Deducing Improper Movement from Phase-Based C-to-T Phi Transfer: Feature-Splitting Internal Merge

Miki Obata and Samuel David Epstein
University of Michigan, Ann Arbor

1. Introduction

The main goal of this paper is to present a new 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$_3$ seems [$_{CP}$ <John$_2$> [$_{TP}$ <John$_1$> [$_{VP}$ is intelligent]]].
A $\longleftarrow$ A' $\longleftarrow$ A (disallowed)

Improper movement phenomena provide us with an interesting puzzle regarding learnability: why is it that A-to-A, A'-to-A' and A-to-A' movement are allowed, yet A'-to-A movement is by hypothesis excluded? How do children "learn" this exception? The question: "What is the nature of the mechanism or constraint responsible for the prohibition?" has been a central concern since Chomsky (1973). Although May's (1979) Condition C analysis and Fukui's (1993) Chain Uniformity analyses are both insightful and intriguing[1], the account we will present in this paper is one instead based only on derivational and local computations as exemplified particularly in 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", which we argue is the most natural implementation of Chomsky's phi-feature inheritance system and Richards' (2007) value-transfer simultaneity. In addition, this analysis is empirically supported by and seeks to explain, without stipulation, A'-opacity intervention effects discussed in Rezac (2003). Furthermore, the proposed account enables us to rule out improper movement without appeal to the Activity Condition (see Nevins, 2005 and Bošković (2007), for arguments against the Activity Condition).

The paper is organized as follows: Section 2 amplifies some theoretical background and suggests the feature-split analysis. Section 3 illustrates how feature-split works especially regarding bans on improper movement and also spells out several empirical and theoretical consequences. Section 4 concludes our study.

---

[*] We are very grateful to Klaus Abels, Christopher Becker, Željko Bošković, Cedric Boeckx, Noam Chomsky, Betty McNulty Epstein, Gerardo Fernández-Salgueiro, Catherine Fortin, Samuel Gutmann, Mark Hale, 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.

[1] For other universal prohibitions against "improper movement", see Abels (2007) and Williams (2002). In Obata and Epstein (2008a), we argue that the legitimacy of improper movement is parameterized, suggesting that accounts based on universals, including the fixed architecture of UG, are too restrictive.

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 its predictive content. Since *On Phases*, the treatment of T (and also V) has radically changed: T does not bear phi-features inherently, but rather inherits phi-features from C.[2] 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 non-lexically 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, 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 (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 an unclarity regarding the status of T as a phase head: phases are only CP and vP, but not TP 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 behind 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 timing of transfer and feature valuation. (See Ouali, 2006 for other well-motivated and interesting possibilities concerning feature inheritance.) According to Richards' argument, transfer 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 valued [uF] and [iF] disappears in the eyes of Transfer (See Chomsky, 2001, 2007 and Epstein and Seely, 2002 for analysis and possible problems). This leads to crash of the derivation because the computational system (transfer) fails to remove 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, this operation [= Transfer, MO, SDE] 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 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 [iF]. This entails that these two operations must occur simultaneously for a derivation to converge. Given this argument, Richards 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 [uPhi] on C, if C did NOT transfer its phi to T) are valued [uF], indistinguishable from [iF]. Transfer will not know

[2] We assume that the same feature inheritance system applies to v/V following Chomsky (2007, 2008): V inherits its features from v in the process of the derivation. For the limited scope of this paper, however, we mainly focus on C-to-T inheritance.

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 transfer accomplishes.[3]

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 a convergent derivation. But, what about goals such as DP? DP also bears [uF], namely [uCase], which is valued by T (or V). If a phase head is not allowed to bear valued [uF], the same should be true of elements moved to phase-edge positions, where transfer is also suspended until the domain of the next higher phase.[4] In the next section, we will address this issue and propose a new idea.

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, [uCase] on DP never occurs at Spec-CP in a convergent derivation:

(3) Who do you think bought the book?[5]

a. Embedded CP

$[_{CP}$ C $[_{TP}$ T$_{[uPhi]}$ $[_{vP}$ who$_{[iPhi][uCase][Q]}$ $\boxed{[_{VP} \text{ bought the book}]}$]]]
TRANSFER

b. $[_{CP}$ ***who*** C $[_{TP}$ who T$_{[uPhi]}$ $[_{vP}$ who$_{[iPhi]\cancel{[uCase]}[Q]}$ $[_{VP}$ . . .]]]]

c. $[_{CP}$ ***who*** C $\boxed{[_{TP} \text{ who T}_{[uPhi]} [_{vP} \text{ who}_{[iPhi]\cancel{[uCase]}[Q]}}$ $[_{VP}$ . . .]]]]
TRANSFER

Let us focus only on the derivation of the embedded CP. The embedded C is introduced into the derivation bearing [uPhi] which are then inherited by T. T bearing [uPhi] agrees with the subject DP "who" at Spec-vP bearing [iPhi] as in (3a). As a consequence of this phi-agreement, [uCase] on "who" is valued. The edge feature (EF) 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' 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 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 inheritance deduction, since it is inconsistent with it). In this paper, we will pursue the former possibility and will maintain Richards' (2007) claims. (See Epstein, Kitahara and Seely, 2008 for further discussion regarding the latter possibility.)

What mechanism makes it possible that [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), EF on C and on T each independently attracts the single element "who" occupying edge of vP, which explains suppression of the subject condition (since A-movement does not feed A’-movement). What happens

[3] One might think that languages which have C-agreement are problematic for this view in that C cannot trigger Agree because it lacks [uF]. Obata and Epstein (2008a,b) discuss this issue in detail and suggest that there seem to be two types of languages concerning C-agreement. One is exemplified in certain Bantu languages such as Kilega and Lussamia (See Carstens, 2005 and 2008), where wh-phrases get morphemes from C and improper movement phenomena (A→A’→A) are permissible. The other is exemplified by languages such as Dutch and West Flemish (see Zwart, 2006), where there is apparent C-agreement but improper movement is not allowed. (See Obata and Epstein, 2008a,b for further analysis.)

[4] See Obata (2008) for potential exceptions to this assumption in matrix clauses.

[5] We will ignore the accusative Case assignment to make the discussion simpler.

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

[Q] C [uCase][iPhi] T [vP WHO $\boxed{\text{[iPhi][uCase]}}$ $\boxed{\text{[Q]}}$

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 like two sides of the same coin and should be paired. As in the above configuration, 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 Richard's (deductive) condition is satisfied, the derivation converges. Furthermore, the feature-split analysis is compatible with Chomsky's (1964/1995a) view that non-branching lexical wh-phrases are in fact composed of two distinct morphological feature sets: WH on the one-hand and an indefinite QP "something". Given this view, we naturally hypothesize that WH and "something" each moves to a different place.

One might ask the following two things regarding feature-splitting: [1] What mechanism makes it possible that [uCase]/[iPhi] are attracted by T but not by C?, and [2] Why can a single feature or feature set land at edge/spec positions?.[6] One possible answer to the first question is that T only attracts features it has agreed with. In (4), T phi-agrees with [iPhi] on DP and [uCase] is valued as its reflex. That is, both [uCase] and [iPhi] on DP "participated" in agreement with T. In contrast, [Q] is not involved in this phi-agreement. Therefore, T attracts only [uCase]/[iPhi] which it phi-agreed with while C attracts the rest, namely [Q]. In fact, this view is fully compatible (if not forced) with Chomsky's (2007) 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. (Chomsky, 2007:24)" In other words, an A-position is one that results from attraction by EF under Agree in contrast to A'-movement which is triggered solely by EF. The difference between A and A'-movement has been widely accepted. This way of feature-splitting, whereby T attracts only features which it agreed with, enables us to capture two different types of movement. In this sense, our feature-split analysis is a natural implementation deduced from Chomsky's A/A'-distinction and 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 vs A' in the interface representation, which presumably does not have access to the movement type that created the position "back in" the NS. As for the second question, a single feature or feature set 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^0 (Chomsky, 1995b:396)." In the configuration of (4), the category D 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.[7,8]

In a nutshell, for Richards' deduction based on Chomsky's system to go through, there is a need to

[6] This is an extension of Chomsky's (1995) MOVE-Feature proposal. Also see Toyoshima (2000) and Matushansky (2006) for related ideas.

[7] Hisatsugu Kitahara (p.c.) suggests that feature-splitting may correspond to DP decomposition into D and NP: DP bearing [Q] moves to an edge of CP/vP and NP bearing [phi]/[Case] moves to Spec-TP/VP.

[8] 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 [Phi] and [Case], so that it does not affect the merged syntactic objects. That is, feature-split is compatible with these two conditions.

avoid the appearance of valued [uF] on a phase head (Richard's deduction of inheritance) and, as we note here, there is more generally a prohibition against valued [uF] appearing at ANY edge position. As for a phase head, Chomsky's feature inheritance system forces those features on C to lower to T. Regarding 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. 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 embraced by Condition C and Chain based approaches, and without appeal to the Activity Condition). There are two types of improper movement as follows:

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

These two types of improper movement are distinguished in terms of timing of Case-valuation. 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 current approach (indices on "who" for expository purposes only):

(6) *Who seems will leave?
  a. Embedded vP
    $[_{vP} <\text{who}_{1[Q][Phi][Case]}> [_{VP} \text{ leave}]]$
  b. Embedded CP
    $[_{CP}<\text{who}_{3[Q]}> \text{C } [_{TP}<\text{who}_{2[iPhi][uCase]}> \text{T } [_{vP}<\text{who}_{1[Q][iPhi][uCase]}> [_{VP} \ldots]]]]$
  c. Matrix CP
    $[_{CP} \text{ C}_{[EF]} [_{TP} \text{ T}_{[uPhi]} \text{ seems } [_{CP} <\text{who}_{3[Q]}> [_{TP} \ldots]]]]$
    Embedded TP-Transfer

In (6a), the embedded subject "who" is externally-merged into Spec-vP. In (6b), after T phi-agrees with "who", EF on C and on T each attracts the single element "who" occupying Spec-vP. As mentioned in the last section, T only attracts a feature subset of "who" which it agreed with and C attracts the rest.. That is, Features on "who" are split into the edge of CP and Spec-TP: [uCase]/[iPhi] 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), [uPhi] on the matrix T seeks a Matching Goal with [iPhi]. However, "who" at the edge of the embedded CP has already lost [iPhi] as a consequence of feature-split. In addition, "who" at the embedded Spec-TP is not in the minimal search domain of the matrix T because of the Phase Impenetrability Condition. Therefore, [uPhi] on matrix T is not valued, which causes crash of the derivation.[9] The derivation in (6) is successfully excluded. (See Nevins,

[9] If "who" at the edge of the embedded CP in (6b) bears [phi] by some additional mechanism, [uPhi] on the matrix T is valued. That is, it is predicted that the derivation converges. In fact, we think that this happens in *tough*-constructions. See Obata and Epstein (2008c) for more detail. (cf. Rezac, 2006 for a related previous analysis of *tough*-constructions.)

2005 for discussion of the redundancy between PIC and Activity in such derivations.)

What happens if there is no feature-split? If there is no feature-split, it is predicted that "who" still has [iPhi], putting aside Richards' value-transfer simultaneity issue. Therefore, in the configuration of (6c), the matrix T can phi-agree with “who”. The derivation converges, yielding the wrong prediction. What mechanism rules out this derivation if our feature-splitting is not assumed? A separate principle, namely the Activity Condition, needs to be stipulated, where the computational system cannot see elements without [uF]. Even if "who" at the edge of the embedded CP has [iPhi], there is no [uF] on it. Therefore, the Activity Condition prohibits the probe (i.e. the matrix T) from agreeing with the inert "who". As a result, [uPhi] on the matrix T is not valued, which causes crash. If feature splitting does not exist, the Activity Condition is crucial for excluding improper movement phenomena. But the Activity Condition is a stipulation, that is, “who” in embedded Spec-CP bearing [iPhi], 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 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.[10]

In this section, we have 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.

3.2. A'-Opacity Effects

As discussed in the last section, our idea of feature-splitting makes it possible to exclude improper movement without appeal to the Activity Condition. The crucial property of feature-splitting is that DP loses [iPhi] once it undergoes A'-movement to an edge-position. DP lacking [iPhi] fails to value [uPhi] on an upstairs T, causing crash. This is what happens in the improper movement derivation. In addition, the idea of no [Phi] at the edge of a phase enables us to capture so-called "A'-opacity effects" discussed in Rezac (2003): elements bearing phi-features at A'-positions do not block Agree occurring between A-positions. As an illustration, consider the following data from Icelandic.

(7) Strákarnir$_2$ höfðu [**engu grjóti**]$_1$ [$_{vP}$ *t*$_2$ [$_{VP}$ hent ***t*$_1$** í bílana.]]
The boys had no rock thrown in the cars
"The boys had thrown no rocks at the cars."
(See Svenonius, 2000)

In (7), the derived position of "engu grjóti" is between T and the in-situ position of the subject "strákarnir (t$_2$)" with phi-features. However, "engu grjóti" is (somehow) invisible as an intervener to phi-agreement between T and the subject (i.e. nominative assignment). Why does this intervening DP not induce intervention effects? The same scenario, in fact, applies to English:

(8) What cars does John buy?
([$_{CP}$ what cars does [$_{TP}$ John T [$_{vP}$ **<what cars>** [$_{vP}$ John [$_{VP}$ buy <what cars>]]]])
(pl) (sing)

When T phi-agrees with the subject "John", the intervening DP "what cars" bearing "plural" can be skipped. As in the Icelandic data, elements at A'-positions do not function as interveners of T-agreement in (8).[11] Why do A’-interveners not block A-agreement relations? If the A’-position bears [phi] and Minimal Search --a third factor notion (Chomsky 2005) by hypothesis-- is correct, then A’ positions hosting phi features SHOULD block phi-agreement. Any departure from this maximally

[10] See Nevins (2005) and Bošković (2007) for attempts to eliminate the Activity Condition.

[11] Our system enables us to recapture A/A'-distinctions in terms of features inside categories. That is, A/A'-position types are no longer stipulated under this view, but instead, presence/absence of [Phi] on categories distinguishes position-types featurally. See Obata and Epstein (2008b) for more detail.

simple assumption is potentially stipulative.

With respect to this point, Chomsky (2001) suggests: If the fronted “what car” in (8) undergoes further movement to the edge of CP, phi-agreement between T and the subject “John” is, as a result, viable ignoring the intervener-copy of “who’ .[12] More precisely, the analysis refers to phonological features on “what car”, and says: if the intervening element has no phonological features, T-agreement is not blocked. However, it is not clear to us why phonological features care about phi-agreement applying in the NS. Also, the analysis seems not to be viable without look-ahead. That is, it is not clear how the computational system “knows” that “what car” moves further (given copy theory) or stays in-situ at the NS-internal time of T agreement with the subject, an operation which precedes the determination of which parts of the wh-chain will be +/-phonological (nor is it clear to us why Agree intervention effects should be sensitive to the presence vs. absence of phonological features at all). Our system of feature-splitting straightforwardly explains these phenomena (without look-ahead, or dependence upon phonological features) by hypothesizing that elements attracted by EF 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. The data in (7)-(8) further confirms our feature-split analysis.

4. Conclusion

In this paper, we proposed the feature-split analysis, which enables us to account for improper movement locally and featurally (without unbounded representational conditions/relations such as the Condition C approach of May (1979) or the Chain uniformity approach in Fukui (1993). Moreover our account need not appeal to Activity, itself an arguably undesirable condition. We also emphasized that the proposed analysis is a very natural way to implement the current system of transfer and feature-valuation, so that the idea seems to be conceptually well-grounded. In Obata and Epstein (2008b), we further explore this idea and claim that feature-splitting can be parameterized as improper movement is, e.g. in languages such as Kilega and Lussamia reported by Carstens (2005) and (2008). Moreover we argue that English *tough*-constructions, long noted for their enigmatic display of improper movement relations, are “simply” constructions in which English is “acting” no different from e.g. Kilega or Lussamia in allowing an upstairs T to agree with an (unsplit) Phi-bearing wh-phrase at the edge.

References

Abels, Klaus. 2007. Towards a Restrictive Theory of (Remnant) Movement. Ms. University of Tromso and University College London.

Bośković, 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 and Linguistic Theory*23: 219-279.

Carstens, V. 2008. Raising in Bantu. Ms. University of Missouri.

Chomsky, N. 1964. *Current issues in linguistic theory*. The Hague: Mouton and Co.

Chomsky, N. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*, ed. S. Anderson and P. Kiparsky, 232-286. Holt, Reinhart and Winston, New York.

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 and 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.

[12] See Kitahara (2006) for more detailed discussion regarding this issue.

Chomsky, N. 2007. Approaching UG from Below. In *Interfaces + Recursion = Language?* ed. U. Sauerland and H-M Gärtner, 1-29. 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, and M. L. Zubizarreta, 133-166. Cambridge, Mass.: MIT Press.

Epstein, S. D., H. Kitahara and T. D. Seely. 2008. Uninterpretable Features: What are they and what do they do?. Ms. University of Michigan, Keio University and Eastern Michigan University.

Epstein, S. D. and 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 and T. D. Seely, 65-89. Blackwell Publishing.

Epstein, S. D. and T. D. Seely, 2006. *Derivations in Minimalism*. Cambridge University Press: Cambridge.

Fukui, N. 1993. A note on improper movement. *The Linguistic Review* 10: 111-126.

Kitahara, H. 2006. Some Notes on the Minimalist Program. In *Minimalist Essays*. ed. C. Boeckx, 3-15. Amsterdam: John Benjamins.

Lasnik, H. and T. Stowell. 1991. Weakest Crossover. *Linguistic Inquiry* 22: 687-720.

Matushansky, O. 2006. Head-movement in linguistic theory. *Linguistic Inquiry* 37: 69-109.

May, R. 1979. Must COMP-to-COMP movement be stipulated? *Linguistic Inquiry* 10: 719-725.

Nevins, A. 2005. Derivations without the Activity Condition. *MIT Working Papers in Linguistics* 49: 283-306.

Obata, M. 2008. Determining Phonetic Feature Distribution: How to "Pronounce" Three Types of A'-Movement. Ms. University of Michigan, Ann Arbor.

Obata, M. and S. D. Epstein (2008a) Phasing Out Improper Movement as Featural Crash. Paper presented at 31th Generative Linguistics in the Old World. Newcastle University, Newcastle, UK.

Obata, M. and S. D. Epstein (2008b) On the (Possible) Elimination of A/A'-Positions: Improper Movement and Intervention Effects. Ms. University of Michigan, Ann Arbor.

Obata, M. and S. D. Epstein (2008c) (Proper) Improper Movement is Tough, We AGREE. Ms. University of Michigan, Ann Arbor.

Ouali, H. 2006. *Unifying Agreement Relations: A Minimalist Analysis of Berber*. Ph.D. dissertation, University of Michigan.

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. 2007. On feature inheritance: An argument from the phase impenetrability condition. *Linguistic Inquiry* 38: 563-572.

Svenonius, 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.

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.

Proceedings of the 27th West Coast Conference on Formal Linguistics

edited by Natasha Abner and Jason Bishop

Cascadilla Proceedings Project Somerville, MA 2008

Copyright information

ISBN 978-1-57473-428-7 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Obata, Miki and Samuel David Epstein. 2008. Deducing Improper Movement from Phase-Based C-to-T Phi Transfer: Feature-Splitting Internal Merge. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*, ed. Natasha Abner and Jason Bishop, 353-360. Somerville, MA: Cascadilla Proceedings Project.

or:

Obata, Miki and Samuel David Epstein. 2008. Deducing Improper Movement from Phase-Based C-to-T Phi Transfer: Feature-Splitting Internal Merge. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*, ed. Natasha Abner and Jason Bishop, 353-360. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1850.