The A-/A'-distinction as an epiphenomenon

Ken Safir, Rutgers University*

1.0 Introduction

One of the central ideas of the 1970's and 1980's was that conspiracies of principles generate constructions as epiphenomena, such that even broad surface generalizations and their apparent exceptions could, in the best case, be deduced from principles that look nothing like the generalizations that emerge. Conspiracies of this sort have had an enduring influence in the design of linguistic theory. Phrase structure rules and transformational rules were reduced first to X'-theory and Move α, respectively, and then Chomsky (2004) introduced the idea that Merge encompasses both forms of structure-building distinguished only by the choice of terms in the Merge relation. Insofar as the most parsimonious theory of exclusively syntactic relations would rely on nothing more than Merge to generate structure, the rich generalizations about construction types that have been explored for the last 50 years must now emerge from the way the geometry generated by Merge interfaces with other components, at least in the best case. From this perspective, the persistence of the A- vs. A'-distinction cannot have any independent status in the theory (see, e.g., Chomsky, 2004:125 fn.30). It is the goal of this paper to demonstrate that the A-/A'-distinction is an epiphenomenon that emerges from independently necessary properties of Merge and the interpretive components. If the A-/A'-distinction is to be truly derived rather than stipulated, then

A. the characteristic that distinguishes the two classes of structures must emerge from a conspiracy of independently motivated principles and
B. the distinction should explain why the contrasts between A-/A'-constructions are precisely those contrasts and not other contrasts.

The key theoretical proposal that underlies my account is that a slight independently motivated alteration in the terms available to Merge results in a crucial difference between the two construction types. The analytic idea is that once a wh-phrase DP moves, it must be altered on the way to its landing site or else it will interfere with normal Case and Agreement relations (as first proposed by Rezac, 2003). An optional instance of Merge, late attachment of a quantificational or prepositional head to the moved DP, 'insulates' that DP from Case and agreement, but alters its ability to antecede other DPs, with wide consequences for what an insulated DP can antecede and or license. It is argued that the possibility of late attachment, as well as other operations that violate Chomsky's (1995) 'Extension' is independently necessary and that a slight revision of Extension permits them, so the possibility to Merge an insulator is no addition to the theory, but rather is a consequence of Merge. Although insulation is optional, its timing and distribution is limited by interface requirements. The resulting distribution of insulated and uninsulated expressions will explain what is crucially different about A- vs. A'-constructions and why the contrasts between A and A'-structures differ in just the ways they do and not in other ways.

The paper is organized as follows: Section 2 outlines the empirical effects that divide A- vs. A'-constructions. Section 3 introduces the Peak Novelty Constraint, which replaces Chomsky's (1995) Extension as the factor determining what can be a term in the Merge relation. The theoretical and analytic consequences of this revision, in particular the potential

* The first version of this paper was presented in 2010 in Beijing at the University and Culture as part of GLOW in Asia 8. Over the many years this project has gestated, quite a number of conversations were of particular importance, including discussions with Mark Baker, Vicki Carstens, Carlo Linares, Jim McCloskey, Tom McFadden and audiences at Queen Mary University of London, Zentrum für Sprachwissenschaft in Berlin, and several different audiences at Rutgers University.

for late merger of an insulating head, are explored in the remainder of section 3, including consequences for A'-opacity and pied piping. Antecedent relations affected by insulation are used to derive A-/A'-contrasts in anaphor binding, bound variable interpretation and parasitic gap licensing in section 4. Asymmetries in reconstruction are addressed and derived in section 5. Scrambling and languages where IMB does not hold are discussed in section 6 and 7, respectively. Section 8 briefly compares the insulation account with competing theories and then I conclude.

2.0 The A-A'-distinction in practice

Over the years, linguists appeal to the A-/A'-distinction, first so-designated by Chomsky (1981:184) to distinguish construction types that contrast in the ways described in Table 1.

Table 1: ***The A-/A'- contrasts***

	1. Case can be assigned to the landing site	2. Can agree with T in landing site	3. By-passes inter-vening subjects	4. Allows pied-piping	5. Landing site can bind anaphors	6. Licenses parasitic gaps	7. Can induce weak crossover	8. Allows anti-recon-struction of comple-ments
A'	-/%-	-	+	+	-	+	+	-
A	+	+	-	-	+	-	-	+

Examples (1-8) illustrate these contrasts as they are found in English.

1a)*I don't know whom arrested John.
b) He was arrested.
2a) How many boys is/*are it clear that Mary likes?
b) The boys happen(*s) to be guilty.
3a) **The accountant* seems that it is beginning *t* to find fraud.
b) *What sort of fraud* is the accountant beginning to find *t*?
4a) About whom did John speak?
b) *About Mary was spoken (compare: Mary was spoken about)
5a) **Which boys* did *each other* trust?
b) *The boys* seem to *each other t* to be greedy.
6a) *Who* did John trust *t* before he spoke to *pg*?
b) **Mary* seemed *t* to be happy before John spoke to *pg*?
7a)?**Who* does *his* mother love?
b) *Everyone* seems to *his* mother *t* to be a good boy.
8a) Whose account of *his* arrest does *every prisoner* reject [whose account of *his* arrest]?
b) A policeman's account of *his* arrest seems to *every prisoner*
[[a policeman's account of *his* arrest] to be suspect].
c)*Which attack on Hillary's integrity does she consider to be unfair?
d) The attack on Hillary's integrity seemed to her to be unfair?

The first pair of examples shows that Case can be assigned to the landing site of A-movement (though A-movement is not always to a Case-marked position) but except under special circumstances, wh-movement does not show the Case of its landing site.[1] The contrast in

[1] Relative clauses in some languages show case-attraction, whereby the wh-phrase matches the case of the modified nominal rather than the position it is extracted from, as discussed for Bavarian by Bayer (1984). Any account of the Case property of the A-/A-distinction will require a compatible theory of case-attraction, though that is not attempted here.

(2a,b) shows that raised subjects can agree in their landing sites for subject-verb agreement, but wh-phrases do not (but see section 7). A-movements can never skip a subject position, even an expletive one, but A'-movements can, as illustrated in (3a,b). The wh-question in (4a) permits pied-piping whereas passive does not. Examples (5a,b) shows that a subject moved by A-movement, raising in this instance, can bind an anaphor, but a wh-moved object cannot. Parasitic gaps are supported by wh-movement, as illustrated in (6a), where the direct object trace and the prepositional object parasitic gap can both be identified as *who*, but raising, once again, does not support a comparable parasitic gap in (6b). Wh-question movement induces weak crossover, where the pronoun *his* cannot be bound by *who* in (7a), but in a subject-raising structure like (7b), the raised and quantified subject *everyone* can bind the experiencer complement of *seem*. Finally, although the literature is not unanimous, it is generally assumed that A'-movement permits reconstruction for bound pronoun anaphora in both A- and A'-constructions (see, e.g., Sauerland, 2008), as indicated in (8a,b), where the reconstructed reading of (8b) can be paraphrased as one where *every prisoner* believes that a policeman's account of *his* arrest is suspect. This should be possible if *his* embedded in the displaced operator acts as if it is in the scope of *every prisoner* (that is, if it 'permits reconstruction', allowing the contents of the operator to be evaluated in its pre-movement position). However, complement clauses must reconstruct for A'-movement, while A-movement does not seem to require it, as illustrated by the purported Principle C effect in (8c) as opposed to (8d).

Another restriction that is often invoked as part of the difference between A-movement and A'-movement is that movement from an A'-position to an A-position is usually deemed 'improper'.

9) ***Improper Movement Ban* (IMB)**: A'-movement of a constituent X cannot be followed by movement of X to an A-position.

On this syntactic analysis, both (10a) and (10b) should constitute improper movement and be banned, because Spec CP is an A'-position where there has been wh-movement past an overt subject (which A-movement cannot do), and then the wh-phrase has landed in the starred and bolded (trace) subject position, as shown by verb agreement, before continuing on to the matrix clause. It is usually assumed that *tough*-constructions cannot be generated by improper movement as in (9b), even though the right result seems to be achieved.

10a)*[$_{TP}$*How many people* did he say [$_{CP}$ *t* [$_{IP}$ ***t*** are unclear
[$_{CP}$ *t* (that) [Mary spoke to *t*]]]]]

b) [$_{TP}$*How many people* did he say [$_{CP}$ *t* [$_{IP}$ ***t*** are tough [$_{CP}$ *t* for [Mary to speak to *t*]]]

I will assume that the correct theory should derive the IMB in the languages where it is in force, but I will return to the IMB in section 7 where languages improper movement is allowed. Finally, there are other A-/A'-differences that are more language particular, and I will not explore these here (e.g., there are contexts where P can be stranded by wh-movement but not by passive in English).

In order to capture the A-/A'-distinction, several ways of distinguishing the landing site of A- vs. A'-movement have been proposed, or distinctions between the triggers for such movements have been proposed. The following quotes are excerpted from Obata and Epstein (2010: 11), who review the history of this distinction and propose a new one.

- 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 h-position. The position of subject may or may not be a h-position, depending on properties of the associated VP. Complements of X¢ are always h-positions.…An A'-position is that of an adjunct of one sort or another.

- Chomsky (and 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].…
- Chomsky (2007:24): A-movement is IM (internal merge) contingent on probe by uninterpretable inflectional features, while A'-movement is IM driven by EF [Edge Features].

None of these distinctions explain why all the distinctions between the two classes of constructions are just the ones that they are, although there has long been a general theoretical intuition that A-movement is driven by the need to satisfy Case and agreement and A'-movement is driven by something else, often scope, or, in the case of Chomsky (2007), perhaps information structure. In a very interesting paper, Obata and Epstein (2010) suggest that some of the A-/A'-distinctions follow from feature-splitting to satisfy different triggers (also deriving A'-Opacity, see below): EFs like wh-features, on the one hand, and phi-features, on the other, are attracted and neutralized separately. Althouh they claim to derive the IMB they still do not explain why WCO and parasitic gaps are associated with the A-/A'-distinction, and why movement driven by EFs is different in this way and not some other way. To meet the explanatory burden established at the beginning of this essay, not only must the factor that makes the A-A'-distinction be independently motivated, but the distinctions in Table 1 should all follow from it.

3.0 Insulation and A'-Opacity

The inspiration for the proposal defended here originates in Rezac (2003), where it is proposed to account for the analytic problem that arises when A'-movement proceeds by phases and ought to interfere with A-movement, if nothing is said. Rezac observes that extraction of a wh-argument from vP by adjunction to vP, which permits subsequent movement to see the wh-phrase in the vP edge, puts that wh-phrase in a position where it c-commands Spec vP, which is the position of the external argument (EA). The EA is destined to move to Spec TP to satisfy EPP or to get Case, depending on the theory. However, since the adjoined wh-DP is more local to probing T than Spec vP, wh-DP adjoined to vP should agree with T, be assigned Nominative Case, and be attracted to Spec TP to fulfill EPP instead of the contents of Spec vP. Instead, it is the EA that is attracted by T, as if the vP-adjoined wh-phrase were invisible (opaque) to the trigger for A-movement. For example, in a sentence with object extraction like *Who did John praise?*, with its schematized derivation in (11), T can 'see' the wh-direct object (wh-DO) adjoined to vP as the closest DP, then T must attract wh-DO and not EA, as in (11d), rather than the desired outcome, which is (11d').

11a) $[_{vP}$ EA [v [V [wh-DO]]]]
b) $[_{vP}$ [wh-DO] $[_{vP}$ EA [v [V [wh-DO]]]]]
c) [T $[_{vP}$ [wh-DO] $[_{vP}$ EA [v [V [wh-DO]]]]]]
d) *$[_{TP}$ [wh-DO] [T $[_{vP}$ [wh-DO] $[_{vP}$ EA [v [V [wh-DO]]]]]]]
d') $[_{TP}$ EA [T $[_{vP}$ [wh-DO] $[_{vP}$ EA [v [V wh-DO]]]]]]]

Presumably, (11d) will be ruled out for a number of reasons, namely, [wh-DO] gets Nom and Acc Case (double Case assignment is plausibly * in general), EA gets no Case, and T agrees with [wh-DO], or else fails to agree with wh-DO (perhaps) because it cannot Agree with an Accusative DP. However, it is not clear that syntax has the option to prefer the derivation in (11d').

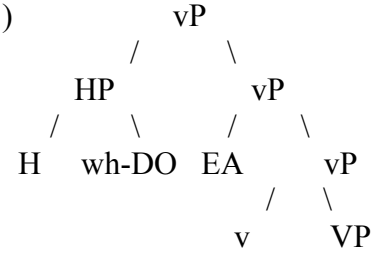
Rezac's strategy is to find a way to make the wh-DO opaque to T. This can be achieved, he proposes, by merging a head (H) to [wh-DO] after it adjoins to vP. The merger of this head intervenes between T and the wh-DO and thereby insures that the intermediate positions of A'-movement are opaque to probing by T for agreement and Case assignment.

12a) $[_{vP}$ EA [v [V [wh-DO]]]]

b) [$_{vP}$ [$_{DP}$ wh-DO] [$_{vP}$ EA [v [V [$_{DP}$ wh-DO]]]]]
c) [$_{vP}$ [$_{HP}$ H [$_{DP}$ wh-DO]] [$_{vP}$ EA [v [V [$_{DP}$ wh-DO]]]]] ***insulation - late head attachment***
d) [T [$_{vP}$ [$_{HP}$ H [$_{DP}$ wh-DO]] [$_{vP}$ EA [v [V [$_{DP}$ wh-DO]]]]]]
e) [$_{TP}$ EA [T [$_{vP}$ [$_{HP}$ H [$_{DP}$ wh-DO]] [$_{vP}$ EA [v [V [$_{DP}$ wh-DO]]]]]]]
f) [$_{CP}$ [$_{HP}$ H [$_{DP}$ wh-DO]] [C [$_{TP}$ EA [T
[$_{vP}$ [$_{HP}$ H [$_{DP}$ wh-DO]] [$_{vP}$ EA [v [V [$_{DP}$ wh-DO]]]]]]]]]

The unorthodox move is (12c), where a head is merged to the moved element rather than to the undominated node of the tree. As illustrated in (13), where (12c) is diagrammed, the introduction of H by Merge is a violation of Chomsky's (1995) Extension Condition (which I formulate as in (14)), because H is not merged to the undominated node of the tree (which is the highest vP), but is merged instead lower than that.

13)



14) **Extension Condition**: Only merge to the undominated node.

I call Rezac's hypothesized head merger to moved wh-DPs ***insulation***, I provide more motivation for insulation and then I show that it is the key to the A-/A'-distinction.

The point of A'-opacity is that Case and agreement do not see an intervening insulated wh-phrase, but regular DPs are indeed visible to agreeing or Case-assigning heads. Thus the reason that A'-moved elements do not get Case in their landing site positions is that they are not DPs after insulation. DPs that can move without insulation, however, are indeed visible to Case and agreement, so the locality conditions on those relations apply. Thus there is (normally) no uninsulated movement across an intervening subject, as uninsulated movement to a position between T and its agreement partner would leave the subject without Case, while the direct object might receive more than one Case (which I assume is *) and it is unclear if Agree has a goal it can agree with. This is not surprising, since the ability of so-called A'-movement to move past a subject and the inability of so-called A-movement to do so are just different sides of the A'-opacity coin. If insulation is a cost-free option that can apply to any sort of constituent and if the cases where it does apply (and its timing in the derivation) are predictable from general principles, then the first three properties on Table 1 now follow from the option to insulate.

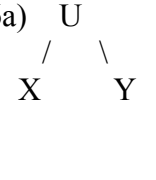
3.1 Insulation is cost-free: The Peak Novelty Condition vs. Extension

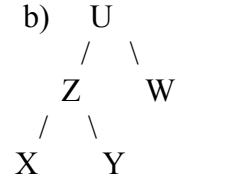
Although insulation violates Chomsky's Extension Condition, it is only one of a variety of frequently-appealed-to operations that violate Extension. Extension does not permit late attachment, which is appealed to by many to account for reconstruction effects (Lebeaux, 1990, Chomsky, 1995, Sauerland, 1998, Safir, 1999, Fox and Nissenbaum, 1999, Bhatt and Pancheva, 2004, Takahashi and Hulsey, 2009), tucking in (Richards, 1999), and head movement by adjunction of a head to a head (e.g., Baker,1988, although challenged, but see Roberts, 2010, and Safir and Bassene, 2015, for a defense of head-to-head adjunction), to which I return later in this section. One of the motivations for Extension is that it blocks counter-cyclic movement and in fact all movement downward, since the result of such internal Merge operations would not create a new undominated node. All of these exceptions just listed, however, merge to nodes just below the top one. If we loosen what counts as the top of the tree to which Merge applies, then the ban on counter-cyclic movement can still be

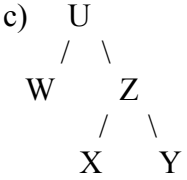
preserved while allowing the set of proposals cited above that violate Chomsky's 1995 version of Extension. I revise Extension as in Safir (2010) (renamed here):

15) **Peak Novelty Condition** (PNC): After every instance of Merge, M_i, the undominated node U of the resulting structure immediately dominates a node that U did not immediately dominate before M_i.

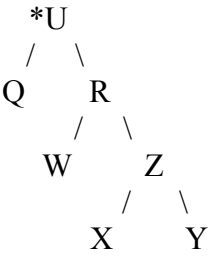
The PNC permits Merge to form all of the trees in (16) where 'X' is merged to the structure by M_i.[2]

16a) 

b) 

c) 

When the undominated node is new[3] (the conventional case) after Merge of X (16a), PNC is satisfied. When X is adjoined just below the undominated node (16b,c), which I describe henceforth as 'penultimate Merge', then the node created by that instance of Merge (Z) is new, hence PNC is satisfied. If Merge involves a term not dominated by U (external Merge), then (16b) and (16c) model late attachment (aka 'late merger') and insulation, respectively. When internal Merge has applied, that is when X has a copy in W in (16b), or where X has a copy in Y, as in (16c), then the structure as presented models head-to-head adjunction in (16b) and tucking in (16c). However, Merge cannot apply to X as in (17).

17) 

Merge of X to Z does not change R, so U does not dominate a novel node and the PNC is violated.

PNC now permits head movement by adjunction to a higher head, as illustrated schematically in (16b), it permits tucking in, illustrated in (16c), as proposed by Richards (1999) to account for superiority effects (but see also Safir and Bassene, forthcoming, who apply tucking in to clitic movement), and it permits late attachment, which is proposed to account for anti-reconstruction effects, as in (16c).

Thus insulation is just another instantiation of penultimate Merge, just like head to head adjunction, late attachment, and tucking in, which have been permitted in practice without formal justification for the operation involved (particularly with respect to late attachment). In fact, ***once the PNC permits penultimate merge, penultimate Merge of a head is expected, i.e., it would take a stipulation to rule out insulation as an instance of Merge.*** In subsequent sections, I will show that the distribution and timing of insulation also accounts for contrasts 4-8 in Table 1.

Notice that we could stipulate that all A'-movements are insulated in order to make the distinction A and A'-constructions, e.g., we could stipulate that movements triggered by EFs require insulation and that would be enough to insure that the A'-opacity effect would

[2] I assume there is actually no left vs. right in syntax (i.e., linearity is introduced only at the morphological/phonological interface), and if so, (16b,c) are not really distinct in syntax, but I are provided for convenience of presentation only.

[3] The 'newness' of what U immediately dominates does not correspond to a new syntactic category, just a node that was not in structure before the instance of Merge in question. In other words, I abstract away from the label of the new node.

arise as a consequence (as is proposed by Rezac). However, even if insulation accounts for all the distinctions in Table 1 (and I will show that it does), the distinction is still not explained if the feature that determines the difference does not itself follow from general principles. I have suggested an explanatory path for the occurrence of insulation – if insulation does not apply to a DP adjoined to vP, then the derivation will crash, i.e., be filtered out, either at the PF interface or at the semantic interface (or both) and if so, it is not necessary to stipulate that insulation must take place for A' movement. In the next subsection explores some of the key assumptions of this approach.

3.2 Free Merge and the 'motivation' for movement

The distinction between EFs and other features matters for movement if movement (a term used here interchangeably for internal Merge) is feature-driven, as in most versions of minimalist theorizing until very recently. Following Chomsky (2013) and Safir (2010), I assume that movement (I-Merge) is not triggered by features, but is completely optional. The result of the operations that generate syntactic trees must, however, meet interface conditions or the derivation will crash. In this respect, this is a return to the filtering model of the Move alpha era (early 1980s), when it was assumed that conspiracies between principles and parameters, particularly those enforced at the semantic and phonological interfaces, determined whether a sentence is well-formed and what it could mean. The abandonment of triggers in the syntax means that there is no 'activation' condition and that Agree does not condition Merge. In other words, Merge is free. However, if constituents need to enter into local relationships in order to receive properties or features or to have certain interpretations, then the failure to form these relationships will be apparent at the interfaces and either certain interpretations will not be possible, or certain morphological outputs will not be possible (i.e., the derivation will crash).

The elimination of triggers for movement removes major complicating factors on the Merge operation, which now is only about how one constituent combines with another. Two motivations for triggers that used to be evoked are no longer relevant. First, the notion that movement is more expensive than external Merge disappears if Chomsky (2004) is right that movement and external Merge are the same operation. The view that triggered movement could render derivations fully deterministic (e.g. crash-proof, as in Frampton and Gutmann, 2002) has generally been set aside as a design feature (rather than refuted), as it is not conceptually necessary to explain competence. Although I do not assume that any Agree relation conditions movement, the Agree relation is still assumed here and is still limited by phases, just as all operations in syntax are. Since Agree has always been motivated independently from movement (e.g., in Chomsky, 2000, 2001), removing Agree as a trigger for Merge certainly does not add to theoretical complexity.

Moreover, the elimination of triggers also removes the opportunity to characterize the A-/A'-distinction as a differentiation based on triggers, and in that respect removes descriptive power. Output conditions will still act as *indirect triggers* for movement, however, so the elimination of triggers in syntax does not reduce the possibilities for stipulations that restrict outputs. If features are not interpretable at the interfaces, they are presumably invisible (not poisonous, see Safir, 2010), but if features cannot be interpreted because they are unvalued, and internal Merge creates the conditions where Agree can value features, then unvalued features are indirect triggers for movement. EPP, which requires that the Spec TP position must be filled (however it is formulated) is such an output condition and it is a key indirect trigger for (what usually turn out to be) A-movements, thus motivating movement in this theory as well. It is important to keep in mind, however, that if my attempt to eliminate the A-A'-distinction results in a surface filter that distinguishes the two kinds of constructions, then my explanatory project will have failed.

The requirement that movement be triggered has already been noted as problematic for A'-movement insofar as movement to intermediate vP-adjoined positions must be triggered by 'spurious' features assigned to v (see McCloskey, 2002, for a defense of spurious features in the C phase edge, but there is no attempt there to justify spurious features on vP). In other words, spurious features are necessary in intermediate positions only to facilitate long distance movement, and so the elimination of this theory-internal problem is welcome. From the present perspective, if constituents do not reach the landing sites where morphological or semantic conditions can be met, then movement cannot stop until the destination is reached, that is, no intermediate trigger for movement is necessary (but see Chomsky, 2013, where it is argued that the failure of labeling assures that movement must take place in the next phase – in other words, he proposes an independent account that permits spurious features to be eliminated).

The indirect trigger for many so-called A'-movements is that a constituent must be in a scopal relation to the clause or must be in a local relation to a head that determines how it should be interpreted with respect to the complement of that head. It is often assumed that movement of wh-question phrases is movement to a scopal position, and where wh-phrases are in situ, abstract movement is often posited to achieve the same result at the semantic interface. An interrogative head or interrogative features on a head are often taken to be what determines that a clause is to be interpreted as a question and if the head bearing interrogative features has a Spec position filled with a wh-phrase then it will be interpreted as a direct or indirect question (see May, 1985, on the wh-criterion and extended by Haegeman, 1995, to the NEG Criterion). To receive the right interpretation, then, movement, overt or covert, must take place in syntax to place the wh-question phrase in the position where it needs to be to be interpreted. All of these assumptions about interpretation are essentially uncontroversial in generativist models and have been codified by Rizzi (1997), who regards certain positions as 'criterial'. In order to make the connection between scope and criterial satisfaction more clear for the proposals to follow, the assignment of a criterial domain is formulated as in (18).

18) **Assignment of Criterial Domain**: If a head H has first-merged to Y, Y a non-terminal node, such that Z immediately dominates H and Y, then Y is the domain for H, H_D. If R is the first constituent merged to $[_Z$ H $H_D]$, R is assigned H_D as its criterial domain.

19) **Criterial Well-formedness**: If R is assigned its criterial domain in construction with H, then H and R must have matching features.

'Spec CP' (the constituent merged above C and its complement) which are the positions that are the landing sites for wh-movement and relative operator movement, require an interpretation that is only achievable if the operator is assigned an appropriate domain (which could be scopal or thematic). Movement of an operator immediately above interrogative-marked C and its complement (or whatever head, if it is other than C, that hosts interrogative marking in the left periphery), then establishes the domain for the wh-phrase, which is, in this instance, its scope. The interrogative head may require that the wh-phrase bear matching features, and where this is the case, merger of an inappropriate non-terminal will crash at the semantic interface. The restriction in (19) could be executed in a number of ways (not necessarily with features), but the point of it is that it restricts criterial domain assignment and allows for a certain descriptive leeway as to how 'matching' is achieved (e.g., it could be a semantic requirement for some heads for thematic assignment, a morphological requirement for others), but it is a sort of leeway that is commonly allowed.

For whatever reason, the position first-merged to T and its complement must be filled, which is another way of stating the frequently appealed-to EPP.

20) EPP: A non-terminal must be assigned a domain by T.

Rizzi has suggested that the subject position bears an aboutness relation to its domain, and if so, there is a semantic value to the domain assignment, but the mechanism in (18) does not legislate the content of a criterial relation. On this interpretation, EPP must be satisfied at the semantic interface by an appropriate non-terminal. If the non-terminal is forced to be a DP, by virtue of the Case assignment, control or agreement relations that it must enter into, then T need have no special categorial features that must be matched. It may be, however, that T requires a DP subject, and perhaps what T requires in this context varies parametrically (see section 7). Where EPP requires a DP, it can then be an indirect trigger for A-movement, since the elements destined for this position, in most cases, must be DPs, but no insulated DP can then satisfy EPP, since insulated DPs, by hypothesis, are not DPs.

This much said, the IMB follows without further assumptions. Once a wh-phrase moves to a position where it has to be insulated, it is no longer a DP that can satisfy EPP. Thus movement from vP-adjoined position to subject position would be ruled out by EPP, which requires a DP. If the matching condition imposed by EPP is loosened, then it might be expected that IMB could be violated, and I will consider such a case in section 7.

It is possible that a theory committed to triggered movement could adapt the insulation account and still achieve the same results, but any attempt to write a trigger for insulation itself would fail the explanatory criteria I have held myself to, as it would constitute a stipulation that introduces the distinction. The trigger for insulation must, therefore, be indirect and, as it happens, it must occur at a point in the derivation in advance of any appeal to opacity. In the derivation of (12), repeated below, notice that insulation occurs in (12c) because that is the only point in the derivation where the PNC will permit it.

12a) $[_{vP}$ EA [v [V [wh-DO]]]]
b) $[_{vP}$ $[_{DP}$ wh-DO] $[_{vP}$ EA [v [V $[_{DP}$ wh-DO]]]]]
c) $[_{vP}$ $[_{HP}$ H $[_{DP}$ wh-DO]] $[_{vP}$ EA [v [V $[_{DP}$ wh-DO]]]]] ***insulation - late head attachment***
d) [T $[_{vP}$ $[_{HP}$ H $[_{DP}$ wh-DO]] $[_{vP}$ EA [v [V $[_{DP}$ wh-DO]]]]]]
e) $[_{TP}$ EA [T $[_{vP}$ $[_{HP}$ H $[_{DP}$ wh-DO]] $[_{vP}$ EA [v [V $[_{DP}$ wh-DO]]]]]]]
f) $[_{CP}$ $[_{HP}$ H $[_{DP}$ wh-DO]] [C $[_{TP}$ EA [T
$[_{vP}$ $[_{HP}$ H $[_{DP}$ wh-DO]] $[_{vP}$ EA [v [V $[_{DP}$ wh-DO]]]]]]]]]

If T is merged in (12d) before the wh-phrase is insulated in (12c), then penultimate Merge cannot apply, so T is not even in the derivation to be a direct trigger for insulation. The requirement that insulation occur before the vP phase is complete will also have an interesting consequence for antecedency that is pointed out in section 4.

Before concluding this section, it is perhaps useful to address the potential for overgeneration that penultimate Merge permits. In keeping with the interface filtering design adopted here, it should be expected that penultimate Merge while be filtered out in many contexts where it is formally permitted. The 'first merge' priority conditions on criterial assignment can play an important role here. If the criterial position receives scope/thematic assignment as the first-merged non-terminal (YP) to [H XP] to form $[_{HP}$YP [H XP]], then penultimate Merge of ZP to [H XP] to create $[_{HP}$ YP $[_{HP}$ ZP [H XP]]] has to be interpretable at the interface. If ZP can be absorbed into the criterial relationship, it should be well-formed, as in tucking in contexts, where two quantifiers are interpreted together (e.g., *who saw what*), especially in languages where both wh-phrases front overtly (e.g., Bulgarian as discussed by Richards, 1999). The interpretation of the second wh-phrase depends on the first, but the result is still a pair-list or single pair interpretation, etc. (see, e.g., Dayal, 2002, for discussion and references). Apparently, there is no way to interpret a second external argument if it is penultimate merged under the first one - the result is incoherent at the interface unless a conjunction is used instead. Moreover, adjunction of DP to DP without some means of interpreting the relationship between them should also be filtered out, unless conjunction is

introduced or some sort of apposition relation is licensed (e.g., by prosodic isolation, as in *John, an avid fan, was first in line*.

For better or worse, translating direct triggers into indirect ones does not require anything more than is needed for direct triggers, and this includes the association with Agree. While Agree is not always appealed to as a trigger for movement, many heads seem to require internal Merge into their specifiers for the goals they probe and agree with. In other words, Agree picks a 'nominee' among the candidates that could satisfy the criterial relation by choosing the one that the domain-assigning head Agrees with (which will be the nearest one, assuming minimal search). All that is necessary for the indirect trigger theory is to say that Criterial Assignment in (18) is conditioned, for certain heads or classes of heads, by a requirement that the domain assigning head must Agree with a copy of the criterial nominee. Not all domain-assigning heads will have this condition. Since T in English does not always require movement (as in *there*-constructions), the condition must not hold for T, but by contrast, it appears that the domain-assigning head must always Agree in Bantu, which is how one could translate Carsten's (2005) assertion that EPP features triggering movement are always associated with phi-features in Bantu languages. In general, then, penultimate Merge is free to overgenerate as long it produces a result that is acceptable at the interfaces. If this strategy ultimately places too much burden on the interfaces in the form of stipulated filters, then the appeal of this strategy diminishes, but no special burden is obvious so far, as any theory must have an account of why absorption is possible for some instances of multiple quantification but not for multiple external arguments, and if movement and agreement have to be associated in some cases and not others, any theory that does not derive the association in every case will need some way to distinguish the cases.

The key result of this section is that free Merge eliminates all appeal to EFs, to spurious features that force intermediate movements, indeed to triggers of any sort, including those in the form of uninterpretable features or EPP features that require internal Merge. Merge is free to generate structure and only those structures that satisfy interface requirements are those that are well-formed. The assumption that insulated DPs are different from uninsulated ones is what leads to what is described as the A-/A'-distinction. Since insulation is not stipulated, but typically derived by the need for A'-opacity (or the result crashes), the IMB is a consequence of insulation that is explained without appeal to any assumption that mentions an A-/A'-distinction.

3.3 On the insulating head

Having introduced Merge α and the PNC, the possibility of merging a head to a moved constituent is now unavoidable in the approach outlined here. If it does not happen, A'-opacity is not achieved and the result will crash at the interface. But what is it about the late merged head that makes it an insulator?

Insofar as a head that takes a nonterminal as its sister is always taken to project its label, the penultimate-merged head H will change the phrase type of what it attaches to from the perspective of operations outside of HP. If the insulating head is not a D, it will mean that, to subsequent operations, the vP-adjoined element that has moved is not a DP when T probes for DPs. To achieve this result about what projects, a variety of labeling conventions will do as long labeling of the constituent formed of a head H of type x, H^x, and a phrase of any sort results in a phrase of type x, HP^x. All labeling conventions that project syntactic category agree about this case of labeling, so nothing needs to be added to the theory from the perspective of labeling.

If the syntactic category of the insulating head is not D, then what is it? Three possibilities seem at least to be adequate. For example, H may be a preposition or a

quantificational head or a combination of the two, a preposition with quantificational force. We might expect the insulating head to be visible in some languages.

In fact, an analysis very close to this one has been proposed by Cable (2010), where it is argued that an overt Q-particle takes the *wh*-phrase as its complement and it is the maximal projection of the Q-particle that is then moved. The Q-particle *sá* in Tlingit, as illustrated in (21a,b), must immediately follow the questioned element (wh-phrase) and the wh-phrase must be fronted. Moreover, the presence of *sá* is obligatory, as illustrated in (21c) (p.570)

21a) Aadóoch sá kwgwatóow yá x'ux? (Cable 2010:5690
who.erg Q will.read this book
'Who will read this book?

b)*yá x'ux akwgwatóow aadóoch sá
this book will.read who.erg Q

c) Daa *(sá) aawaxaa i éesh? (Cable 2010:570)
what Q he.ate.it your father
'What did your father eat?'

Cable argues that the Q-particle is crucial to wh-question interpretation and that it is the Q-particle and the phrase that counts as its complement that must satisfy the criterial position for wh-interrogatives. If the Q-particle is not present, then the criterial position is not saturated and the sentence fails. Cable extends his analysis not only to other languages with overt wh-movement and comparable overt Q-particle forms such as Sinhala and Japanese, but also to English, where he posits that the insulator (my term) is null, as in, (22) (see also Safir, 2010).

22a) Whose father's cousin's uncle did you meet at the party?

b) [**QP** [[[[**whose**] father's] cousin's] uncle] **Q**] did you meet at the party?

As Cable points out, quantificational head (identified here as the insulating head) also derives the pied-piping property of the A-A'-distinction.

> Under this analysis, a pied-piping structure in English is derived exactly like the pied-piping structures of Tlingit. In such sentences, the (null) Q-particle takes as sister a phrase properly containing the wh-word, which entails that the fronted phrase of the wh-question properly contains the wh-word. (Cable 2010: 577)

In languages with in-situ wh-structures, he assumes that the Q-particle, if present, does not take the wh-phrase as a complement, but rather the Q-particle is an adjunct and thus does not provide the projected property that could satisfy the criterial position. Cable then proposes that the Q-particle can take a phrase of almost any type as its complement and any phrase it can take as a complement can move with it. In English, this will mean that both DPs and other constituents, such as PPs, APs, and DPs with the wh-phrase in the possessor position can all be complements to the covert Q-particle and the QP is what moves to satisfy the criterial position in question (e.g., in relatives or questions, for example). Thus pied-piping reduces to movement of QP since QP of the right sort is what is needed to satisfy the criterial position where it ultimately lands.

Cable does not address whether or not the Q-particle plays a role in Case assignment or opacity. If the Q-particle is introduced directly on the wh-phrase DP when it is in direct object position, and the predicate takes the wh-phrase as a complement, then the wh-phrase would never get Case and nor is it obvious it would get theta-assigned either, since it is neither the right syntactic category (when insulated by Q) nor of the right semantic type. Thus if the Q particle is (hypothetically) introduced while the DP is in situ, the predicate cannot take the wh-phrase as a complement, as illustrated in (23). This consequence is consistent with Cable's proposal (and provides some motivation for it).

23) [$_{CP}$ C [$_{TP}$ T [$_{vP}$ EA [v [V [[$_{QP*}$ *Q [$_{DP}$ wh-DP]]]]]]]

For the same reason, even in languages without overt movement, the Q-particle must not be introduced until the wh-DP it is out of the Case-assigning and thematic position, i.e., when the wh-phrase optionally moves to the vP phase edge. At that point the Q-particle is introduced by penultimate Merge, Q takes the wh-phrase as a complement, and A'-opacity results. Moreover, since what the insulator projects is not a DP, the moved DP retains just the Casemarking it has received in its pre-movement position.

In the context of my proposal for how criterial domains are assigned and conditioned, it can now be said that, for Cable's cases, Q-head insulation enables a DP to be an operator adequate for a criterial position. The same reasoning applies to the questioning of subject positions, since EPP will not be satisfied if insulation occurs before the phase where EPP applies has closed. Thus insulation must occur after movement to the edge of the CP to satisfy the criterial position assigned by C. When the CP edge position that the subject wh-phrase moves to is not criterial (as in the case where it is cyclic movement through the CP edge on the way to a higher criterial position), insulation must occur as soon as the wh-DP would be in danger of receiving a second Case assignment or thematic role. When the matrix predicate is a verb taking a complement clause, there is danger of Accusative assignment. However, the same issue does not arise for an adjectival predicate that takes a clause, at least until vP adjunction in the matrix clause, since adjectives do not assign Case in English.[4]

24a) Who is Gladys glad (*that) ate the fish?

b) [$_{CP}$ [$_{QP}$ *Q DP*] [$_{TP}$ Gladys [$_{vP}$ [$_{QP}$ Q *DP*] [$_{vP}$ is Gladys glad
[$_{CP}$ DP [(that) [$_{TP}$ *DP* ate the fish]]]]]]]

It is not clear that derivations like the one in (24b) should be ruled out generally, but if they are ruled out, it may be because the DP in the CP edge is visible to the theta-assigning embedding predicate and the DP in the CP edge introduces a fatal conundrum for theta assignment - unless it is insulated[5]. After all, insulated phrases are never candidates for complement arguments, but DPs are. It is a natural assumption that criterial assignment would hide a constituent on the phase edge, as in the case of indirect question complements, where the wh-phrase is in a criterial position, or for Exceptional Casemarking complements where the complement subject is in the EPP position, even though it receives Case from the higher clause. However, in the absence of criterial assignment, a DP on the phase edge of a complement clause would count as an additional argument inducing incoherence at the interface unless it is insulated. Thus it could be that insulation is always required for wh-phrases as soon as they move to a phase edge.

I have now introduced five different reasons why insulation might be motivated to occur and they are listed in (25) and (26).

25) Insulation at the vP edge

a) Prevents wh-DP adjoined to vP from blocking assignment of NOM to Spec vP.

b) Prevents wh-DP adjoined to vP from blocking Agree between T and Spec vP.

c) Prevents double assignment of Case to wh-DP which is already Case-assigned.

26) Insulation at the CP edge

a) Allows Q-head to facilitate criterial satisfaction.

b) Prevents incoherence for thematic role selection.

---

[4] See Kayne (1983:3), who noted that such sentences are degraded with or without *that*. Here we are concerned with the possibility that (22) could be acceptable with subject extraction, but if it is always blocked for other reasons, then the issue addressed in this paragraph is potentially orthogonal.

[5] Chomsky (1986:16) remarks in that Kyle Johnson suggested to him that confusion of theta-assignment could arise when a DP adjoins to CP. Chomsky used this suggestion to forbid adjunction to CP as opposed to movement into Spec CP. With the absence of a theoretical commitment to a Spec CP position, the issue comes down to one of labeling. In this case, Chomsky (2013) will require the wh-phrase to move further in order to allow CP to project. The question is whether thematic assignment only applies at the end of the phase or sooner, and if it is sooner, then insulation will still be necessary. I leave the matter open, but see section 6.

c) Prevents double assignment of Case to DP that is already Case-assigned (in lower clause).

It may be that double Case assignment is not always excluded and may not be necessary to motivate insulation which is overdetermined at the CP edge by (26b) alone, and at the vP edge by (25b,c). It is possible that some of these motivations may be moot in some languages based on how Case and Agree apply, as discussed in sections 7 and 8. The absence of insulation is motivated by theta-assignment to DP and by EPP,[6] which normally requires a DP for criterial satisfaction, another point addressed in sections 7 and 8.

The marriage of Rezac's approach to A'-opacity and Cable's approach to pied piping shows that insulation, as it is formulated here, now accounts for why pied piping is possible for so-called A'-movement, but not for so-called A-movement. A-movement does not require insulation and usually is not compatible with it, because elements that move to satisfy EPP or Case or agreement relations must be DPs, and if the relevant relations are not satisfied, the derivation leads to crash at the interface. However, insulation in the form of Q may not always be necessary to meet the matching condition on a given criterial domain. In such a case, insulation is still necessary to achieve A'-opacity, but the choice of head may be more flexible. I return to such cases later in section 4.4. The timing of insulation, when it is necessary for a result acceptable at the interfaces, must always at least precede the close of the phase because as soon as a head is merged above the phase, insulation is no longer permitted by the PNC for an element in the phase edge. Thus Cable's proposal, united here with Rezac's proposal, combine to derive both A'-opacity and the pied piping property of so-called A'-movements.

4.0 Binding mismatch

The existence of penultimate Merge allows for insulation to take place freely in keeping with free Merge, but the interface conditions that must be met constrain its timing and distribution. So far, I have shown how insulation derives A'-opacity for Case and agreement, the pied piping distinction and the IMB. The key assumption has been that an insulated DP is not a DP. If we now assume that antecedents must somehow match the positions that they bind, then it becomes possible to explain the anaphor binding distinction, the weak crossover distinction, and the parasitic gap distinction.

27) **Binding Mismatch:** If [H [DP]] is not a DP then it cannot antecede a DP.

For this result to hold, I must rely on a common assumption about phases, namely, that the phase head complement is only submitted to the interfaces for spell-out and interpretation after the end of the phase, which would be the point at which all Merge applying to the phase edge is complete. Reconsidering once again the derivation in (12a-c)

12a) $[_{vP}$ EA [v [V [wh-DO]]]]

b) $[_{vP}$ $[_{DP}$ wh-DO] $[_{vP}$ EA [v [V $[_{DP}$ wh-DO]]]]]

c) $[_{vP}$ $[_{HP}$ H $[_{DP}$ wh-DO]] $[_{vP}$ EA [v [V $[_{DP}$ wh-DO]]]]] ***insulation - late head attachment***

At the moment the vP phase is complete, then, interpretation can only treat the wh-phrase as an HP, not a DP. In other words, no appeal can be made to the earlier point in the derivation, (12b), where the wh-phrase is still a DP, but the vP phase has not been completed. Thus any determination of eligible antecedents for an element in the vP complement must treat the wh-phrase as an HP, not a DP. The EA, which is in the vP edge, is not interpreted until the next phase closes, and so can only count as a DP antecedent (even if it is a wh-phrase in EA position).

[6] It is sometimes suggested that EPP for the Spec TP position is not universal. Bobaljik and Wurmbrand (2005), for example, has argued that does not hold in German, but arguments concerning control infinitives in German suggest the opposite (see Safir, 1985; 208). I will assume that EPP is universal, but interesting questions may arise for the distribution of insulation depending on how the absence of EPP is argued for.

The next natural question is how binding relationships are established between a DP trace and the HP that, in my account of antecedency, should not be a potential antecedent for a DP. There is an important difference, however, between dependency of one DP on another and the relation between copies resulting from internal Merge. The dependency relation is asymmetric (see section 4.2) insofar as one partner in the relationship acts as the antecedent. The relationship between copies is not one of antecedency, but indistinctness. While copies created by internal Merge may have different locations in structure, they cannot otherwise be different unless they are in some way transformed. In other words, copy relations do not create dependent identity relations unless something further takes place in interpretation to distinguish an operator from its trace. Whether trace interpretation involves the sort of type shifting proposed by Sauerland (2004) or definite interpretation (as in Fox, 2003), or some other operation, the treatment of lowest copies as variables (of whatever type) when an expression is moved is already a separate path to antecedency specific to copies. I rely on the plausible assumption that this path to antecedency is not sensitive to HP-DP binding mismatch since the copy relation is established by internal Merge and holds whether c-command holds or not (e.g., for penultimate Merge).[7] What is needed for my account to go through is summarized in (28).

28) If X is created by internal Merge applying to Y such that X and Y are copies,
then Y is interpreted as identity dependent on X.

The conversion of copy relations to dependency relations is assumed in this account and thus represented by a dependency arrow (introduced in 4.2) – the lowest copy depends on the highest one or whatever contains the highest one. Movement without insulation is no different, except that the highest copy directly c-commands any lower copy. This conversion of a copy relation to a dependency relation can also apply to head movement. For example, when a head A internally Merges to head B and then [A B] is internally Merged to a head C, the relation of the copy in [[A B] C] to the lowest copy of A is determined as dependency by the same mechanism. Thus (28) does not correspond to the A-/A'-distinction since it applies to A-movement and head movement as well, but is rather a consequence of the way that relations between copies are converted into dependency relations.

4.1 Anaphor binding

Binding mismatch leads us to expect insulated phrases cannot antecede DPs, so insulated phrases will never be able to bind anaphors. As soon as DPs move to positions where they are insulated, they are disqualified as binders for anaphors. Only their positions before movement and insulation, what used to be called A-positions, are eligible to bind anaphors. Thus a wh-phrase can permit anaphora binding just in its highest position before insulation. For example, the lowest DP trace/copy of *which students* c-commands *each other* in (29) but not in (30).

29a) *Which students* did John say spoke to *each other*?
b) [H [*which students*]] [$_{TP}$ did John [$_{vP}$ [H-*which students*] [$_{vP}$ say
[$_{CP}$ [H-*which students*] [$_{TP}$ [*which students*] spoke to *each other*]]]]]]

30a)**Which students* did *each other* say spoke to Bill?
b) [H [*Which students*]] [$_{CP}$ did *each other* [$_{vP}$ [H [*which students*]] [$_{vP}$ say
[$_{CP}$ [H [*which students*]] [$_{TP}$ [*which students*] spoke to Bill]]]]]

In (29a), there is no insulation in the lower TP since the wh-phrase must be a DP for Case, agreement and EPP, but as soon as the wh-phrase moves up to the lower CP, the moved subordinate subject must be insulated before it either causes thematic confusion and/or is

[7] This matter is specifically taken up in Safir and Bassene (under review) for head movement with respect to the algorithms that determine which copy is pronounced when internal merge does not result in a c-command relation distinguishing a higher and lower copy.

doubly Case-assigned with ACC from *say*, so the wh-phrase is already insulated after its first movement to the lower CP in both examples. The insulation it already has insures A'-opacity in the matrix clause adjoined-to-vP position. Thus the insulated operator is not an appropriate antecedent by the time it locally c-commands *each other* at the end of the derivation in (30).

4.2 Weak crossover

The antecedent matching requirement between elements in distinct chains applies to insulated operators and pronouns just as it applies to insulated operators and DP anaphors. Movement without insulation should permit pronoun-binding, but insulated moved constituents are no longer DPs, so they cannot directly bind pronouns. Embedded within an independently motivated theory of weak crossover (WCO), insulation not only derives the fact that only insulated movement induces WCO, but derives a key feature of most WCO accounts that is normally stipulated.

Reinhart (1983) argued that pronouns are only bound by c-commanding antecedents and WCO arises when a pronoun is not c-commanded by the quantifier (or quantifier extraction site) is supposed to be bound by. Higginbotham (1983:402) and Safir (2004) argue that WCO arises when a pronoun or a constituent that contains it c-commands a quantifier or extraction site that the pronoun depends on (and we will follow the latter account). Higginbotham (1983:404-405) observes that the 'depends on' relationship is asymmetric, a fact that is not captured by co-indexing notation, which is symmetric. English reflexive *himself*, for example, gets its value from whatever binds it, but it does not transmit that value to its binder. For this reason, Higginbotham (and Safir, 2004, following him) proposes that in lieu of indices, dependencies should be represented by asymmetric connections, such as the ones in (31), where the anchor, ├, indicates the antecedent and the hook, ┘, indicates the element that depends for its value on the antecedent.

31a) Percy praised himself
├──────────┘
b) Alice praised her brother
├──────────┘

The bound pronoun reading in (31b) is also captured by the linking notation. Developing and revising Higginbotham's account, Safir proposes the Independence Principle.

32) **Independence Principle** (INP)
α cannot depend on β if α is (embedded in) γ and γ c-commands β.

Thus the Higginbotham-Safir account is based on freely assignment of dependency relations but a negative condition on dependency, whereas the Reinhart account is based on allowing relations to be established if they meet a certain condition (licensing). This difference is not crucial for what follows.

On Safir's (2004) account, as in Reinhart's, it is not possible to link a moved quantifier directly to a pronoun and it is stipulated that the only way a pronoun can be a bound variable is through its relationship to the quantifier in situ or the gap it leaves after movement. Safir formulates the condition as follows:[8]

33) **Quantifier Dependency Condition** (simplified from Safir, 2004)
If X is not the trace of OP, then X cannot depend directly on OP, but only on its trace.

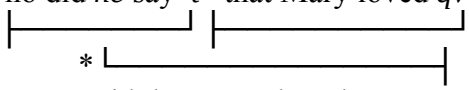
This stipulation is now unnecessary as it is derived by antecedent matching, insofar as moved quantifiers, overt or covert (where covert quantifiers are moved in syntax but pronounced low – see Bobaljik, 2003, and Safir, 2010), are ruled by the same logic as overt wh-movement,

[8] Koopman and Sportiche (1983) assumed that displaced operators could bind pronouns in principle, but limited how that was done (one-one restriction for Bijection, parallelism for Safir's PCOB). Safir (1984) required pronouns bound in this way to be somehow parallel. I reject these accounts, but I accept the possibility that an uninsulated DP quantifier could bind a pronoun if it is not insulated, even in displaced positions.

namely, such moved constituents must be insulated or they will fail to deliver A'-opacity, leading to crash at the interface. If an insulated quantifier cannot bind a pronoun, then the pronoun must be bound or somehow depend on the site it originates in, since only that site is a DP.[9]

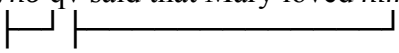
Assuming, then, that Binding Mismatch requires a bound pronoun to depend on the extraction site, the relations that must hold for a pronoun to be interpreted as a bound variable are illustrated in (34a,b) for strong crossover and for WCO in (35a,b) (where qv=quantifier variable.

34a) Who did *he* say *t* that Mary loved *qv*

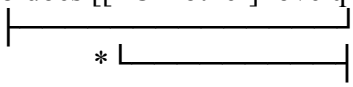


*

b) *Who* qv said that Mary loved *him*

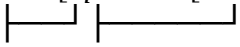


35a)**Who* does [[*his* mother] love qv]?



*

b) *Who* [*qv* loves [*his* mother]]?



Insulation leading to binding mismatch is responsible for feeding (34a) and (35a) to the Independence Principle which in turn rules out both dependency patterns. In (34a), the bound reading is excluded because *he* depends on *qx* and *he* c-commands *qx*, while in (35a) [*his* mother] c-commands *qx* and *his* depends on *qx*.

The mismatch approach proposed here recalls Sauerland's (2004) approach which is based on a semantic type mismatch between wh-phrases, which, according to him, assign individuals to properties (i.e., they are choice functions), and pronouns, which are individual variables. Some appeal to c-command is still required for DPs in situ (containing choice function variables) to bind only those pronouns (individual variables) that they c-command, as in other accounts (e.g., Reinhart's licensing account). The choice function account is motivated by non-crossover phenomena, so it may be that WCO is overdetermined, but if the insulation story is right, then an insulatoin account is not independently motivated by WCO.

Notice, however, that choice function account of wh-variables does not explain cases where resumptive pronouns are bound in restrictive relative clauses. From the syntactic perspective, insulation is not necessary in such cases, so it is predicted that resumptive pronouns, where no movement is involved, do not induce WCO.

36) She was always going out with the kind of guy who, after you meet him, you have the feeling that he is just like the last one she went out with.

As mentioned earlier in fn.8, nothing in the insulation account prevents a DP operator from binding a pronoun as long as the operator is not insulated. If the restrictive operator *who* is a choice function, it cannot bind either pronoun, unless the pronouns are shifted to be variables of properties to which individuals are assigned. However, this would predict that the pronouns would have trace-like structure even where they cannot be extracted (which leads to bad reconstruction effect predictions). Moreover, if pronouns can be converted to trace-like interpretations for the semantic mismatch theory, then something must be introduced to prevent this from happening in all WCO contexts, which otherwise would then be expected to be grammatical. The insulation approach is not committed to a choice function analysis, but

[9] The binding mismatch story also explains why PRO can never be bound directly by the operator in PRO gate constructions. Such constructions are only possible if a control relation is possible between the extraction site and the position of the PRO independent of wh-movement. See Safir (2004) for discussion.

predicts no WCO for the two pronouns, since the operator does not have to be insulated – it never moves – and so can be a DP binder for both pronouns.

4.3 Parasitic gaps

The same antecedent matching requirement explains why only so-called A'-constructions license parasitic gaps. As Chomsky (1986) argues, parasitic gaps are traces of operators that have moved within the constituent, be it an adjunct or a subject, and the parasitic operator is c-commanded by the operator that licenses the parasitic construction. In the account proposed here, this means that the parasitic operator that binds the parasitic gap must be an insulated operator, and if the licensing operator must bind the parasitic one, this will only be possible if they satisfy antecedent matching. The structure in (37b) shows that the highest [H [*which mayor*]] c-commands the adjunct (adjoined here to TP for presentational purposes). The bolding in in (37b) shows that the categories of the null adjunct operator and the fronted wh-phrase match. The dependencies involved for (37a) are illustrated in (37c) (where pg=parasitic gap)

37a) Which mayor did John praise after talking to?

b) [$_{CP}$ **[$_{HP}$ H [$_{DP}$ *which mayor*]]** [$_{TP}$ did John [$_{vP}$ [$_{HP}$ H [$_{DP}$ *which mayor*]]
[$_{vP}$ praise [*which mayor*]]
[$_{PP}$ after [$_{CP}$ **[$_{HP}$ H-*OP*]** [$_{TP}$ PRO [$_{vP}$ [$_{HP}$ H-*OP*] [$_{vP}$ PRO talking to *OP*]]]]]]]

c) [*which mayor* [did John praise *t*]] [*OP* before PRO talking to *pg*]],…

(*OP* depends on *which mayor*)

Not every movement that requires insulation supports parasitic gaps, but uninsulated movement never supports parasitic gaps, and it is this latter contrast that this theory derives. DP movement cannot license pgs because the DP does not match, and thus cannot antecede, the insulated operator that would have to be present in the parasitic gap clause. Once again, insulation makes the key difference.[10]

Notice, however, that the account of parasitic gap licensing based on insulation does not explain the inability wh-movement from the matrix subject to license parasitic gaps.

38a)*Who [did [Sarah see t] [before OP [pg left]]]?

b)*Who [t left [before [OP Sarah saw pg]?

Movement from matrix subject position to the matrix fronted position does not require insulation for A'-opacity in a matrix clause (there is no danger of an additional Case or thematic assignment from a higher phase head), but if a question requires the wh-phrase to be embedded in the QP projected by a Q-particle, insulation is obligatory. If insulation occurs in (38a,b), parasitic gaps are licensed, contrary to fact. Thus the account of parasitic gaps offered here explains why insulated operators are capable of supporting parasitic gaps and DP antecedents are not, but it is not a complete account of the restrictions on the distribution of parasitic gaps.

[10] It could be said, however, that a DP in subject position should be able to antecede another DP, just as an insulated DP operator can antecede another insulated DP operator. In fact relationship might be the correct way to characterize subject-oriented adjunct control, as in i.

i. Wanda was walking down the street eating an ice cream one.

What is at stake here is not whether or not *Wanda* has moved (though it presumably has moved as the EA of the vP ultimately headed by *walk*), but simply that it is the right kind of phrase to be an antecedent for the PRO subject of *eating an ice cream cone.*

4.4 Weakest crossover

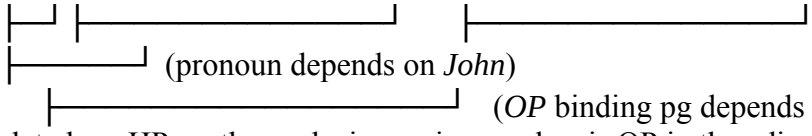
It is well known that has in a variety of A'-constructions, the WCO effect does not appear even though the structural configuration seems to match those contexts where WCO effects are in force. Lasnik and Stowell (1991) compiled a number of such cases described them as 'weakest crossover'. The constructions are illustrated with English examples in (39), where parasitic gaps are supported even though the WCO effect is absent.

39a) *John* is tough *OP* for *his* mother to talk to *t* without insulting *pg*
b) *John, his* mother can talk to *t* without insulting *pg*
c) *John* is too angry *OP* for *his* mother to talk to *t* without insulting *pg*
d) *John*, *who* we told *his* mother to talk to *t* without insulting *pg* …

From the perspective of my proposal, the absence of WCO suggests that the operator (OP) should be a DP and thus a matching antecedent for pronouns, but support for parasitic gaps suggests that OP is not behaving as a DP. This analysis of weakest crossover might suggest that an internal contradiction arises for the insulation-based theory supported here, but in fact no such contradiction arises.

Lasnik and Stowell suggested that the quantificational status of the operator may be the key factor, whereby only 'true quantifiers' are taken to induce WCO in A'-constructions. When the fronted element is not a true quantifier, the effect disappears, according to them. The notion 'true quantifier' has never been clear, however, and as a result there is a certain fuzziness about when the effect should be expected. By contrast, the approach developed here treats the failure of antecedency to be a thoroughly structural matter based on binding mismatch. The binding mismatch account can be rescued from this apparent paradox by appeal to the account of weakest crossover in Safir (2004, with antecedents cited there), where it is argued that weakest crossover constructions are not exempted by the kind of quantification that is involved, but rather by the availability of a nominal other than the operator that can serve as an antecedent. In (39d), there is a nominal, the proper name *John*, which is not restricted by the clause over which the OP presides.

40) *John* [*who* [[*his* mother] [talks to *t*]] [*OP* without PRO insulting *pg*]],…



Who is insulated, an HP, as the analysis requires, and so is OP in the adjunct, so the HP *who* can bind the HP OP and a parasitic gap is licensed. Neither OP nor *who* is a DP in its landing site and so neither can bind pronouns, as usual. The pronoun *his* in (39d) does not depend on the operator or its trace, but rather depends directly on *John*. The same analysis can easily be extended to (39a) and (39c). For topicalization, a null operator analysis would produce the same result, but it is not clear that a null operator is warranted. Rather the topicalization is probably insulated, hence not a possible antecedent for *his* in (39b), but since topics are discourse familiar, the pronoun can access the discourse referent the topic refers to, as in cases of apparent backwards coreference (see Safir 2004:53 and fn.4, 173).

Thus weakest crossover is not a counterexample to the insulation-based theory, but is rather another consequence of the interaction between insulation (which leads to binding mismatch) with independent properties of constructions. The analysis remains uncompromised.[11]

[11] I offer no proposal as to how *tough*-constructions license the matrix subject .The smuggling analysis of *tough*-movement proposed by Hicks (2009) is potentially compatible with this analysis. On that account, the OP is a constituent larger than a DP and a DP is extracted out of it from Spec CP and moved to Spec TP. Even if this is correct for *tough*-constructions, this form of analysis does not appear to extend to (39b-d).

5.0 Reconstruction

One of the purported advantages of the copy theory of movement is that the so-called reconstruction effect can be explained by the presence of a copy in the launching site position of a moved constituent which is accessible to the semantic interface. Thus the acceptability of (41) is licensed because the pronoun *his* can be bound by *any man* in the position where it first entered the derivation.

41) [How many of *his* sins] does *any man* ever admit to [how many of *his* sins]?

There are, however, contexts where the availability of a copy in the lowest position, or the necessity to leave one, makes the output sensitive to negative effects, such as the Principle C effect, or whatever induces WCO (e.g., the Independence Principle) as in (42a), with the relevant copy relations presented in (42b).

42a)*Which evidence that *Dean's* accomplices were guilty did *he* expect would incriminate Erlichman?

b) [Which evidence that *Dean's* accomplices were guilty] did *he* expect [which evidence that *Dean's* accomplices were guilty] would incriminate Erlichman?

Since Lebeaux (1990), it has been argued that examples like (42a) contrast with (43a), in that the coconstrual of *he* and *Dean* is considered better, or fully acceptable, in (43a) (e.g., as argued for analogous cases in Freidin, 1986 and Lebeaux, 1990).[12]

43a) Which evidence that implicated *Dean's* accomplices did *he* expect would incriminate Erlichman?

b) [How much evidence that implicated *Dean's* accomplices] did *he* expect [how much evidence that implicated *Dean's* accomplices] would incriminate Erlichman?

Lebeaux pointed out that the gapless relative in (42a) is generated as a direct sister to N for a class of nouns that can license gapless relatives as their complements, while relatives with a gap such as in (43a) are compatible with any sort of noun, and hence is not a selected complement, but rather a kind of adjunct modifier. Lebeaux proposed that the adjunct modifier is not a sister to N, thus it can be added later in the derivation when the N is encased in a DP. This 'late attachment' is an instance of adjunction that is equivalent to penultimate Merge as formulated here. However, the direct complement cannot be added later and still be a complement to N, which is too deeply embedded.

Lebeaux did not say exactly why lower attachment would be too low for a counter-cyclic operation. Consider the derivation in (44) as a possible source for (42a).

44a) $[_{CP}$ $[_{DP}$ which] [did *he* expect $[_{DP}$ which] would incriminate Erlichman]]

b) $[_{CP}$ $[_{DP}$ $[_{D}$ which] [evidence that *Dean's* accomplices were guilty] [did *he* expect $[_{DP}$ which] would incriminate Erlichman]]

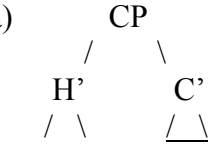
In (44b), a complement clause has been late attached (penultimate Merged) to the D *which*. If this derivation is possible, then the reconstruction asymmetry should disappear, as late attachment can apply equally to complements and adjuncts – that is, (42a) would incorrectly be predicted to be acceptable. However, if *which* must be insulated in the course of the derivation, as argued here, then late attachment of the complement of *which* will violate the PNC, since the sister position to *which* is too deeply embedded under the insulating head.

45a) $[_{CP}$ [H [which]] [did *he* expect [which] would incriminate Erlichman]]

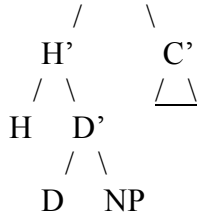
b)*$[_{CP}$ [[H [which]] [evidence that *Dean's* accomplices were guilty]] [did *he* expect [which] would incriminate Erlichman]]

Thus (45b) cannot be generated. In order to generate (42a), the complement to *which* must be introduced low, as in (42b), if it is to enter the derivation as a complement at all, as schematically illustrated in (46a,b), where late attachment applies to (46a), yielding (46b).

[12] The generalization about sentential complements is disputed by Safir (1999), but Safir ultimately accepts a complement adjunct distinction on other evidence. Illustration of the effect with contrasts between sentential complements and adjuncts is offered here for convenience of presentation.

46a)


b) *CP late attachment of NP violates PNC



By contrast, there is no difficulty in applying penultimate Merge (=late attachment) to the insulated wh-phrase as long as the interpretation is not that of complementation, but modification, as in (47)

47a) [$_{CP}$ [H [which evidence]] [did *he* expect [which evidence] would incriminate Erlichman]]

b) [$_{CP}$ [[H [which evidence]] [that implicated *Dean's* accomplices]] [did *he* expect [which] would incriminate Erlichman]]

Thus insulation *explains* the adjunct/complement distinction for reconstruction, i.e., why only adjuncts, and not complements, can be late-attached to fronted wh-elements.

There is, however, a further distinction that has been claimed by Sauerland (1998) to distinguish A- vs. A'-structures. He takes the purported generalization in (48a,b) as a starting point.

48a) A-movement *optionally* leaves a trace.

b) A'-movement *obligatorily* leaves a trace.

The reason that Sauerland assumes that A-movement does not have to leave a copy at the point in the derivation where a DP is merged is because there is no Principle C reconstruction effect.

49a) The evidence that *John* is guilty seems to *him* to be without merit.

b) [The evidence that *John* is guilty] seems to *him*
[$_{TP}$ [the evidence that *John* is guilty] to be without merit]

On the assumption that *him* c-commands *John* within the lower copy in spite of the position of the pronoun within a PP, Sauerland reasons that a copy left by A-movement should induce a Principle C effect, contrary to fact ((49a) is acceptable). Instead of allowing A-movement to leave no copy, he accepts that movement must always leave a copy, but he suggests that A-movement always allows for a late attachment derivation. When the argument of the verb is first merged, it can consists only of a D head (e.g., [the]). The D head moves to the position of the top of the A-chain, and then [*evidence that John is guilty*] is attached late by late attachment (penultimate Merge).

50a) [the] seems to *him* [$_{TP}$ [the] to be without merit]

b) [[the [evidence that *John* is guilty]] seems to *him* [$_{TP}$ [the] to be without merit]

Notice that this is very much like the derivation in (44) that we ruled out on the grounds that the PNC would not allow the complement to wh-D could to be late-attached to the *insulated* wh-DP or the PNC would be violated. This is precisely what is different about so-called A-movement, however. Late attachment of [*evidence that John is guilty*] is a simple case of penultimate Merge, precisely because *the* has not been insulated. If it were insulated, it would not satisfy EPP, which requires a DP subject. Thus the absence of a Principle C effect is due to the availability of late attachment at the top of a so-called A-chain in the absence of

insulation. Once again, no appeal is made to a stipulated distinction between A-chains and A'-chains, neither in distinguishing two classes of movement by virtue of the sorts of copies they leave (all movement leaves a copy), nor by any stipulation distinguishing between classes of feature types that trigger movement.

Takahashi and Hulsey (2009) develop a version of Sauerland's account of the absence of reconstruction effects in A-chains (see their fn.13 for other predecessor accounts). They appeal to Case theory to distinguish why late attachment is possible for the complements of the highest position in an A-chain but not for the complements of an A'-chain. T&H claim that A'-movements of D heads cannot get their restrictors/complements by late attachment because N must get its Case in a Case-assigning position. The top of an A-chain is such a place, but not the top of an A'-chain. If the NP complement of D is added to the top of the A'-chain it cannot get Case and the derivation fails. This is consistent with the theory presented here, but one of Takahashi and Hulsey's enabling assumption does not, in fact, have to be stipulated – it follows from insulation. A wh-phrase that has been insulated cannot subsequently get Case, which explains why the top of an A'-chain is not a position where Case is assigned. Since a Casemarked D in an A-position is not insulated, late merger can apply to it and satisfy the PNC as in (49). Nothing has to be said (with respect to the A-/A'-distinction) about differences between highest A'-chain positions vs. highest A-chain positions with respect to Case, and indeed nothing needs to be said about how N gets Case, at least not to distinguish where late attachment can or cannot apply.[13]

To summarize this section, the insulation theory provides the key to explaining the distribution of reconstruction effects that have been attributed to the A-/A'-distinction and the complement/adjunct contrast. Not only does it explain why so-called A'-movement only allows late attachment of adjuncts (after insulation, the PNC would be violated), but it allows late attachment of D-complements if the D has not been insulated (as in so-called A-chains). Both the possibility of late attachment, here defined as penultimate Merge, and the possibility of insulation follow from redefining what Merge can generate according to the PNC. Where penultimate Merge is well-formed and when it can occur in a derivation are all determined by interaction with other factors, such as phase theory and where Case assignment, thematic role assignment, criterial satisfaction and/or agreement must take place.

6.0 Scrambling

It has long been noted that some languages described as having short-distance scrambling allow displacement constructions that permit the union of output sentences that would be well-formed as A-constructions and those sentences that would be expected to be well-formed if scrambling is considered an A'-construction. In some languages, even if local scrambling has only A-properties, long distance (LD) scrambling always has A'-properties (see Déprez, 1989, Mahajan, 1990 and Webelhuth, 1992). To address the full range of issues that scrambling constructions raise is beyond the scope of this paper, but whether or not scrambling constructions present special problems for the insulation approach to the A-/A-distinction as opposed to some of the feature-based proposals mentioned in section 1 is a question that needs to be addressed.

[13] Miyagawa (2010) argues that reconstruction is optional for A-chains because they do not involve movements across 'transfer domains', that is, they involve internal Merge to positions inside phase heads, rather than by internal Merge in the phase edge. The proposal, as applied to topicalization and focalization in Japanese requires an architecture of feature-inheritance that involves the presence or absence of an additional α-projection that facilitates feature inheritance differentially in Japanese, as opposed to English. This difference does not exclude the need for late attachment for wh-movement, however, as Miyagawa (2010:117) notes, and thus the feature-based A-/A'-distinction potentially adds more than it takes away, even if the α-projection proposal turns out to be correct.

Scrambling in Japanese is used here to illustrate the challenge that any explanatory theory of the A-/A'-distinction must face and I compare my account to that of Miyagawa (2010) who also attempts to address the A-/A'-distinctions in a principled way. In Japanese, scrambling within a clause (local scrambling) can move an object past a subject, which is an A'-property in comparison to English, and local scrambling of an object creates a binder for an anaphor in the subject, as in (51) and neutralizes WCO, as in (52) (both (51) and (52) are from Saito, 2003; 485)

51) ?[TP *Karera-o* [[*otagai-no* sensei]-ga *t* hihansita]] (koto)
they-ACC each.other-GEN teacher-NOM criticized fact
'*Them each other's* teachers criticized.'

52a)?*[TP [*Sono* tyosya]-ga *dono hon-ni-mo* keti-o tuketa]
its author -NOM which book-to-even gave-criticism
'*Its* author criticized *every book*'

b) [TP *Dono hon-ni-moi* [[*sono* tyosya]-ga *t* keti-o tuketa]]
which book-to-even its author-NOM gave-criticism
'*Every book*, [*its* author criticized *t*]'

Scrambling outside the clause of origin, that is long distance (LD) scrambling do not allow anaphor binding, as in (53) and WCO is still in force (54).

53) *[TP *Karera-o* [[*otagai-no* sensei] -ga [CP[TP Tanaka-ga *t* hihansita] to] itta]] (koto)
they-ACC each other-GEN teacher-NOM -NOM criticized that said fact
'*Them*, [*each other's* teachers] said that Tanaka criticized *t*'

54) ?*[TP *Dono hon-ni-mo* [[*sono* tyosya] -ga [CP[TP Hanako-ga *t* keti-o tuketa]to] itta]]
which book-to-even its author-NOM Hanako-NOM gave-criticism that said
'*Every book*, *its* author said that Hanako criticized *t*'

Following the logic of insulation theory developed here, conditions must conspire to insure that insulation must be forced where the construction (descriptively) shows A'-properties (resulting from binding mismatch), prevented where it only shows A-properties, or optional where the union of well-formed A- and A'-construction outputs are allowed. Since A'-opacity is what drives insulation in most cases, the natural question is to see if Japanese requires A'-opacity. Although Japanese has Case particles on DPs, it has no visible agreement, so if Case can be assigned successfully without appeal to opacity within the clause, then fronting to a topic or focus position, which arguably do not require a Q-particle, then fronting of an object past a subject should be possible without insulation.

Miyagawa's (2010) account draws on the presence of an extra projection between CP and TP, αP, such that α can inherit the probing potential from C, following Chomsky's, (2007, 2008) and Richard's (2007) theory of feature inheritance. Under that theory, the phi-features that make T a probe do not originate on T, which is not a phase head, but on C, which is a phase head. For T to act as a probe, the features of C must be inherited by T, and then T probes for Case and phi-features. The presence of the α head between C and T means that α inherits the ability to probe instead of T, but α probes for topic or focus as well as traditional phi-features (not a factor in Japanese), and unlike T, α is stipulated not to probe for Case.[14] If Case is not probed for, then there will be no locality difficulty in Miyagawa's approach if an object moves past the EA to adjoin to vP. before it then moves to Spec αP.

An insulation analysis is potentially compatible with Miyagawa's approach. If the EPP feature resides where the probing features are - not on T in Japanese, but on α - then EPP is satisfied by moving the DO into Spec αP position without any insulation being required. If there is no insulation, then there should be no binding mismatch and scrambled DPs should

[14] Miyagawa is very spare in his appeal to Case, arguing that only in languages where α does not inherit the probe features and T does is Case probed for. He claims that Case is a lexical relation, not a functional one, as he takes Agree to be, but it is not so clear in his account what the status of Case is in α-probing languages.

be licit antecedents for both anaphors and bound variable pronouns, as in (51) and (52b), respectively. The A'-effects for LD scrambling suggests that insulation must occur before the scrambled DP leaves its clause of origin. Here I appeal to the need to insulate DPs in the CP edge from argument structure assignment by a higher clause, as I did with respect to adjectival clausal complements in (24). Thus movement out of a clause with a CP edge will always result in insulation (and here double Case assignment would not play a role).

I am not claiming that Miyagawa's theory is correct and that the insulation approach to the A-/A'-distinction is therefore correct, but rather that at least one thoughtful approach to licensing the properties of scrambling structures can easily be interpreted in such a way as to be consistent with the insulation-based approach. Nor am I claiming that insulation necessarily provides a better account of scrambling phenomena than those that exist, although that could turn out to be the case. It may be that different accounts of scrambling or scrambling in languages other than Japanese may involve more serious challenges to the insulation analysis and this I leave to future research. However, if there are accounts of scrambling do not pose any special problem for a theory that derives the A-/A'-distinction only by positing that penultimate Merge is possible, then it is not obvious why any approach to scrambling that does not permit the insulation theory to succeed would be appealing.

7.0 Languages where movement can be 'improper'

The insulation approach does not depend on binding mismatch to block improper movement, but rather the assumption that non-DPs cannot satisfy EPP in the languages that we have examined so far. However, if EPP can vary crosslinguistically with respect to the kinds of constituents which can satisfy it, then it could be the case that an insulated DP can move to the 'subject', that is, the specifier position of the EPP head, typically T. Alternatively, if a wh-phrase is never insulated, then it should be an adequate specifier to satisfy the EPP even in languages where only DPs will do. Carstens (2005) argues that Kilega is a language where A'-opacity does not hold and improper movement is permissible. At issue is whether or not languages like Kilega (or Dinka, as described by van Urk, 2015) pose any special difficulty for the insulation theory that other theories do not face, or if there is some advantage for insulation theory in the way that such languages are accounted for.

Many Eastern and Southern Bantu languages are liberal about what can appear in apparent subject position, insofar as verb agreement in Bantu is an indication of which position is the subject. Bantu languages are noun class languages and Kilega (and Lusaamia) is like other Bantu languages in having rich agreement with distinct paradigms for most of the noun classes in both their singular and plural forms. Moreover, in many Bantu languages, locatives behave like DPs insofar as they belong to a distinct noun class (e.g., c17) and they agree with the verb for c17 when they are in the immediate preverbal position, as illustrated in (55) (all data from Carstens, 2005, unless specified otherwise).

55) 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.'

Also like many other Bantu languages, questions can be formed in situ (as in (56a) or by fronting (56b), but what matters for the present discussion is that the verb shows agreement with the wh-phrase noun class when it appears in immediate preverbal position and the notional subject is post-verbal. The inversion happens in long distance movement as well.

56a) Bábo bíkulu b-á-kás-íl-é mwámí bı´kı´ 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?'

c) Bikı´ bi-a´ -te´nd-ı´le´ ba´na bi-a´ -gu´ l-ı´le´ nina-bo´ ?
8what 8CA-A-say-PERF 2child 8CA-A-buy-PERF mother-their
'What did the children say their mother had bought?

Kilega contrasts with other Bantu languages like Kinande (see, e.g., Schneider-Zioga, 2007) where the subject still precedes and agrees with the verb even when wh-fronting applies.[15] Carstens (2005) analyzes these constructions as involving but C agreement and T agreement and she assumes that the wh-phrase passes through the Spec TP position, then proceeds to the Spec CP position, but only one of the two agreements appears because the agreements would be the same. Since the element satisfying traditional EPP (to fill Spec TP) must agree and an extracted object would be adjoined to vP, it would be outside the EA and be attracted first, followed by attraction to Spec CP, and so the effect will be seen in each clause of a long distance extraction as well, as illustrated in (56c).

The insulation theory analysis might be applied to these facts in one of two ways: Wh-moved elements (not near subjects) must move to the edge of vP if they are to escape the vP in subsequent movement, as required by phase theory. Since the wh-phrase is adjoined above the EA, T sees the wh-phrase and agrees with it. The EA remains in situ, in post-verbal position and the verb does not agree with it (so far, this is Carstens analysis). In English, such movement would fail because the wh-phrase would get two Case assignments (Accusative in vP and Nominative from T) and the EA would not get a Case. Movement to the EPP position by the wh-phrase would not be problematic, because it is still a DP, but double Case assignment at minimum would rule it out.[16] Baker (2003) suggests, however, that Bantu does not require DPs to have Case (particularly the post-verbal notional subject when the Spec TP subject position is filled). Thus there is no reason to insulate the wh-DP when it has moved to the edge of vP. It will agree with T and can satisfy EPP just as any other nominal. There is reason to believe, however, that Case is still a factor in Bantu, at least for Kinande (Schneider-Zioga 2007), and that may be why Kinande has a pattern more like English (agreement is with the notional subject in preverbal position). If so, then by this logic, Kilega stands apart from Kinande in not requiring DPs to have case. Carstens (2005) assumes that Case is still required in Kilega, but it is assigned differently and Obata and Epstein (2010) propose a parameter to this effect. Either way, if T does not assign Nominative, then the wh-phrase is not exposed to double Case assignment and so is eligible to agree with T and move to Spec TP as an uninsulated DP.

A second approach would be to treat EPP as more liberal in languages like Kilega such that an insulated DP can fill the TP subject position (on Carsten's account, on its way to Spec CP). For the second approach, it could be assumed, contrary to the first approach, that double case assignment forces insulation just as in English. Much more investigation is needed to see if either of these approaches if fully viable, but the challenges for insulation theory do not appear exceptional or more complex than any other theory of the A-/A'-distinction would require.[17]

---

[15] Obata and Epstein (2010:22) point out further that wh-fronting in Lusaamia, which shares the basic properties in question with Kilega does not induce WCO, even for wh-fronted matrix questions, which suggests that they are A-movements, but the example they give (provided by Vicki Carstens) shows the subject agreeing with the verb, not the wh-phrase. The facts need much more investigation before any conclusions can be firm.

[16] If double Case assignment is not a motivation for insulation, as suggested in section 3, this motivation does not need to be neutralized in this context.

[17] O&E require a parameter that distinguishes two types of edge features. If Chomsky (2013) and Safir (2010) are right, there are no edge features, and so no parameter of this kind can be stated (as a means of avoiding [uF] at the phase edge, contravening Richards, 2007). See also Miyagawa (2010) for an analysis of the Kilega pattern which appeals to the α-projection he uses to model A-movement effects in scrambling contexts. Similar accommodations to an insulation approach could be used to support an insulation analysis.

Given the spare nature of the facts available, it is not possible to be certain of all the factors that must be taken into account if an insulation-based version of Carstens' analysis is to be upheld in its particulars, but the key point is that neither Carsten's account nor Miyagawa's account of scrambling (which Miyagawa also extends and adapts for the Kilega pattern) are inconsistent with an insulation-theory account of why IMB does not hold in Kilega, though the question deserves much more study.

Another language where it appears that improper movement is not ruled out is Dinka as discussed by van Urk (2015). In Dinka, it looks like all the properties of A-movement hold of movement of any sort in Dinka and that long distance movement is possible proceeding cyclically where the voice marking and agreement indicates that the preverbal position has been passed through by long distance movement, moving objects and lower subjects past matrix subjects, for example. The moved element is the one agreed with in both clauses. Van Urk (2015: 16) summarizes the situation as follows:

> …movement in Dinka behaves for the purposes of binding like A-movement, even though it has the locality profile of A'-movement. I showed that long-distance movement in Dinka differs from long-distance movement in other languages in being accompanied by phi- agreement. The resulting mixed behavior then provides an argument that the features that distinguish A-movement derive from the Agree relation that it involves.

Van Urk takes the Dinka pattern to be a function of the inseparable relationship between phi-features and what others have called edge features (that trigger A'-movement), whereas these features are separated in languages that show the A-/A'-distinction. He takes this to be evidence for the 'featural' approach to the A-/A'-distinction.

From the perspective of the insulation approach, it appears that insulation never occurs in Dinka and if so, all the effects described would be predicted. Any DP moved to the left periphery of vP would be closer to T than the Spec vP and would be the goal of Agree. It would appear that insulation in this language, if it occurs, is somehow stripped or leads to ill-formedness where it occurs. Some evidence for this is seen in cases where a PP is fronted. Van Urk points out that when PPs are fronted in Dinka, the preposition must be suppressed. In (57b) (van Urk, 105), a topicalized PP becomes nominal.[missing diacritics under vowels and over ɛ for Dinka examples to be added in final draft].

57a) Bòl à-thɛɛt nè tòony.
Bol 3s-cook.SV P pot
'Bol is cooking with a pot.'

b) Tòony à thɛɛtè Bôl
pot 3s-cook.OBLV Bol.GEN
'A pot, Bol is cooking with.'

A similar stripping effect is seen when notional subjects, marked with Genitive when they are not preverbal, lose their Genitive Case in preverbal position (which van Urk takes to mean that they bear Absolutive Case).

58a) Àyèn à-cé cuîin câam nè păal
Ayen 3s-PRF.SC food eat.NF P knife
'Ayen has eaten food with a knife.'

b) Cuîin à-cíi Áyèn câam nè păal
food 3s-PRF-.OV Ayen.GEN eat.NF P knife
'Food, Ayen has eaten with a knife.'

Van Urk does not offer any explanation of this effect, but it would make sense if there is no position in Dinka that permits a non-nominal specifier, in which case there could be no criterial satisfaction if a nominal is insulated. The ban against double Case assignment (if this ban is necessary, see fn.16) would have to be rethought in terms of stripping P, perhaps

abstract in the Genitive Case, but this alternative approach seems plausible and does not appear to add an additional assumption in Dinka beyond what van Urk must say about the necessary coexistence of phi-features and edge features (here, criterial satisfaction requires a nominal) and what he doesn't say about the P-stripping phenomenon.

Clearly more work needs to be done on languages where IMB is not in force, including languages that van Urk goes on to discuss where some of the same effects hold. At this point, however, it is not clear that there are effects that require more of the insulation account as opposed to others.

8.0 Competing explanations of the A-/A'-distinction

The A-A'-distinction was defined in terms of the position of the landing site for most of the past 30 years and since then there has been a tendency to define it in terms of the different sorts of features that trigger the movement. As mentioned at the outset, merely isolating a factor that distinguishes the two kinds of constructions without explaining why the distinction arises at all, or why it involves the particular distinctions that it does, is not amount to an explanation. However, Obata and Epstein (2010) (henceforth, O&E) and van Urk (2015) both propose to derive the A-/A-distinction from the features that trigger movement and their proposals address the explanatory criteria that have been laid out here.

Obata and Epstein (2010) (henceforth, O&E) take as a point of departure two key ideas introduced by Chomsky (2007, 2008): First, the proposal that phi-features and uninterpretable features are introduced on phase heads (C and v, primarily) and that those features must be inherited by T and second, that criterial freezing (no movement out of a criterial position) applies to subjects, on the assumption that the subject position (here, Spec TP) is a criterial position. A consequence of the second idea is that wh-movement cannot move from subject position to Spec CP; rather the Spec vP position must be the launching point both for movement to Spec TP and movement to Spec CP. This means that internal Merge applies to the same position twice, such that it introduces a copy of the Spec vP in Spec TP position, and a separate internal Merge introduces a copy of Spec vP in Spec CP position. Without going into the merits of the analysis,[18] O&E suggest that when internal Merge is triggered by Agree with T, uninterpretable features ([uF]) and phi-features are attracted and [uF] are removed from the DP, but when internal Merge is triggered by interrogative C in these cases, only the interrogative Q features on the wh-DP are attracted and removed. The feature set of the wh-phrase in Spec vP are thus split and removed separately in separate positions. This feature split characterizes the A-A'-distinction in their story, but they argue further that the distinction arises (is derived) because a restriction on phase edges precludes [uF]s, Case features in this instance, from occurring on a phase head at the close of a phase (at the moment of transfer) (following Richards, 2007)). This is the same assumption that forces features to be inherited from the phase head downward onto T in the first place, but it also means that only edge features, like interrogative Q, are left to attract the constituents that lead to A'-binding, and that the internally Merged copy in Spec CP thus lacks phi-features. They choose to then characterize the difference as follows.

55) An A-position is a category bearing phi-features, whereas an A'-position is a category lacking phi-features.(O&E, 2010:12)

If the internally Merged copy of wh-DP in Spec CP lacks phi-features and [uF], then it cannot be subsequently moved into a higher Spec TP position because it cannot satisfy the [uF] on matrix T. In other words, feature-splitting in the O&E account derives the IMB.

[18] See Haegeman and van Koppen (2012) for a critique of the feature inheritance approach and, in particular, for convincing evidence that C-agreement and T-agreement are distinct. Criterial Freezing is not assumed in this paper.

Although much of their work is oriented toward the IMB, O&E have some speculations about why the distinction they propose would have just the effects they describe and not others, but these do not extend to binding, pied piping, reconstruction or parasitic gaps. Insofar as binding mismatch could be captured in their theory by the absence of phi-features on phrases attracted by probes with edge features and not phi-features, it might be possible for them to extend their account to anaphor-binding, WCO and parasitic gaps in the same fashion that the insulation theory addresses those phenomena (mismatch). As proposed here, the binding relations between higher and lower copies will require some reinterpretation at the interface in any case, and so the same issue of distinguishing binding by operators containing copies from those that do not is thus also available to their approach in principle. It is not at all apparent how the feature-splitting approach would simultaneously make the reconstruction distinction between A- and A'-movements and the distinction between adjuncts and complements for A'-constructions, and so it would necessarily be less general. Much depends on the longevity of the assumptions that phi-features and [uF] are only introduced on phase heads and that [uF] cannot be in the edge during transfer. Both of these highly technical assumptions lose their force if Chomsky (2013) and Safir (2010) are correct that Merge should apply freely without triggers.

Van Urk (2015)[19] attempts to address all of the A-/A'-distinctions discussed here and does so without appeal to feature splitting, but instead he argues that

> the factor that distinguishes A-movement from other movements is that it is driven by obligatoryfeatures of nominals, such as phi-features. In contrast, dependencies like *wh*-movement and relativization are triggered "by optional properties of lexical items, such as *Wh* or Top(ic) (p.31).

This bears some similarity to the O&E distinction, but it is executed differently. The pied-piping available to A'-structures is driven by non-nominal-specific features and thus permits non-DP movements, following the general outline of Cable's theory. Van Urk adopts the choice function account of WCO, with the same advantages and disadvantages discussed in section 4.2. Broadly speaking then, van Urk's approach is also based on binding mismatch and the possibility of binding match is also used in his approach to parasitic gap licensing, though his execution within the choice function account is different from that proposed here. The sharpest contrast between van Urk's account and the one proposed here concerns reconstruction effects, for which he essentially adopts the Takahashi and Hulsey approach, which appeals to Case assignment rather than insulation. The Takahashi and Hulsey approach assumes late attachment without ever explaining why it should be allowed to apply and violate Extension, and this may be the most crucial distinction that divides feature-based accounts from insulation-based ones.

The insulation account only introduces one new and crucial assumption to derive all the A-/A'-distinctions, namely, the PNC allows penultimate Merge. The distribution and timing of insulation follows from free Merge as it interacts with phases, Agree and Criterial Satisfaction, etc. Van Urk's comprehensive approach is the most complete competitor to the insulation account so far, but it falls short by introducing an additional assumption which is precisely the one that the insulation approach uses to derive everything else.

[19] Van Urk's dissertation is reached me as I was completing this manuscript , so his account did not influence mine, though it forms an interesting point of comparison to the insulation analysis and suggests new directions for research bearing on which approach is better.

9.0 Summary

Table 1: ***The A-/A'- contrasts***

	Landing site binds anaphors	Can induce WCO	Can induce parasitic gaps	Allows pied-piping	Case can be assigned to landing site	Can agree with T in landing site	Permits reconstruction
A'	-	+	+	+	-/%	-	Mostly yes
A	+	-	-	-	+	+	Mostly no

Every distinction has been addressed by the proposal that 'A'-constructions' are just those constructions where movement requires insulation after theta-assignment and A-constructions are just those constructions, whether involving movement or not, that do not induce insulation of DP. Insulated [Q/P DP] cannot bind anaphors or pronouns, but can support parasitic gaps, while uninsulated DP movement can bind anaphors and pronouns but not operators (which might bind parasitic gaps) under the 'like binds like' assumption. The placement of the Q/P-head allows for pied piping of insulated phrases, but not simple DPs. A'-opacity, which typically drives the timing of insulation, follows from Q/P being an intervener for Agree that would assign Case or trigger agreement with the D of DP. Other motivations for insulation involve interpretation (adding a particle with a Q-feature) or avoiding theta-assignment conundrums. A constituent on the edge of a phase must be insulated before the phase closes, or else penultimate Merge cannot apply subsequently. This means that insulation must apply blindly and freely since the heads that may require it are not yet in the derivation (part of the argument for free Merge and against triggered movement). Improper movement derives from insulation which makes [Q/P-DP] an inappropriate satisfier of EPP, except in languages where insulation of wh-movement is not required or EPP can be satisfied by a wider range of constituents, as in Kilega, where A'-opacity does not hold.

Although issues remain, the possibility of truly eliminating the A-/A'-distinction from linguistic theory, while predicting the properties of the two descriptively different classes, now seems feasible. No distinction between edge feature types or types of criterial positions is necessary, nor any notion of 'A'-position', since the actual distinction is about the internal structure of phrases that predicts their external behavior. Only the PNC has been introduced, which permits penultimate Merge in a system where Merge is free. The existence of penultimate Merge is justified by phenomena analyzed using late attachment, tucking in and head to head movement. If penultimate Merge is possible, insulation is freely available and is sufficient to eliminate the A-/A'-distinction from syntactic theory.

References

Baker, Mark. 1988. *Incorporation: A theory of grammatical function changing.* Chicago: Chicago University Press.

Baker, Mark. 2003. Agreement, dislocation and partial configurationality. In A. Carnie, H. Harley and M. Willie (eds.) *Formal Approaches to Function.* 107-134. Amsterdam: John Benjamins.

Bayer, Josef. 1984. COMP in Bavarian syntax. *Linguistic Review* 3: 209-274.

Bhatt, Rajesh and Roumyana Pancheva. 2004. Late merger of degree clauses. *Linguistic Inquiry*: 1-45.

Bobaljik, Jonathan. 2002. A-chains at the PF-interface: Copies and 'covert' movement. *Natural Language and Linguistic Theory* 20, 197–267.

Bobaljik, Jonathan and Susi Wurmbrand. 2005. The domain of agreement. *Natural Language*

and Linguistic Theory 43:371-321.
Cable, Seth. 2010. Against the existence of pied-piping: Evidence from Tlingit. *Linguistic Inquiry* 41: 563-594.
Carstens, 2005. Agree and EPP in Bantu. *Natural Language and Linguistic Theory*: 23:219-279.
Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Dordrecht: Foris.
Chomsky, Noam. 1986. *Barriers*. Cambridge: MIT Press.
Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge, MA: MIT Press.
Chomsky, N. (2000). 'Minimalist Inquiries: The Framework.' In R. Martin, D. Michaels and J. Uriagereka (eds), *Step by Step: Essays on Minimalist Syntax in honor of Howard Lasnik*. Cambridge, MA: The MIT Press.
Chomsky, N. (2001). Derivation by Phase. In M. Kenstowicz, ed., *Ken Hale: A Life in Language*, 1-54. Cambridge, MA: The MIT Press.
Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti, ed., *Structures and Beyond: The Cartography of Syntactic Structures*, vol. 3, 104–131. New York: Oxford University Press.
Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax–Semantics* (Studies in Generative Grammar 89), 1-29. Berlin: Mouton de Gruyter.
Chomsky, Noam. 2008. On Phases. In R. Freidin et. al. eds. *Foundational Issues in Linguistic Theory*, 291-321. Cambridge: MIT Press.
Chomsky, Noam. 2013. Problems of Projection. *Lingua* 130:33-49.
Déprez, Viviane. 1989. On the typology of syntactic positions and the nature of Chans: Move α to the specifier of functional projections.
Dayal, V., 2002. Single-pair versus multiple-pair answers: WH-in situ and scope. *Linguistic Inquiry* 33, 512–520.
Fox, Danny. 2003. On logical form. In Randall Hendrick (ed.), *Minimalist Syntax*, 82–123. Malden, MA: Blackwell.
Fox, Danny and Jon Nissenbaum. 1999. Extraposition and scope. A case for covert QR. In Sonya Bird, Andrew Carnie, Jason D. Haugen, and Peter Norquest, eds., *WCCFL 18: Proceedings of the 18th West Coast Conference in Linguistics*, 132-144. Somerville, MA: Cascadilla Press.
Frampton, John and Sam Gutmann. 2002. Crash-proof syntax. In *Derivations and explanations in the minimalist program*, eds. Samuel David Epstein and T. Daniel Seely, 80-105. Oxford: Blackwell Publishing.
Freidin, Robert. 1986. Fundamental issues in the theory of binding. In Barbara Lust, ed., *Studies in the acquisition of anaphora*, 1:151-198. Dordrecht: Reidel.
Haegeman, Liliane. 1995. *The Syntax of Negation*. Cambridge: Cambridge University Press.
Haegeman, Liliane and Marjo van Koppen. Complementizer agreement and the relation between C^o and T^o. *Linguistic Inquiry* 43:441-454.
Hicks, Glyn. 2009. Tough constructions and their derivation. *Linguistic Inquiry* 40:535-536.
Higginbotham, James. 1983. Logical Form, binding and nominals. *Linguistic Inquiry* 14:395-420
Higginbotham, James. 1985. On Semantics. *Linguistic Inquiry* 16:547-593.
Kayne, Richard. 1983. *Connectedness and binary branching*. Dordrecht: Foris Publications.
Koopman, Hilda and Dominique Sportiche. 1983. Variables and the Bijection Principle. *The Linguistic Review* 2: 139-160.
Lasnik, Howard, and Tim Stowell. 1991. Weakest Crossover. *Linguistic Inquiry* 22:687-720.

Lebeaux, David. 1990. Relative clauses, licensing, and the nature of derivation. In J. Carter, et.al., eds. *NELS 20*, 318–332. GLSA, U. of Massachusetts, Amherst.
Mahajan, Anoop. 1990. The A/A' distinction and movement theory. Ph.D. dissertation, MIT.
May, Robert. 1985. *Logical Form: It's structure and derivation*. Cambridge: MIT Press.
McCloskey, James. 2002. Resumption, successive cyclicity and the locality of operations. In Samuel David Epstein & T. Daniel Seely (eds.), *Derivation and the Explanation in the Minimalist Program*, 184–226. Malden, MA: Blackwell.
Miyagawa, Shigeto. 2010. *Why agree? Why move?* Cambridge: MIT Press.
Obata, Miki and Samuel D. Epstein. 2011. Feature-splitting internal merge: Improper movement, intervention and the A/A'-distinction. *Syntax* 13:1-26.
Reinhart, Tanya. 1983. *Anaphora and semantic interpretation*. Chicago: University of Chicago Press.
Rezac, Milan. 2003. The fine structure of cyclic Agree. *Syntax* 6.2:156-182.
Richards, Norvin. 1999. Featural cyclicity and the ordering of multiple specifiers. In David Samuel Epstein and Norbert Hornstein, eds. *Working Minimalism,* 127-158. Cambridge: MIT Press.
Richards, M. 2007. On feature inheritance: An argument from the Phase Impenetrability Condition. *Linguistic Inquiry* 38:563–572.
Roberts, Ian. 2010. *Agreement and Head Movement*. Cambridge. MIT Press.
Rizzi, Luigi. 2006. On the form of chains: Criterial positions and ECP effects. In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *WH-Movement: Moving On*, 97–133. Cambridge, MA: MIT Press.
Safir, Ken. 1984. Multiple variable binding. *Linguistic Inquiry* 15:603-638.
Safir, Ken. 1985. Missing subjects in German. In Jindrich Toman, ed., *Studies in German grammar*, 193-229. Dordrech: Foris Publications.
Safir, Ken. 1999. Vehicle change and reconstruction in A'-chains. *Linguistic Inquiry* 30:587-620.
Safir, Ken. 2004. *The syntax of (in)dependence*. Cambridge: MIT Press.
Safir, Ken. 2010. Viable syntax: Rethinking minimalist architecture. Biolinguistics 4.1. http://www.biolinguistics.eu/index.php/biolinguistics/issue/view/15/showToc.
Safir, Ken and Mamadou Bassene. 2015. The Morphosyntax of Verb Stems in Jóola Eegimaa: Stem-internal movement and compositional interpretation, ms., Rutgers University.
Saito, Mamoru. 2003. A derivational approach to the interpretation of scrambling chains. *Lingua* 113:481-518.
Schneider-Zioga. Patricia. 2007. Anti-agreement, anti-locality, and minimality: The syntax of dislocated subjects. *Natural Language and Linguistic Theory* 25:403-446.
Sauerland, Uli. 1998. The meaning of chains. Ph.D. dissertation. MIT, Cambridge, MA.
Sauerland, Uli. 2004. The interpretation of traces. *Natural Language Semantics* 12: 63-124.
Takahashi, Shoichi and Sarah Hulsey. 2009. Wholesale late merger: Beyond the A/A'-distinction. *Linguistic Inquiry* 40:387-426.
Van Urk, Coppe. 2015. A uniform syntax for phrasal movement: A case study of Dinka Bor. MIT Ph.D. dissertation.
Webelhuth, Gert. 1992. *Principles and parameters of syntactic saturation.* Oxford: Oxford University Press.