suspended until the domain of the next higher phase is transferred, based on the Transfer/Spell-out system suggested in Chomsky 2000. But syntactically valued features appearing at the edge (such as valued [u $\phi$ ] on C, which would appear if C did *not* transfer its  $\phi$  to T) are valued [uF]. These valued [uF] are indistinguishable from inherently interpretable features [iF]. Transfer will not know to remove them at the next phase level, and the derivation will crash. Therefore, Richards argues, convergence is

possible only when all of the valued [uF] are included in the domain of a phase. This is precisely what C-to-T feature inheritance accomplishes.<sup>4</sup>

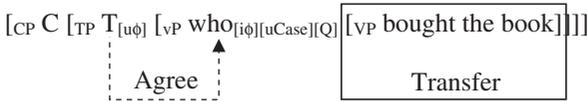
This analysis of simultaneity of Transfer and feature valuation has an interesting implication. Richards only focused on features on a phase head (probe), so that feature inheritance must happen for convergence. But, what about goals such as DP? DP also bears [uF], namely [uCase], which is valued by T (or V). If a phase head (which, by definition, occupies the edge) is not allowed to bear valued [uF], the same should be true of elements *moved to* phase-edge positions, because Transfer of these objects is also suspended until the domain of the next higher phase.<sup>5,6</sup> In the next section, we will address this issue and propose a new analysis.<sup>7</sup>

### 2.2 The Mechanics of Feature Splitting

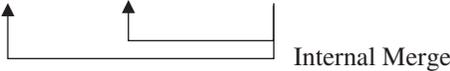
As discussed in the last section, no valued [uF] is allowed to appear at a phase edge, given the Chomsky-Richards analysis. For instance, valued [uCase] on DP never occurs at Spec,CP in a convergent derivation:

(3) Who do you think bought the book?<sup>8</sup>

a. Embedded CP

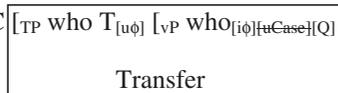


b. [CP **who** C [TP **who** T[uφ] [vP **who**[iφ][uCase][Q] [VP...]]]



(A and A'-movement occur separately; i.e., A-movement does not feed A'-movement)

c. [CP **who** C [TP **who** T[uφ] [vP **who**[iφ][uCase][Q] [VP...]]]



<sup>4</sup> One might think that languages that have C-agreement are problematic for this view in that C cannot trigger Agree because it lacks [uF] given feature inheritance from C to T. We will discuss this issue later in this paper.

<sup>5</sup> A reviewer pointed out a potential problem with our extension of Richards 2007 to [uCase] on moving elements: As opposed to  $\phi$ -features, Case features do not have the interpretable counterpart, so that valued [uCase] is never confused with [iCase], which does not exist. However, we assume, following Chomsky, that Transfer just “blindly” eliminates valued [uF] unless some additional mechanisms—such as Transfer “knowing” the (look-ahead) CI interpretability of features—are stipulated. In other words, Transfer only “sees” that feature values [uF] or [iF] because interpretability is the concern of only the interfaces (see Epstein et al. 1998). Also, it seems far from clear that there is no [iCase] if inherent Case is taken into account.

<sup>6</sup> See Obata 2009 for potential exceptions to this assumption in matrix clauses.

<sup>7</sup> In later discussion, we will assume that features are inherited before Agree takes place, following Chomsky’s original claim.

<sup>8</sup> We will ignore the accusative Case assignment to make the discussion simpler.

Let us focus only on the derivation of the embedded CP. The embedded C is introduced into the derivation bearing [u $\phi$ ], which are then inherited by T. T bearing [u $\phi$ ] agrees with the subject DP *who* at Spec,vP bearing [i $\phi$ ] as in (3a). As a consequence of this  $\phi$ -agreement, [uCase] on *who* is valued. The edge feature on C and on T each (independently and simultaneously) attracts *who* to its edge position as illustrated in (3b). Then, TP is transferred. If Richards's analysis is on the right track, *who* at the edge of CP must *not* bear any valued [uF]. If it did, the derivation would incorrectly crash.

There appear to be at least two ways to overcome this problem. One is valued [uCase] on *who* "disappears" by some mechanism, so that it is not copied to the occurrence of *who* at the edge of CP. The other way is that valued [uCase] does appear at the edge, but (contra Chomsky) the computational system, specifically, Transfer, still can see the difference between valued [uF] and [iF]. This assumption would in turn deprive us of maintaining Richards's inheritance deduction, because it is inconsistent with Richards's deduction by virtue of allowing valued features on the edge. In this paper, we will pursue the former possibility and will maintain Richards's (2007) claims. (See Epstein, Kitahara & Seely 2008 for further discussion regarding the latter possibility.)

What mechanism makes it possible that valued [uCase] on *who* at the edge of vP is not copied to the edge of CP? Recall that T can work as a probe only by receiving features from C. After features are inherited from C, T finally begins to work as a probe. Then, as exemplified in (3b), the edge feature on C and on T each independently attracts the single element *who* occupying the edge of vP, which explains the suppression of Subject Condition effects (because A-movement does not feed A'-movement). (See Chomsky 2008:10 for more detail.) What happens in the simultaneous attraction of the single element by these two different heads? We propose that features on the attractee are "split" into the two different landing sites (= feature splitting).

(4) (= (3b)): Feature Splitting



As a consequence of this simultaneous attraction by C and T, features on *who* are decomposed into the two positions as displayed above. Since Chomsky 2000, Case valuation on DP has been a reflex of  $\phi$ -valuation, so that [Case] and [ $\phi$ ] are unified. As in (4), if T attracts [Case]/[ $\phi$ ] and C attracts [Q], valued [uCase] on DP is transferred as a part of the TP domain of the phase head C and therefore this valued [uCase] never makes it to the edge of CP, and Richards's (deductive) condition is satisfied. Additionally, the derivation converges, as required for empirical adequacy. Furthermore, the feature-split analysis is compatible with Chomsky's (1964/1995a) view that

nonbranching lexical *wh*-phrases are in fact composed of two distinct morphological feature sets: [wh] on the one-hand and an indefinite QP like *something*. Given this view, we naturally hypothesize that [wh] and *something* each move independently.

One might ask the following two things regarding feature splitting: (1) what mechanism makes it possible that [uCase]/[iφ] are attracted by T but not by C? and (2) why can a single feature or featural proper subset land at edge/specifier positions?<sup>9</sup> One possible answer to the first question is that T only attracts features it has agreed with. In (4), T φ-agrees with [iφ] on DP and [uCase] is valued as its reflex. That is, both [uCase] and [iφ] on DP “participated” in agreement with T. In contrast, [Q] is not involved in this φ-agreement. Therefore, T attracts only [uCase]/[iφ], which it φ-agreed with, whereas C attracts the rest, namely [Q]. In fact, this view is fully compatible with (if not forced by) Chomsky’s (2007:24) characterization of the A/A’ distinction: “A-movement is IM (internal merge) contingent on probe by uninterpretable inflectional features, while A’-movement is IM driven by EF.” In other words, an A-position is one that results from attraction by the edge feature under Agree, in contrast to A’-movement, which is triggered solely by the edge feature. The difference between A and A’-movement has been widely accepted. Feature splitting, whereby T attracts only features that it agreed with, enables us to capture (derivationally) two different types of movement. In this sense, our feature-split analysis is a natural implementation of Chomsky’s A/A’ distinction and is deduced from the (explicable) prohibition against valued [uF] on the edge. Note also that if interpretation is at the interface—these distinct features allow us to featurally represent A versus A’ in the interface representation, which presumably does not have access to the movement type that created the position “back in” the narrow syntax. In this sense, perhaps feature splitting is the narrow syntax optimally exploiting independent mechanisms (Agree vs. pure edge attraction) to featurally distinguish position types, rendering stipulative position-type definitions (e.g., “Spec,CP = A’-position, and Spec,TP = A-position”) eliminable.

As for the second question, a single feature or featural subset *can* be regarded as a maximal projection under bare phrase structure: “a category that does not project any further is a maximal projection XP and one that is not a projection at all is a minimal projection X<sup>0</sup>” (Chomsky 1995b:396). In the configuration of (4), the moved [wh] D<sup>max</sup> does not project further at the edge of CP and at Spec,TP. Therefore, those split features can be regarded as maximal projections in their own right. In this sense, our analysis also lends support for bare phrase structure.<sup>10,11</sup>

<sup>9</sup> This is an extension of Chomsky’s (1995b) MOVE-Feature proposal. Also see Toyoshima 2000 and Matushansky 2006 for related ideas.

<sup>10</sup> Hisatsugu Kitahara (p.c.) suggests that feature splitting may correspond to DP decomposition into D and NP (assuming that DP is not a phase): D bearing [Q] moves to an edge of CP/vP and NP bearing [φ]/[Case] moves to Spec,TP/vP.

<sup>11</sup> Notice that feature split violates neither the inclusiveness condition nor the no-tampering condition. Feature split involves no new features but only splits existing features observing the former condition. With respect to the latter condition, in the configuration of (4), [Q] is internally merged by C. The no-tampering condition says that merge of [Q] and C leaves the two syntactic objects unchanged. Feature split only makes it possible that [Q] moves separately from [φ] and [Case], so that it does not affect the merged syntactic objects. That is, feature split is compatible with these two conditions.

In a nutshell, for Richards’s deduction based on Chomsky’s system to go through, there is a need to avoid the appearance of valued [uF] on a phase head (Richards’s deduction of inheritance) and, as we note here, there is entailed a more general prohibition against any valued [uF] including those of a goal appearing at an edge position. As for a phase head probe, Chomsky’s feature-inheritance system forces those features on C to lower to T. Regarding goals internally-merged to phase edges, our analysis splits off valued [uCase], and just like C-to-T feature inheritance (deduced by Richards), keeps valued [uF] “off the edge” and within the transferred phase head complement ensuring convergence hence empirical adequacy. In the next section, we present empirical and theoretical advantages provided by the feature-split analysis by considering improper-movement phenomena as a case study.

### 3. Improper Movement as an Agreement-Failure Phenomenon Causing Featural Crash

#### 3.1 Agreement Failure of T

In this section, we demonstrate how the ban on improper movement is recaptured locally under the current phase-based derivational approach equipped with feature splitting (without appeal to unbounded dependencies expressed by Condition C and Chain-based approaches, and without appeal to the Activity Condition). There are two types of improper movement. Consider (5), for example.

- (5) a. \*Who seems it is likely to leave?
- b. \*Who seems will leave?

In (5a), [uCase] on *who* is valued *after* movement to the edge of the embedded CP. By contrast, [uCase] on *who* in (5b) is valued *before* movement to the edge of the embedded CP. The former case can be explained by saying that [uCase] on *who* is transferred unvalued along with the lowest TP, which causes crash of the derivation. The same scenario does not go through for the latter case because [uCase] on *who* is properly valued by the embedded T. Let us see how the latter case is derived under the feature-splitting approach (indices on *who* for expository purposes only):

- (6) \*Who seems will leave?

- a. Embedded vP

[vP <who<sub>1[Q][φ][Case]> [VP leave]]</sub>

- b. Embedded CP

[CP <who<sub>3[Q]> C [TP <who<sub>2[iφ][uCase]> T [vP <who<sub>1[Q][iφ][uCase]> [VP...]]]]]</sub></sub></sub>

- c. Matrix CP

[CP C<sub>[EF]</sub> [TP T<sub>[uφ]</sub> seems [CP <who<sub>3[Q]> [TP...]]]]</sub>

Embedded-TP Transfer

In (6a), the embedded subject *who* is externally merged into Spec,vP. In (6b), after embedded T  $\phi$ -agrees with *who*, the edge feature on C and on T each attracts the single element *who* occupying Spec,vP. As mentioned in the last section, T attracts only a featural subset of *who* that it agreed with and C attracts the rest. That is, features on *who* are split into the edge of CP and Spec,TP: [uCase]/[i $\phi$ ] move to Spec,TP while [Q] moves to the edge of CP. Then, the embedded TP is transferred and the derivation goes on to the matrix clause. In (6c), [u $\phi$ ] on the matrix T seeks a Matching Goal with [i $\phi$ ]. However, *who* at the edge of the embedded CP has already lost [i $\phi$ ] as a consequence of feature split. Additionally, *who* at the embedded Spec,TP is not in the minimal search domain of the matrix T because of the Phase Impenetrability Condition (PIC).<sup>12</sup> Therefore, [u $\phi$ ] on matrix T is not valued, which causes crash of the derivation. The derivation in (6) is excluded as desired.<sup>13</sup>

What happens in analyses lacking feature split? If there is no feature split, it is predicted that *who* in Spec,CP still has its inherent [i $\phi$ ]. Therefore, in the configuration of (6c), the matrix T can  $\phi$ -agree with *who* in embedded Spec,CP. The derivation converges,<sup>14</sup> yielding the wrong prediction. What mechanism excludes this derivation if feature splitting is not assumed? A separate principle, namely the Activity Condition, needs to be stipulated, where the computational system cannot see elements that lack unvalued features. Thus, even if *who* at the edge of the embedded CP has [i $\phi$ ], there is no [uF] on it. Therefore, the Activity Condition prohibits the probe (i.e., the matrix T) from agreeing with the deactivated *who*. As a result, [u $\phi$ ] on the matrix T is not valued, which causes crash. If feature splitting does not exist, the Activity Condition is crucial for excluding such improper-movement phenomena. But the Activity Condition is a stipulation—that is, *who* in embedded Spec,CP bearing lexically inherent [i $\phi$ ] is in the minimal search domain of upstairs T and can value T's  $\phi$ -features under Match. In other words, feature splitting, which is naturally induced from Chomsky's system, enables us to explain improper movement without appeal to the SMT-violating Activity Condition. Given that Nevins (2005) demonstrates that the Activity Condition is empirically problematic, and partially redundant with PIC, our approach is arguably preferable in this respect.<sup>15</sup>

In this section, we demonstrated how improper movement is excluded as a direct consequence of the feature-split analysis. In the next section, we will suggest further empirical advantages obtained from our analysis.

<sup>12</sup> Phase Impenetrability Condition (Chomsky 2000:108): In phase  $\alpha$  with head H, the domain of H is not accessible to operations outside  $\alpha$ , only H and its edge are accessible to such operations.

<sup>13</sup> An anonymous reviewer notes as a potential problem the question of how the operator-variable binding relation can be established, given that the moved *wh*-phrase does not itself bear  $\phi$ -features under feature split. The treatment of binding relations is a general issue that needs to be closely considered within an inclusive minimalist approach and requires further investigation. (It is not clear to us that operator-variable binding invariably requires that  $\phi$ -features appear on the operator, as in, e.g., *wh*-adjunct variable binding.)

<sup>14</sup> We propose that precisely this happens in English *tough*-constructions. See Obata & Epstein 2008b for more detail (cf. Rezac 2006 for a related previous analysis of *tough*-constructions).

<sup>15</sup> See also Bošković 2007 for attempts to eliminate the Activity Condition.



narrow syntax (not in the phonological component) cares about or can detect phonological features without narrow syntax look-ahead to the phonological component.<sup>17,18</sup> That is, it is not clear how the computational system “knows” that *what cars* moves further (given copy theory) or stays in situ at the narrow syntax-internal time of T agreement with the subject, an operation that precedes the determination of which parts of the *wh*-chain will be  $\pm$ phonological (nor is it clear to us why in general Agree intervention effects should be sensitive to the presence vs. absence of phonological features at all). Our system of feature splitting explains these phenomena (without look-ahead or Agree-sensitivity to intervening phonological features) by hypothesizing that elements attracted by the edge feature on phase heads, C or v, cannot keep their  $\phi$ -features. If an intervening element has no  $\phi$ -features, it does not serve as a blocker of  $\phi$ -agreement. In this sense, we can say that our analysis unifies improper movement and A'-Opacity effects, explaining each as the result of the fact that A'-elements lack  $\phi$ -features.<sup>19</sup> The data in (7)–(8) further support the feature-split analysis.<sup>20,21,22</sup>

<sup>17</sup> See Kitahara 2006 for more detailed discussion regarding this issue.

<sup>18</sup> The problems regarding Chomsky's (2001) account appealing to phonological features are also pointed out in Richards 2004.

<sup>19</sup> One might wonder how the current system deals with a sentence like *Whom did you see?* in that the topmost copy of *whom* has neither [Case] nor [ $\phi$ ], which are already split off in the course of the derivation, but still exhibits the Case morpheme *-m*. With respect to this problem, one might adopt Lasnik & Sobin's (2000) analysis, where the addition of the morpheme *-m* is an independent operation of Case valuation. That is, *-m* is not an accusative Case morpheme accompanying Case valuation but rather is attached by an “extra” grammatical operation, which they call a “grammatical virus.” On the analyses of such examples, see also Chomsky 2007:18 and for a counterproposal, see Epstein, Kitahara & Seely 2008.

<sup>20</sup> See Heck & Muller 2003 for (what an anonymous reviewer calls) an “eventually possibly problematic paradigm” regarding A'-opacity.

<sup>21</sup> A reviewer suggests that the same correct results follow from maintaining the activity condition, (a condition we seek to eliminate following Nevins 2005, Chomsky 2007, Bošković 2007). The reviewer writes with respect to (7) and (8) that “the activity condition straightforwardly rules out A'-intervention. The A'-intervener is simply  $\phi$ -inactive by virtue of already being Case-marked.” In (8), for example, we agree with the reviewer that the activity condition correctly prevents object agreement between T and (the inactivated) *what cars*. However, we disagree with the reviewer that the activity condition allows the necessary T-Subject agreement that is blocked in the configuration (8) by the defective intervention constraint in analyses lacking feature splitting.

<sup>22</sup> Another type of data from Icelandic, pointed out by a reviewer, seems to be worth thinking about. Svenonius (2000) presents the following data involving quantifier movement in Icelandic.

- (i) a. Á prófinu mun hann snast geta margt.  
       on the-test will he seem solve many  
       ‘On the test he will seem to be able to solve many [problems].’  
       b. Á prófinu mun hann margt snast geta.  
       on the-test will he many seem solve  
       ‘On the test he will seem to be able to solve many [problems].’

Here, the subject ‘he’ moves from the subject position of ‘solve’. Given that the moved quantifier ‘many’ lands at the edge of a phase in (ib), the subject ‘he’ is transferred before the matrix T agrees with it because ‘he’ is included in a domain of the vP phase headed by ‘seem’. That is, our system predicts that this type of Icelandic data is wrongly ruled out.

However, though this is a problem for us, we think this is not a problem only for us but rather a general problem for this particular formulation of the timing of Transfer, as the reviewer also points out. Regardless of assuming feature splitting, these data cause a problem because as long as ‘many’ lands at the phase edge, a domain including ‘he’ is transferred along with [uCase], causing crash. Although we find these data important, we will leave this issue for future research.

3.3 *Eliminating the Concept of A/A'-Positions: A Conjecture*

The current approach to improper movement and intervention also throws new light on the A/A' distinction. The existence of two different types of positions has been widely accepted given distinct behaviors in, for example, binding. At least three ways to distinguish these two position types have been proposed:

(9) Chomsky (1981:45):

An A-position is one in which an argument such as a name or a variable may appear in D-structure; it is a potential  $\theta$ -position. The position of subject may or may not be a  $\theta$ -position, depending on properties of the associated VP. Complements of  $X'$  are always  $\theta$ -positions....An A'-position is that of an adjunct of one sort or another.

(10) Chomsky & Lasnik (1995:64):

Given a lexical head L, we say that a position is L-related if it is the specifier or complement of a feature of L. The L-related positions are the former A-positions [i.e., the non-L-related positions are A'-positions]....<sup>23</sup>

(11) Chomsky (2007:24):

A-movement is IM (internal merge) contingent on probe by uninterpretable inflectional features, while A'-movement is IM driven by EF.

In Chomsky 1981, A- and A'-positions are defined in terms of potential  $\theta$ -positions, as shown in (9). As pointed out in Chomsky & Lasnik 1995, however, this definition is not straightforward in that a certain position needs to be regarded as an A-position despite the fact that there is no  $\theta$ -relation to its head because the definition is based on transderivational comparison. Chomsky & Lasnik (1995) redefine it in a less global way by introducing the concept of "L-relatedness," as in (10). The L-related positions include specifiers and complements of Agr and T in that those heads bear features of V inherently but not C. The definitions in (9) and (10) refer to phrase structure representations. In contrast, in Chomsky 2007, the differences between A-movement and A'-movement are explained by appeal to what features are involved in Internal Merge, as in (11). That is, this definition is for distinguishing not "positions" but rather different types of movement by making use of features involved in Internal Merge. (Although (11) captures the A/A' distinction for Internal Merge in terms of triggering features, it does not characterize the A/A' status of externally-merged elements, such as adjuncts or arguments, given that A/A' is defined in (11) only with respect to Internal Merge. The A/A' distinction for External Merge is captured by Chomsky's Set-Merge vs. Pair Merge.)

As mentioned in section 3.1, feature splitting decomposes features on a single lexical element. If the subject *wh*-phrase *who* is taken as an example, *who* bears [uCase], [i $\phi$ ] and [Q]. Then, T attracts [ $\phi$ ]/[Case], which it agreed with, and C attracts

<sup>23</sup> Interestingly, note that (10) refers to "complement of a feature" as opposed to complement of a head (= bundle of features).

[Q] purely by EF. That is, a copy of *who* bearing [ $\phi$ ]/[Case] is located at Spec,TP and the other copy bearing only [Q] is placed at the edge of CP. Based on the feature-split analysis, we suggest the following definition:

- (12) An A-position is a category bearing  $\phi$ -features, whereas an A'-position is a category lacking  $\phi$ -features.

The definition in (12) is distinct from Chomsky's (1981) and Chomsky & Lasnik's (1995) type of definitions in that we recapture position types in terms of  $\phi$ -features without appeal to nonfeatural position-typing stipulations. Another conceptual advantage is that the current definition no longer struggles with answering why position types matter in the CI representation under binary compositional interpretation of categories. Rather, it is preferable that interpretable  $\phi$ -features (or Q-features), which are all sent to the CI interface, not to the sensory-motor (SM) interface, play important roles for interpretation. In particular, traditional A/A'-positions behave differently with respect to binding phenomena, which are presumably computed in the semantic component. It seems conceptually more plausible that the availability of  $\phi$ -features serves as the crucial factor in distinguishing the two kinds of positions. Also, in an interface representation, we cannot see that Spec,TP was earlier created by an Internal Merge operation applied in the narrow syntax in which T, then bearing uninterpretable features that are crucially absent in the convergent CI representation, probed DP. Recapturing A/A' distinctions in terms of features on goals, but not in terms of probes, enables the CI interface to detect the A/A' status of a certain element without appealing to features that were present in the narrow syntax but are absent in the CI representation. This is the crucial difference between our approach based on  $\phi$ -features on goals and Chomsky's (2007) approach appealing to features triggering Internal Merge.<sup>24</sup>

Furthermore, Lasnik & Stowell's (1991) weakest-crossover effects lend empirical support for our view:

- (13) \*Mary asked me who to persuade his mother to vouch for?  
 (Mary asked me [ $who_1$  [PRO to persuade [ $his_1$  mother] $_2$  [PRO $_2$  to vouch for  $t_1$ ]]])  
 (Lasnik & Stowell 1991:703)
- (14) Who should be easy to persuade his mother to vouch for?  
 (Who $_1$  should be easy [ $Op_1$  [PRO to persuade [ $his_1$  mother] $_2$  [PRO $_2$  to vouch for  $t_1$ ]]])  
 (Lasnik & Stowell 1991:703)

<sup>24</sup> An anonymous reviewer indicates that our approach based on availability of  $\phi$ -features predicts that the edge of the embedded CP is an A-position in a sentence like \**Who seems it is likely to leave?*, which we discussed in section 3.1. [uCase] on *who* is not valued in the embedded clause, so that features on *who* are never split before moving to the edge of the embedded CP. That is, *who* lands at the edge of the embedded CP still bearing  $\phi$ -features. Although this derivation never converges because [uCase] on the copies of *who* in the embedded TP are transferred following Nevins's (2005) analysis of improper movement, this nonetheless seems to be an interesting prediction made by our analysis. At this point, however, we do not know of any way to test the prediction.

As (13) and (14) show, null-operator movement does not cause weak crossover effects in (14) in contrast to (13), where the lexical *wh*-phrase (i.e., true quantifier) moves. Although weak crossover effects are widely used as a diagnostic for *A'*-movement, Lasnik & Stowell's findings imply that it is descriptively inadequate to say that weak crossover is caused by a certain configuration of *A'*-binding. The weakest crossover effects instead suggest that only a true quantifier, which can be defined as a feature of categories, not merely an *A'*-position, causes weak crossover effects. If so, the CI system needs to see features inside a category to decide whether an *A'*-binder is a true quantifier or not. (See also Koopman & Sportiche 1983:n. 17 regarding topicalization of a proper name and Chomsky 1982:n. 11 regarding relative pronouns suppressing Bijection [WCO] violations). These facts seem to support our hypothesis that *A/A'* properties should be at least in part defined categorially/featurally.<sup>25</sup>

### 3.4 Section Summary

In section 3, we have shown how our feature-split analysis rules out improper movement and explained *A'*-opacity effects, whereby elements at *A'*-positions are invisible to Agree between T and a subject. Moreover, feature splitting, which is compatible with Chomsky's (1964, 1995b) idea of compositional *wh*-phrases, provides us a new possibility for capturing the *A/A'* distinction in terms of features of categories. In section 4, we extend our discussion to the Bantu language Kilega and consider potential problems our analysis confronts. We also discuss morphosyntactic variation regarding feature splitting as well as the parameterization of *A'*-opacity effects and "proper improper movement."

## 4. Feature Split Is Parameterized

### 4.1 Brief Review and Problems

Recall that one of the crucial motivations for the C-to-T feature-inheritance system in Chomsky 2007, 2008, which our system is based on, is that C cannot keep [uF] for a convergent derivation. Rather, [uF] has to be inherited by T. As a result of the inheritance, T attracts features that it agreed with, whereas C attracts the rest of the features purely by the edge feature (i.e., without Agree). That is, the prediction is:

(15) C never directly triggers Agree because C lacks [uF], given feature-inheritance.

One problem we must address then is the apparent phenomenon of C-agreement, which appears inconsistent with (15).

Second, concerning *A'*-opacity, our feature-split analysis renders a DP copy at a phase edge without  $\phi$ -features, thus explaining its invisibility to  $\phi$ -agreement in section 3.2. This mechanism predicts:

<sup>25</sup> As mentioned in fn. 13, we do not make any claim in this paper regarding how to capture binding phenomena under the current minimalist approach. That is, weakest crossover data are used here only to demonstrate that the traditional definition of *A'*-positions referring only to phrase-structure geometry is not sufficient but rather feature specifications serve as defining factors in such phenomena.

- (16) Once an element is attracted to a phase edge, the element can never be the goal of  $\phi$ -agreement.

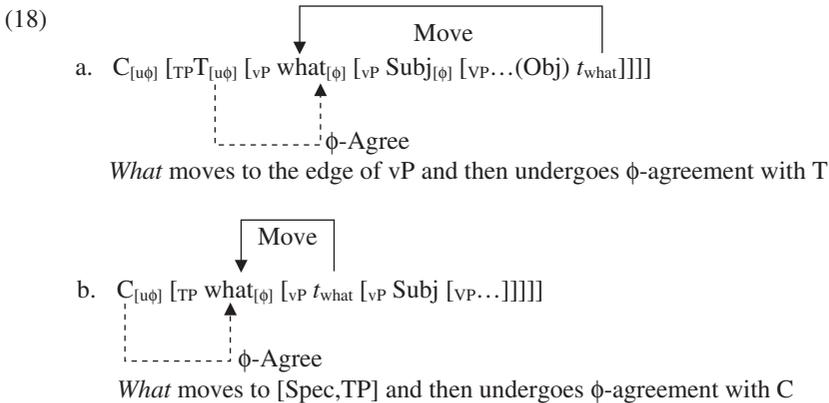
As we will see, this too is not universally true. In the next section, we explore potential problems and a solution based on a descriptive morphosyntactic parameter.

4.2 Potential Problems: C-Agreement Languages

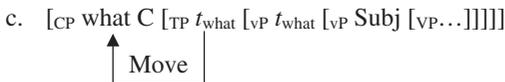
In this section, we consider Kilega, which is a Bantu language that appears to violate our predictions that (i) a phase head never triggers Agree and (ii) an element at a phase edge never undergoes  $\phi$ -agreement. Consider the following *wh*-constructions in Kilega (from Carstens 2005:220).<sup>26</sup>

- (17) a. Bábo bíkulu b-á-kás-íl-é mwámí bíkí mu-mwílo?  
**2.that 2.women 2.SA-A-give-PERF-FV 1.chief 8.what 18-3.village**  
 ‘What did those women give the chief in the village?’  
 b. Bíkí bi-á-kás-íl-é bábo bíkulu mwámí mu-mwílo?  
**8.what 8.CA-A-give-PERF-FV 2.that 2.woman 1.chief 18-3.village**  
 ‘What did those women give the chief in the village?’

Kilega *wh*-constructions allow both in-situ *wh*-questions as in (17a) and overt *wh*-fronting as in (17b). When the *wh*-phrase *bíkí* remains in situ in (17a), the verb agrees with the subject *bábo bíkulu* and bears the subject agreement morpheme. They also agree in noun class, as indicated by the number ‘2’. In contrast, when the *wh*-phrase overtly moves as exemplified in (17b), the verb moves to C and agrees with the fronted *wh*-phrase. As a result of this agreement, the verb gets the complementizer agreement morpheme and also the fronted *wh*-phrase and C share the same noun class number ‘8’. Let us see how Carstens (2005) derives (17b):



<sup>26</sup> The abbreviations used in (17) are: SA = Subject Agreement, CA = Complementizer Agreement, A = Kilega vowel /a/, FV = final vowel of Bantu verbs, and PERF = perfect tense. Arabic numerals denote noun classes.



*What* moves to the edge of CP

In (18a), the object *what* is attracted to the edge of vP after accusative Case is valued. Then, T  $\phi$ -agrees with the fronted *what*. In (18b), the edge feature on T attracts *what* to Spec,TP and [u $\phi$ ] on C agrees with *what*, that is, C-agreement. Finally in (18c), the edge feature on C attracts the *wh*-phrase. Carstens (2005) assumes that Case valuation in Kilega takes place independently of  $\phi$ -valuation, unlike Chomsky 2007, 2008.<sup>27</sup> That is, in (18a), T  $\phi$ -agrees with the fronted *wh*-phrase but Case-agrees with the subject, which remains at Spec,vP.

The predictions in (15) and (16) are false for Kilega *wh*-constructions. Our current system predicts that C never triggers Agree because [uF] cannot be stranded on the edge C. However, in Kilega, C triggers  $\phi$ -agreement. Also, with respect to the other prediction, whereby an element fronted at a phase edge never undergoes  $\phi$ -agreement (as a goal), the *wh*-phrase fronted to the edge of vP does indeed undergo  $\phi$ -agreement with T and also with C in Kilega. If the same feature split as in English occurred in Kilega, [ $\phi$ ] on the *wh*-phrase should already be split off, so that the *wh*-phrase at the edge lacks  $\phi$ -features. That is, [u $\phi$ ] on C and T should never be valued, contrary to fact.

In the next section, we will suggest that feature splitting is, in fact, parameterized between English-type and Kilega-type languages. Also, we will demonstrate that this parameterization predicts *permissible* improper movement in Kilega and *lack of A'-opacity* (i.e., A'-transparency) in agreement patterns, and so further supports our feature-split analysis.

#### 4.3 How to Split Features: The Separation of [ $\phi$ ] and [Case]

As explicated in the last section, our feature-split system has potential problems—namely, the treatment of C-agreement languages such as Kilega. There are at least two issues that seem to be incompatible with our analysis:

- *Problem 1:* The Kilega *wh*-phrase retains  $\phi$ -features at and beyond the edge.
- *Problem 2:* A C head in Kilega has [u $\phi$ ], which triggers C-agreement with a *wh*-phrase, contra Richards's and Chomsky's prediction that valued [uF] can never appear on the edge.

In this section, we consider how to accommodate these incompatibilities into the system proposed in sections 2 and 3.

<sup>27</sup> We will return to the Case-valuation issue in Kilega in section 4.3.1.

4.3.1 Problem 1:  $\phi$ -features at phase-edge positions

Regarding Problem 1, the question is how the system makes it possible that an element moving to a phase edge keeps its  $\phi$ -features. In the previous sections, we claimed that, in the case of C-to-T feature inheritance, T attracts only features that it agreed with and C attracts the rest of the features purely by the edge feature. Since Chomsky 2000, Case valuation has been considered a reflex of  $\phi$ -feature valuation between T and DP. Therefore, features that participated in, or features valued as a result of, Agree with T are [i $\phi$ ] and [uCase] on DP. Given these assumptions, [i $\phi$ ] and [uCase] on DP are attracted to Spec,TP whereas [Q] is attracted to Spec,CP. Recall that feature split was crucial to explain improper-movement phenomena as agreement failure. Now, let us consider what features participate in Agree with T in Kilega. Carstens (2005) suggests that Kilega Case valuation is not tied to  $\phi$ -agreement (contra Chomsky 2001), as exemplified in (19)–(20).<sup>28</sup>

- (19) **Ku**-Lúgushwá **kú**-kili **ku**-á-twag-a **nzogu** maswá.  
 17-Lugushwa 17.SA-be.still 17.SA-A-stampede-FV 10.elephant 6.farm  
 ‘At Lugushwa, elephants are still stampeding over (the) farms.’  
 (Carstens 2005:265)

- (20) a. **Mutu** t-á-ku-sol-ág-á **muku** wénené.  
 1.person NEG-1.SA-PROG-drunk-HAB-FV 6.beer alone  
 ‘A person does not usually drink beer alone.’  
 b. **Maku** ta-má-ku-sol-ág-á **mutu** wéneéné.  
 6.beer NEG-6.SA-PROG-drink-HAB-FV 1.person alone  
 ‘No one usually drinks beer alone.’  
 (Carstens 2005:265)

In (19), the subject is ‘elephants’, whose noun class is marked with ‘10’. This element should be assigned nominative Case by T. However, the verb ‘stampede’ and the auxiliary ‘be still’ on T bear the same noun class markers as the sentence-initial DP ‘Lugushwa’, which is marked with ‘17’. That is, T in this example assigns Case to the subject ‘elephant’ but [u $\phi$ ] on T is valued by the other DP ‘Lugushwa’. Chomsky’s analysis, whereby Case valuation on DP is universally a reflex of  $\phi$ -agreement, incorrectly predicts that T and the subject ‘elephants’ are both marked with ‘10’. The same contrast can be observed in (20). If the subject ‘person’ is placed at Spec,TP as in (20a), T gets the marker ‘1’ from the subject. If the subject stays in the postverbal position as in (20b), the other DP ‘beer’ at Spec,TP assigns its noun class marker ‘6’ to T and the subject ‘person’ keeps the other marker ‘1’. That is, these Kilega data suggest that  $\phi$ -valuation does not necessarily coincide with Case valuation on DP. Based on these kinds of data, Carstens (2005) concludes that Kilega Case valuation takes place independently of  $\phi$ -agreement, unlike in English.

<sup>28</sup> Abbreviations used in (19) and (20): SA = Subject Agreement, A = Kilega vowel /a/, FV = final vowel of Bantu verbs, PERF = perfect tense, NEG = negation, PROG = progressive, and HAB = habitual. Arabic numerals denote noun classes.

Carstens's analysis provides us with a new possibility in feature splitting. Under Chomsky's system of Case valuation, a "single" Agree operation values both  $\phi$ - and Case features. But in Kilega, a single Agree operation involves only  $\phi$ -agreement or only Case valuation. In other words, T triggers two independent Agree operations: one is for [ $\phi$ ] and the other is for [Case].<sup>29</sup> What do these differences predict regarding feature split? In the single Agree operation case, both [Case] and [ $\phi$ ] are involved in Agree at once. Therefore, these two features cannot be separated after Agree, so that they are "fused" in English-type languages. (See section 2.2 for more detail.) As a result, the edge feature on T only has the option to attract this "inseparable" feature set. In Kilega, on the other hand, [ $\phi$ ] and [Case] never combine with each other because Agree applies to them separately. How then does the Kilega edge feature on T decide what features to attract? There seem to be two options, in contrast to English:

- (21) Kilega
- a. Option 1: T attracts both [ $i\phi$ ] and valued [uCase] on DP.
  - b. Option 2: T attracts either [ $i\phi$ ] or valued [uCase] on DP.

The first option is available both in English-type and Kilega-type languages. Although T in Kilega agrees with those two features in a different manner from English, it is obvious that both of the features finally participate in valuation by T. Additionally, Kilega has the second option because of its unique system of Case/ $\phi$ -valuation. As mentioned before, [ $\phi$ ] and [Case] are never amalgamated in Kilega, unlike English, so that the edge feature on T can choose either of the features that it agreed with. If [ $i\phi$ ] is attracted by T, C attracts valued [uCase]. If valued [uCase] is attracted by T, on the other hand, C attracts [ $i\phi$ ]. These three possibilities are logically available for feature splitting. However, it turns out that two of the three options, in fact, cause crash.

For Option 1, this is the same type of feature splitting as in English-type languages. But remember that in Kilega the *wh*-phrase moved to the edge of vP undergoes  $\phi$ -agreement with T and C in the derivation illustrated in (18). Thus, if Option 1 is chosen, [u $\phi$ ] on T or C is never valued, causing crash just the same as in English. This option never converges in Kilega. Similarly, in one of the two possibilities in Option 2, whereby valued [uCase] is attracted by C, the derivation crashes because

<sup>29</sup> One might ask how T triggers Case valuation independently without [uCase]. Under the current system, [uF] on a phase head drives Agree, but if T has [uCase] to execute Case valuation, it is not clear how "valuation" takes place between [uCase] on T and [uCase] on DP because neither of them has a value. (See Epstein & Seely 2006.) How to explain Kilega  $\phi$ -independent Case valuation under the current minimalist approach is beyond the scope of this paper, but for the sake of discussion, we will tentatively assume that the Kilega Case-feature system has something similar to *wh*-phrases in Chomsky 1995b and Grewendorf 2001 in terms of feature distributions, where a *wh*-phrase has both [uWh] and [iQ]. That is, Kilega Case-assigning heads are equipped with [iCase] and [uKase] and DP bears [uCase] and [iKase], so that [uKase] on T triggers Agree with [iKase] on DP and as its consequence, [uCase] on DP is valued independently of  $\phi$ -valuation. Again, this is just a tentative approach to capture Carstens's data and the precise parametric difference between English and Kilega Tense morphosyntax under the current feature-valuation system, so further consideration is required.

the appearance of valued [uF] at a phase edge causes crash at the next phase (as discussed in section 2.1).

On the other hand, the other possibility with Option 2 converges in that a copy at the phase edge has only [iF], namely [i $\phi$ ] (and [iQ]). Under this option, a moving element succeeds in carrying its [i $\phi$ ] along with its [iQ] to the phase edge. When this option is taken, the Kilega derivation can converge. (Notice that a universal Activity Condition would incorrectly predict crash. That is, the Activity Condition incorrectly prevents a higher T from getting its  $\phi$ -features valued by a *wh*-phrase on the edge of CP because the *wh*-phrase lacks Case and so is inactive.) Following Carstens's (2005) analysis of Case valuation as an independent operation, T/C in Kilega induces a slightly different type of feature split from the English type, enabling an element at a phase edge to keep its [i $\phi$ ] rendering it available for  $\phi$ -agreement with a higher probe. On the other hand, the English-type language always chooses Option 1, which makes an element moved to a phase edge lose [i $\phi$ ], causing failure of  $\phi$ -agreement with a higher probe.

#### 4.3.2 Problem 2: [u $\phi$ ] on C

Next, consider Problem 2: a C head in Kilega has [u $\phi$ ], which enters into direct C-agreement with a *wh*-phrase. As we saw in section 2.1, a valued [uF] on a phase head causes crash of the derivation, following Chomsky and Richards. However, C in Kilega directly triggers  $\phi$ -agreement with a *wh*-phrase and they share the same noun class marker. That is, C itself retains valued [u $\phi$ ] and seems not to transfer its  $\phi$ -features to T. (See Ouali 2006 for important discussion regarding possibilities of stranded [uF] at a phase head in Berber.) In this sense, Kilega C-agreement is incompatible with the Chomsky-Richards account. The point to be clarified is what kind of mechanisms allows valued [u $\phi$ ] on C to converge. With respect to this problem, we suggest the following two kinds of edge features:

- (22) Parameter: Two types of edge features
- a. EF: pure edge feature
  - b. EF $_{\phi}$ :  $\phi$ -edge feature

The first type, EF, is “pure” and triggers internal merge without Agree. English-type languages take this type of edge feature as a value of the parameter. The second type, EF $_{\phi}$ , is an edge feature containing [u $\phi$ ]. That is, a  $\phi$ -feature is a part of the edge feature. [u $\phi$ ] within the edge feature triggers  $\phi$ -agreement with a goal equipped with [i $\phi$ ], then the edge feature itself triggers internal merge of the element that agreed with [u $\phi$ ] inside the edge feature. In other words, this type of edge feature triggers  $\phi$ -agreement as well as internal merge. Kilega-type languages choose this type. (See Baker & Collins 2006 for a similar view.)

One might wonder why under the Chomsky-Richards analysis EF can be at a phase head hence on the edge despite the fact that it is uninterpretable. Regarding this issue, Chomsky (2007:11, n. 16) argues:

As an uninterpretable feature, EF cannot reach the interface, so presumably deletion of EF is an automatic part of the operation of Transfer. Note that the same cannot be assumed for the standard uninterpretable features, which can be deleted only when certain structural conditions are satisfied, and will crash the derivation otherwise.

EF is neither phonological nor semantic, so that regardless of it being valued or unvalued it should (but somehow does not) cause crash.<sup>30</sup> That is, EF has to disappear once it is satisfied in order to allow convergence. Given this argument,  $EF_\phi$  in Kilega-type is also automatically deleted by Transfer along with valued  $[u\phi]$ , so that the derivation never crashes. Thus, Kilega  $\phi$ -features, by being embedded as a part of EF, get deleted automatically because EF always does.

Summarizing so far, we suggested that there are two types of edge features: pure EF and  $EF_\phi$ . That is, the edge feature is parameterized and equipped with those two values. English-type languages choose the former, and Kilega-type languages choose the latter. Although EF is  $[uF]$ , it is automatically deleted by the Transfer operation because of its different property from other  $[uF]$ . Therefore,  $EF_\phi$ , as  $[uF]$ , triggers  $\phi$ -agreement with a *wh*-phrase, yet still undergoes automatic Transfer. This parameterization enables us to “solve” Problem 2: a C head in Kilega has valued  $[u\phi]$  under direct C-agreement with a *wh*-phrase. Under this proposal, C in Kilega can trigger  $\phi$ -agreement and bear valued  $[u\phi]$  at the edge, yet can do so without causing crash.<sup>31</sup>

#### 4.3.3 Summary of section 4.3

In section 4.3, we addressed the potential problems involved in C-agreement languages such as Kilega. Table 1 summarizes the claims and identifies two predictions we will discuss in the next section.

The parameters developed in this section enable us to capture the differences between English-type and Kilega-type languages especially regarding C-agreement. The proposed parameters also make strong predictions with respect to both Kilega improper movement and A'-opacity, as illustrated in Table 1. The current system allows an element at a phase edge to bear  $[i\phi]$  in Kilega. That is, improper movement should be “proper” movement in Kilega. Additionally, A'-opacity effects should not be observed—an element at an A'-position (phase edge) itself bears  $\phi$ -features and so should block  $\phi$ -agreement across it. In the next two sections, we argue that these predictions are supported by the data.

**Table 1. Differences between English-type and Kilega-type languages**

	Edge Feature	[Case]/ $[\phi]$ separability	Improper movement	A'-opacity
English	Pure EF	Not separable	Disallowed	Yes
Kilega	$EF_\phi$	Separable	Allowed	No

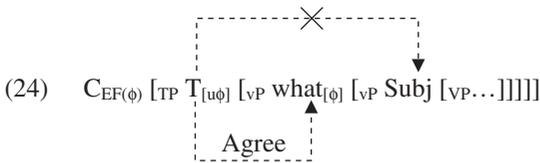
<sup>30</sup> This is the current incarnation of the mysterious EPP (See, e.g., Epstein, Pires & Seely 2005 and Epstein & Seely 2006 and the many references cited for discussion.)

<sup>31</sup> A reviewer notes, and we agree, that embedding  $\phi$ -features within EF is at best a descriptive parameter (if correct and SMT-consistent).

4.4 More Empirical Consequences: A'-Transparency Effects

First, let us examine the predicted suppression of A'-opacity effects in Kilega. In section 3.2, we discussed Rezac's (2003) A'-opacity effects, whereby an element at an A'-position does not block Agree in A-positions, and showed that our feature-split analysis explains this. In English, an element at a phase edge does not bear [iφ], so does not interrupt φ-agreement between T and a subject across it. In contrast, the current system exemplified in section 4 makes it obligatory that an element moved to a phase edge in Kilega retains its [iφ]. This predicts that Kilega does *not* show A'-opacity effects; rather it should show A'-transparency effects, whereby intervening A'-elements *do* block φ-agreement between A-positions. This prediction is confirmed by the following data presented in Carstens 2005:

- (23) \*Bíkí **bi-b-á-kás-il-é** bábo bíkulu mwámí mu-mwílo?  
 8what 8.CA-2.SA-A-give-PERF-FV 2.that 2.woman 1.chef 18-3.village  
 'What did those women give the chief in the village?'  
 (Carstens 2005:237, cited from Kinyalolo 1991)



Example (23) has the same base form as (17) but the verb in (23) contains the morphemes both for subject agreement and for C-agreement. These data indicate that those two agreement morphemes cannot co-occur in a single sentence. Either is possible, but both never occur at once.<sup>32</sup> Why? As illustrated in (24), 'what', which was attracted from the object position, is at the edge of vP intervening between T and the subject 'those women'. But notice that in Kilega [iφ] on DP is attracted by a phase head (v in this case) along with [Q]. That is, the intervening 'what' retains [iφ], unlike in English-type languages. The [uφ] on T then agrees with the closer [iφ], that is, 'what', but not the subject, which rules out (23).

A'-transparency/-opacity effects can be explained by the proposed slight parameterization of feature-splitting operations.

<sup>32</sup> As discussed in den Dikken 2001, however, if the subject were a pronoun in (23), then the verb can undergo both subject agreement and C-agreement. As a reviewer suggests, this may well be due to the possible incorporation of pronominals. (See Baker 1988, Bošković 1997, and discussion of Bošković 1997 in Epstein & Seely 2006:83 for related phenomena concerning pronominal/nonpronominal Case-Agree asymmetries.)

## 4.5 “Proper” Improper Movement

The second prediction made by our analysis is that Kilega should allow improper movement. In the derivation of English-type improper movement,  $[u\phi]$  on upstairs finite T is not valued because a moving element, which is the only potential goal available for T, lacks  $[i\phi]$  by virtue of feature split. (See section 3 for more detail.) In Kilega, on the other hand, moving elements retain  $[i\phi]$  throughout the derivation. That is, the prediction is that Kilega-type feature split renders a moving element available as the goal of  $\phi$ -agreement by a probing upstairs T, and improper movement thought to be universally excluded should be permissible.

Consider again the derivation for Kilega C-agreement illustrated in (18). Movement of ‘what’ occupies the following four positions traditionally (see Chomsky 1981), characterized as follows: Object position (A)  $\rightarrow$  Edge-vP (A')  $\rightarrow$  Spec,TP (A)  $\rightarrow$  Edge-CP (A'). These steps (i.e., A  $\rightarrow$  A'  $\rightarrow$  A) have been traditionally considered improper movement, which is not permissible in English-type languages. But in Kilega, the movement through these positions is allowed and the derivation in (18), in fact, converges.

Another Bantu language, Lusaamia, shows the same kind of C-agreement as Kilega and similarly allows improper movement (Carstens 2008):

- (25) a. Bi-bonekhana Ouma a-bwereo.  
       8.SA-appear Ouma 3.SA-leave  
       ‘It appears that Ouma left.’  
       b. Ouma a-bonekhana a-bwereo.  
       Ouma 3.SA-appear 3.SA-leave  
       ‘Ouma appears as if he left.’

The subject ‘Ouma’ has the noun class marker ‘3’. In (25a), the embedded verb ‘leave’ is marked with ‘3’ but the matrix verb ‘appear’ is marked with ‘8’. This means that ‘Ouma’ stays in the embedded clause and agrees only with the embedded T. In (25b), on the other hand, both of the verbs are marked with ‘3’, which is assigned by the subject ‘Ouma’. That is, ‘Ouma’ first agrees also with the embedded T. Then, it moves to the matrix clause and agrees with the matrix T. The subject movement ends up landing at the matrix Spec,TP. The latter case is the same type of raising as the English data presented in (1c): \**John seems (that) is intelligent*. In Lusaamia, (25b) is grammatical, in contrast to the ungrammaticality in English.<sup>33</sup>

As expected, improper movement is “proper” movement in Bantu languages such as Kilega and Lusaamia. That is, these data suggest that improper movement, which concerns Agree between morphological features under our approach, is in fact parameterized, as is the morphology of these languages. In this sense, our system can be distinguished from previous approaches, where improper movement is universally

<sup>33</sup> We argue elsewhere that the Kilega-/Lusaamia-type derivation occurs even in English, namely in *tough*-constructions, which are notorious for their display of seemingly improper movement. (See Obata & Epstein 2008b.)

excluded, while also indicating a connection between (un)availability of C-agreement and the permissibility of improper movement.<sup>34</sup>

4.6 Remaining Issues: Is Kilega and Lusaamia Wh-Movement A-Movement?

Here, we revisit the issue of the A/A' distinction, which was discussed in section 3.4. The feature-split analysis enables us to recapture A/A'-position types in terms of features of categories. In the English-type derivation, a copy appearing at a phase-edge position lacks [iφ] because of feature split. In the previous section, we suggested: A-positions are categories with φ-features, whereas A'-positions are categories without φ-features. This category/feature-based distinction predicts that phase-edge positions such as Spec,CP bearing φ-features are A-positions in Kilega and Lusaamia. This also entails that *wh*-movement in Kilega and Lusaamia is actually A-movement, in contrast to English-type *wh*-movement construed as A'-movement. With respect to this issue, Vicki Carstens (p.c.) reports that Kilega and Lusaamia informants show reconstruction in *wh*-movement but no weak crossover effects. The following data are from Lusaamia:<sup>35</sup>

(26) Ni esitabu sina khu **biaye**; bidatu si o-para mbwe  
 COP 7.book 7.which LOC **8.his** 8.three 7.wh.agr 2SG-think that  
**buri omwana we esikuli**; a-som-anga?  
**1.of 7.school** 3SG-read-HAB **every 1.child**  
 'Which among his 3 books do you think every student reads?'

(27) **Wina**, yi embwa **eyae**; i-ya-khera?  
**1.who** 1.wh.agr 9.dog **9.his** 9.SA-PRES-love  
 'Who does his dog love?'

In (26), the bound-variable reading is available between 'every student' and 'his' in Lusaamia, like in English. This indicates that the *wh*-phrase is reconstructed to the original position, which 'every student' can bind. In (27), on the other hand, 'his' can receive bound-variable interpretation in Lusaamia, unlike in English. That is, no weak crossover effect is observed in this language. The same behaviors are also reported in Kilega, according to Carstens (p.c.). Based only on such weak-crossover data, it seems premature to say that Kilega and Lusaamia *wh*-movement is A-movement. However, we can at least think that Kilega and Lusaamia *wh*-movement, which bears φ-features, behaves differently from English *wh*-movement, which does not have

<sup>34</sup> Dutch is a language between the English type, which disallows improper movement and does not show C-agreement, and the Kilega type, which allows improper movement and has C-agreement. Dutch shows C-agreement like Kilega but does not allow improper movement. The current system cannot capture this type of language straightforwardly. As for this issue, we will tentatively assume that Dutch also takes the English-type parameter value, so that ungrammaticality of improper movement is explained. With respect to C-agreement, we adopt Zwart's (2006) analysis, where C-agreement in Dutch can be explained without appeal to the Agree operation, and hence is argued not to be an agreement phenomenon in the formal sense.

<sup>35</sup> We are very grateful to Vicki Carstens for providing us with the Lusaamia data.

$\phi$ -features, with respect to traditional A'-movement diagnostics. It might be possible to consider that these differences are attributable to the (un)availability of  $\phi$ -features on *wh*-phrases, although much further research is required to determine the adequacy of our proposal (12) that A-positions are  $\phi$ -bearing categories whereas A'-positions are categories without  $\phi$ -features. (See Obata 2007 regarding the argument/adjunct distinction based on Case features.)

## 5. Summary and Discussion

We presented an account of improper movement based on Chomsky's (2007, 2008) phase-based derivational system. The legitimacy of improper movement is, we claim, parameterized. In some but not all I-languages, it is excluded as a consequence of the feature-split hypothesis and without appeal to the Activity Condition. Feature splitting is parameterized between English-type and Kilega-type languages. The parameters concern feature hierarchies (bundling and separability) of functional features universally present in morphosyntactic feature inventories. We also argued that A'-opacity/transparency effects are parameterized and unified with, hence covary with, improper movement. As a consequence of the proposed analysis, we conjecture that A/A' properties can be recaptured categorially/featurally in terms of the presence or absence of  $\phi$ -features.

However, several problems still remain, necessitating further consideration. One is multiple-Case-assignment phenomena as discussed in Bejar & Massam 1999: In languages such as Hungarian, Case on DP can be "revised" even after the DP undergoes A'-movement:

- (28) *Kiket* mondtad hogy szeretnél ha eljőnnének?  
 who-ACC you-said that you-would-like if came.3PL  
 'Who did you say that you would like it if they came?'

(Bejar & Massam 1999:66, citing Kiss 1985)

*Kiket* 'who-ACC' is extracted out of the subject position of the *if* clause, where it receives nominative Case. On the way to the final landing site, it gets accusative Case from the intermediate verb and the second Case—accusative Case—is always realized. That is, Case is multiply assigned to *kiket*, although this is a construction-specific phenomenon.<sup>36</sup> In our system, such a language is categorized as a Kilega-type, which allows an element moved to an A'-position to still bear  $\phi$ -features (and

<sup>36</sup> Bejar & Massam (1999) categorize English as a language that does not allow multiple Case assignment, unlike Hungarian. On the other hand, Obata & Epstein (2008b) suggest that English also has multiple-Case-assignment phenomena; in particular, in *tough*-constructions. (See Bejar & Massam 1999:n. 4, and also see Chomsky 1981 for relevant discussion.)

so, allowing Case valuation of the DP by the higher v/V under  $\phi$ -agreement).<sup>37</sup> However, Hungarian (contra our expectations) does not show C-agreement. For example, a moved *wh*-object bearing  $\phi$ -features and occupying an outer Spec,vP is expected to (but apparently does not) serve as a possible, goal of a probing C. We have no explanation here for this apparent subparameterization of C-agreement.<sup>38,39</sup>

## References

- Abels, K. 2007. Towards a restrictive theory of (remnant) movement. Ms., University of Tromsø and University College London.
- Adger, D. & G. Ramchand. 2005. Merge and Move: *Wh*-dependencies revisited. *Linguistic Inquiry* 36:161–193.
- Baker, M. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.
- Baker, M. & C. Collins. 2006. Linkers and the internal structure of vP. *Natural Language & Linguistic Theory* 24:307–354.
- Bejar, S. & D. Massam. 1999. Multiple Case checking. *Syntax* 2:65–79.
- Bošković, Ž. 1997. *The syntax of nonfinite complementation: An economy approach*. Cambridge, MA: MIT Press.
- Bošković, Ž. 2007. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry* 38:589–644.
- Carstens, V. 2005. Agree and EPP in Bantu. *Natural Language & Linguistic Theory* 23:219–279.
- Carstens, V. 2008. Raising in Bantu. Ms., University of Missouri, Columbia.

<sup>37</sup> An anonymous reviewer notes a potential problem regarding resumptive pronouns in that an element at an A'-position does not bear  $\phi$ -features because of feature split in our system, so that it fails to capture identity between an element at an A-position and its resumptive pronoun. Although there are several analyses for resumptive pronouns, if those elements at A'-positions are base-generated as Adger & Ramchand (2005) suggest, our system predicts that those elements still have  $\phi$ -features, so that identity between those elements and their resumptive pronouns can be established. Notice that feature split happens as a consequence of Internal Merge by multiple heads. That is, if elements do not undergo Internal Merge but only undergo External Merge (i.e., base-generation),  $\phi$ -features are not split off. We leave this general issue, including predictions regarding A'-opacity of *wh*-phrases externally Merged into Spec,CP, for future research.

<sup>38</sup> Note that Bejar & Massam (1999) propose that movement can strand a Case feature (allowing the now-Caseless mover to receive another Case.) This is a form of feature splitting motivated on independent grounds.

<sup>39</sup> A reviewer suggests that the current system predicts that there are two more kinds of parameter settings in addition to: (i) English-type (= pure EF and Case/ $\phi$  not separable), which has no C-agreement and rules out improper-movement derivations; and (ii) Kilega-type (= EF $_{\phi}$  and Case/ $\phi$  separable), which has C-agreement and allows improper-movement derivations as we listed in Table 1). The other two possibilities would correspond to Hungarian-type languages (i.e., pure EF and Case/ $\phi$  separable), which have no C-agreement and allow improper-movement derivations, and Dutch-type languages (i.e., EF $_{\phi}$  and Case/ $\phi$  not separable), which have C-agreement and exclude improper-movement derivations. The reviewer's idea is that our proposed system can be successfully extended to both Hungarian and Dutch, which we have thus far seen as potential problems. (See fn. 34 regarding Dutch.) Although we need to more carefully analyze those languages, we agree that the Hungarian-type might be captured by extending our system. However, the Dutch-type is not so straightforward. Even if C can trigger Agree by EF $_{\phi}$  in Dutch, if a goal does not have  $\phi$ -features, C-agreement cannot be established. Notice that the option of [Case/ $\phi$  nonseparable] predicts that Dutch  $\phi$ -features are split off like in English, with the result that the Dutch DP cannot keep  $\phi$ -features once it undergoes A'-movement. Hence, this option still predicts that there is no C-agreement. Therefore, contra the reviewer, we believe that the Dutch-type is still problematic for our system. Although we think that these issues are important, due to space limitations we must leave them for future research.

- Chomsky, N. 1964. *Current issues in linguistic theory*. The Hague: Mouton.
- Chomsky, N. 1973. Conditions on transformations. In *A festschrift for Morris Halle*, ed. S. Anderson & P. Kiparsky, 232–286. New York: Holt, Rinehart and Winston.
- Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. 1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.
- Chomsky, N. 1995a. Bare phrase structure. In *Government and binding theory and the Minimalist Program*, ed. G. Webelhuth, 383–439. Cambridge: Blackwell.
- Chomsky, N. 1995b. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, N. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. R. Martin, D. Michaels & J. Uriagereka, 89–155. Cambridge, MA: MIT Press.
- Chomsky, N. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. M. Kenstowicz, 1–52. Cambridge, MA: MIT Press.
- Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry* 35:1–22.
- Chomsky, N. 2007. Approaching UG from below. In *Interfaces + recursion = language?* ed. U. Sauerland & H.-M. Gärtner, 1–29. New York: Mouton de Gruyter.
- Chomsky, N. 2008. On phases. In *Foundational issues in linguistic theory: Essays in honor of Jean-Roger Vergnaud*, ed. R. Freidin, C. P. Otero & M. L. Zubizarreta, 133–166. Cambridge, MA: MIT Press.
- Chomsky, N. & H. Lasnik. 1995. The theory of principles and parameters. In *The Minimalist Program*, N. Chomsky, 13–127. Cambridge, MA: MIT Press.
- den Dikken, M. 2001. “Plurilinguals,” pronouns, and quirky agreement. *The Linguistic Review* 18:19–41.
- Epstein, S. D., E. Groat, R. Kawashima & H. Kitahara. 1998. *A derivational approach to syntactic relations*. Oxford: Oxford University Press.
- Epstein, S. D., H. Kitahara & T. D. Seely. 2008. The value of phonological underspecification in the narrow syntax. Paper presented at the Exploring Crash-Proof Grammars conference, Carson-Newman College, Dandridge, TN (February–March).
- Epstein, S. D., A. Pires & T. D. Seely. 2005. EPP in T: More controversial subjects. *Syntax* 8:65–80.
- Epstein, S. D. & T. D. Seely. 2002. Rule applications as cycles in a level-free syntax. In *Derivation and explanation in the Minimalist Program*, ed. S. D. Epstein & T. D. Seely, 65–89. Cambridge: Blackwell.
- Epstein, S. D. & T. D. Seely. 2006. *Derivations in minimalism*. Cambridge: Cambridge University Press.
- Fukui, N. 1993. A note on improper movement. *The Linguistic Review* 10:111–126.
- Grewendorf, G. 2001. Multiple *wh*-fronting. *Linguistic Inquiry* 32:87–122.
- Heck, F. & G. Müller. 2003. Derivational optimization of *wh*-movement. *Linguistic Analysis* 33:97–148.
- Kinyalolo, K. K. W. 1991. Syntactic dependencies and the SPEC–head agreement hypothesis in Kilega. Ph.D. dissertation, University of California, Los Angeles.
- Kiss, K. 1985. Parasitic gaps and Case. Paper presented at MIT, Cambridge, MA.
- Kitahara, H. 2006. Some notes on the Minimalist Program. In *Minimalist essays*, ed. C. Boeckx, 3–15. Amsterdam: John Benjamins.
- Koopman, H. & D. Sportiche. 1983. Variables and the Bijection Principle. *The Linguistic Review* 2:139–160.
- Lasnik, H. & N. Sobin. 2000. The who/whom puzzle: On the preservation of an archaic feature. *Natural Language & Linguistic Theory* 18:343–371.
- Lasnik, H. & T. Stowell. 1991. Weakest crossover. *Linguistic Inquiry* 22:687–720.
- Matushansky, O. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37:6–109.
- May, R. 1979. Must COMP-to-COMP movement be stipulated? *Linguistic Inquiry* 10:719–725.

- Nevins, A. 2005. Derivations without the Activity Condition. In *Perspectives on phases* (MIT Working Papers in Linguistics 49), ed. M. McGinnis & N. Richards, 283–306. Cambridge, MA: MITWPL.
- Obata, M. 2007. On the formal nature of attraction and the suppression of Superiority effects. In *Proceedings of the 37th annual meeting of the North East Linguistic Society 2*, ed. E. Elfner & M. Walkow, 115–126. Amherst, MA: GLSA Publications.
- Obata, M. 2009. Determining phonetic feature distribution: How to “pronounce” three types of A'-movement. In *Proceedings of the 2nd annual International Conference on East Asian Linguistics* (Simon Fraser University Working Papers in Linguistics 2), ed. J. Brooke, G. Coppola, E. Görgülü, M. Mameni, E. Mileva, S. Morton & A. Rimrott. Vancouver: SFU Linguistics Graduate Student Association. [http://www.sfu.ca/gradlings/wp\\_2.html](http://www.sfu.ca/gradlings/wp_2.html).
- Obata, M. & S. D. Epstein. 2008a. Deducing improper movement from phase-based C-to-T phi transfer: Feature-splitting internal merge. In *Proceedings of West Coast Conference on Formal Linguistics 27*, ed. N. Abner & J. Bishop, 353–360. Somerville, MA: Cascadia Proceedings Project.
- Obata, M. & S. D. Epstein. 2008b. Building (proper) improper movement structures. Paper presented at the Ways of Structure Building conference, University of the Basque Country, Vitoria-Gasteiz, Spain (November). <http://ehutb.ehu.es/es/serial/85.html>.
- Ouali, H. 2006. Unifying agreement relations: A minimalist analysis of Berber. Ph.D. dissertation, University of Michigan, Ann Arbor.
- Rezac, M. 2003. The fine structure of cyclic Agree. *Syntax* 6:156–182.
- Rezac, M. 2006. On *tough*-movement. In *Minimalist essays*, ed. C. Boeckx, 288–325. Amsterdam: John Benjamins.
- Richards, M. 2004. Object shift, scrambling, and symmetrical syntax. Ph.D. dissertation, University of Cambridge.
- Richards, M. 2007. On feature inheritance: An argument from the Phase Impenetrability Condition. *Linguistic Inquiry* 38:563–572.
- Svenonius, P. 2000. Quantifier movement in Icelandic. In *The derivation of VO and OV*, ed. P. Svenonius, 255–292. Amsterdam: John Benjamins.
- Toyoshima, T. 2000. Head-to-Spec movement and dynamic economy. Ph.D. dissertation, Cornell University, Ithaca, NY.
- Williams, E. 2002. *Representation Theory*. Cambridge, MA: MIT Press.
- Zwart, J.-W. 2006. Complementizer agreement and dependency marking typology. *Leiden Working Papers in Linguistics* 3.2:53–72.

Miki Obata  
Mie University  
Faculty of Humanities, Law, and Economics  
1577 Kurimamachiya-cho  
Tsu City, Mie 514-8507  
Japan

Samuel David Epstein  
University of Michigan  
Department of Linguistics  
440 Lorch Hall, 611 Tappan Street  
Ann Arbor, MI 48109-1220  
USA

[obata@human.mie-u.ac.jp](mailto:obata@human.mie-u.ac.jp)

[sepstein@umich.edu](mailto:sepstein@umich.edu)