Th[1]e 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, then

1a) 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. One analytic strategy enabled by this change is that once a wh-phrase DP moves, a conspiracy of factors insures that it is 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 prepositional head to the moved DP, 'insulates' that DP from Case and agreement, but alters the relations it can have with 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, the Peak Novelty Condition proposed here, permits them, so the possibility to Merge an insulator to a moved DP is no addition to the theory, but rather is a consequence of what Merge operations are available. Although insulation is optional, its timing and distribution is limited by interface requirements. The resulting distribution of insulated and uninsulated expressions, as well as asymmetries in reconstruction that also come about as a result of the Peak Novelty Condition, 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 Condition, 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 for late merger of an insulating head, are explored in the remainder of section 3, including consequences for A'-opacity. Antecedent relations affected by insulation are used to derive A-/A'-contrasts in anaphor binding, parasitic gap licensing, and part of bound variable interpretation in section 4. Asymmetries in reconstruction, including the remaining bound variable pronoun asymmetry, 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 have appealed to the A-/A'-distinction, first so-designated by Chomsky (1981:184) to distinguish two construction types in the following ways.

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. Must reconstruct
A'	-/%-	-	+	+	-	+	+	Mostly yes
A	+	+	-	-	+	-	-	no

Examples (2-9) illustrate these contrasts as they are found in English.

2a)*I don't know whom arrested John.

b) He was arrested.

3a) How many boys is/*are it clear that Mary likes?

b) The boys happen(*s) to be guilty.

4a) **The accountant* seems that it is beginning *t* to find fraud.

b) *What sort of fraud* is the accountant beginning to find *t*?

5a) About whom did John speak?

b) *About Mary was spoken (compare: Mary was spoken about)

6a) **Which boys* did *each other's* brothers trust?

b) *The boys* seem to *each other's* brothers *t* to be greedy.

7a) *Who* did John trust *t* before he spoke to *pg*?

b) **Mary* seemed *t* to be happy before John spoke to *pg*?

8a)?**Who* does *his* mother love?

b) *Everyone* seems to *his* mother *t* to be a good boy.

9a) 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 (3a,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 (5a) permits pied-piping whereas passive does not. Examples (6a,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 (7a), 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 (7b). Wh-question movement induces weak crossover, where the pronoun *his* cannot be bound by *who* in (8a), but in a subject-raising structure like (8b), 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, 1998, Takahashi and Hulsey, 2009), as indicated in (9a,b), where the reconstructed reading of (9b) 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 (9c) as opposed to (9d).

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'.

10) ***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 (11a) and (11b) 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 (11b), even though the right result seems to be achieved.

11a)*[$_{TP}$*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) (1995:62): 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. Although 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 (12) (subject-auxiliary inversion is orthogonal to our claims), 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 (12d), rather than the desired outcome, which is (12d').

12a) $[_{vP}$ 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. For one, two Cases would be assigned to the same DP: [wh-DO], which would get ACC Case in the vP phase and NOM in the C phase.

13) No DP can be assigned more than one Case.[2]

Second, if [wh-DO] gets Nominative Case, then EA gets no Case, if Nominative assignment is a one-one relation, as is typically assumed.

14) Case Filter: *DP if it has no Case.[3]

If T agreement with a DP is sensitive only to Nominative Case, then [wh-DO], which bears Accusative, would not be an appropriate agreement partner. Failure of T to agree with anything may or may not be fatal, depending on other theoretical assumptions, so I will not rely on it in what follows. Thus it would appear that if Merge could generate (12d'), it would be ruled out by one or all three of the factors just described.

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 (again, for a sentence like *Who did John praise?*)

15a) $[_{vP}$ 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]]]]]]

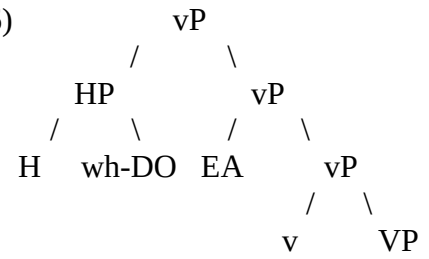
e) $[_{TP}$ EA [T $[_{vP}$ $[_{HP}$ H $[_{DP}$ wh-DO]] $[_{vP}$ EA [v [V $[_{DP}$ wh-DO]]]]]]]

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 (16), where (15c) is diagrammed, the introduction of H by Merge is a violation of Chomsky's (1995) Extension Condition (which I formulate as in (17)), because H is not merged to the undominated node of the tree (which is the highest vP), but is merged instead lower than that.

16)



17) **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 part of the key to the A-/A'-distinction.

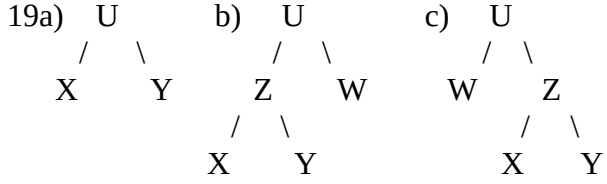
The point of A'-opacity is that Case and agreement do not see an intervening insulated wh-phrase, but regular DPs are indeed visible to agreeing or Case-assigning heads. Thus the reason that A'-moved elements do not get Case in their landing site positions is that they are not DPs after insulation. DPs 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, violating (13), 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, Bhatt and Pancheva, 2004, Takahashi and Hulsey, 2009), tucking in (Richards, 1999), antecedent contained deletion (Fox and Nissenbaum, 1999), Q-insertion (Cable, 2010) 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, 2016, 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):

18) **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.

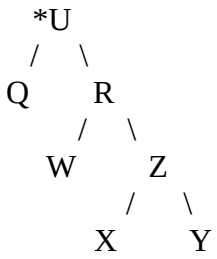
The PNC permits Merge to form (16a-c) where 'X' is merged to the structure by M_i.[4]



When the undominated node is new[5] (the conventional case) after Merge of X (19a), PNC is satisfied. When X is adjoined just below the undominated node (19b,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 (19b) and (19c) model late attachment (aka ‘late merger’) and insulation, respectively. When internal Merge has applied, that is when X has a copy in W in (19b), or where X has a copy in Y, as in (19c), then the structure as presented models head-to-head adjunction in (19b) and tucking in (19c). However, Merge cannot apply to X as in (20).

20)



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

PNC now permits head movement by adjunction to a higher head, as illustrated schematically in (19b), it permits tucking in, illustrated in (19c), as proposed by Richards (1999) to account for superiority effects (but see also Safir and Bassene, 2016, who, like Richards, 1999, apply tucking in to clitic movement), and it permits late attachment, which is proposed to account for anti-reconstruction effects, as in (19c).

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 penultimate Merge also accounts for contrasts 4-8 in Table 1.

Notice that we could stipulate that all A'-movements are insulated to distinguish A- vs. 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 arise as a consequence (as is proposed by Rezac). However, even if insulation were to account for all the distinctions in Table 1 (and I will show that it is only most of the story), the distinction is still not explained if the feature that determines the difference does not itself follow from general principles or relies on stipulations not independently required. 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. 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, but not always, 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 only serves to distinguish 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 – 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 (21).

21) **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.

22) **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. 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. Criterial assignment is not, however, a relation which distinguishes A-positions from A'-positions as it is applied here because it is not unique to operators; it also applies to determine the EA in any sentence that has one. If the first nonterminal adjoined above v and its complement is not an occurrence (copy), then it is the EA for that v.[6] Whether or not the constituent assigned the criterial domain in this way is thematically appropriate is a matter for the semantic component, as it is in all current generative theories. The restriction in (22)

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, descriptively, Spec TP, must be filled, a holdover from the old S→NP VP phrase structure rule that has been known as the Extended Projection Principle.

23) EPP: Spec TP must be filled.

Rizzi has suggested that the subject position bears an aboutness relation to its domain, and if so, there is a criterial relation that assigns a semantic value to the domain assignment, but the mechanism in (21) does not legislate the content of a criterial relation and no such criterial relation is assumed here. In 3.5, EPP will be reformulated as an S-M interface interface requirement, but neither formulation requires Spec TP to be of one syntactic category type or another. If so, an insulated DP should be just as appropriate category-wise to satisfy EPP as in (23) as any DP would be. Nonetheless, it will be argued in section 3.5 that EPP should be formulated differently, following Landau (2007), with the result that insulated DPs do not satisfy EPP for a principled reason distinct from what syntactic category the insulated DP is assumed to be.

If insulated DPs cannot satisfy EPP, the IMB follows without further assumptions. Movement of an insulated DP from vP-adjoined position to subject position would be ruled out by EPP. The EPP also rules out insulation of a DP after it has moved to subject position, a possibility free Merge allows, but that is in this way constrained. If for some reason, insulation is not required for a DP adjoined to vP, 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 in (1a,b), 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 (before T enters the derivation). In the derivation of (15), repeated below, notice that insulation, which is optional, occurs in (15c) because that is the only point in the derivation where the PNC will permit it.

15a) $[_{vP}$ 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 (15d) before the wh-phrase is insulated in (15c), 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 also has 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 will 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.[7] 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*). Conspiracies ruling out other undesired derivations are addressed in other sections.

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. The appeal of this strategy diminishes if it ultimately places too much burden on the interfaces in the form of stipulated filters, but no special burden is obvious so far. Any theory must have an account of why quantifier 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 contexts and not others, any theory will need some way to distinguish those contexts.

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 part of 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 as 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? Although the category of it is underdetermined, I will assume it is a preposition without semantic content. If it has semantic content, then we might suppose the insulating head is motivated not (uniquely) by A'-opacity, but to satisfy a criterial requirement of the specifier where it is assigned a criterial relation. Moreover, there may be languages where the insulating head has morphology. Cable's (2010) analysis of Tlingit suggests that a head introduced after movement, one which has morphology, also has semantic content. While such a case provides more motivation for the PNC, A'-opacity effects will be obscured in Tlingit because of the independently motivated head. Although I discuss Cable's proposal in the next section with respect to the pied-piping property, I will otherwise focus on languages where only A'-opacity conspiracies motivate insulation.

3.4 Pied-piping and so-called A'-movement

Within a Free Merge approach pied-piping is simply expected where it is not blocked by other factors. No trigger needs to be inserted to identify that something larger than a DP can be moved, nor need there be any head attached to the 'goal' that makes it eligible for movement, so the challenge becomes one of restricting over-generation. Whatever moves

must end up in a position where it can be interpreted, however. For example, a phrase fronted to Spec CP must have something in it that satisfies the criterial property of C, namely a wh-phrase, but if so, how does C 'locate' the criterial property in the pied-piped phrase?

Safir (1986: 677-678) argues that there is covert wh-fronting within the pied piped phrase to the edge of the pied-piped constituent. Citing a distinction pointed out by Kayne (1983: 283 and fn. 17), Safir shows that multiple wh-phrases corresponding to the relative clause nucleus are possible within the pied-piped phrase, but not if one c-commands the other (while heavy pied-piping is awkward in a restrictive relative like (24b), the distinction between (24a) and (24b) is sharp).

24a)*No boy whose pictures of whom none of us is allowed to see will be allowed to join our club.

b) No boy whose mother's pictures of whom none of us is allowed to see will be allowed to join our club.

The failure of (24a) is an instance of strong crossover, and we take this example to be evidence for covert movement. If the wh-phrases within the pied-piped phrase must both front to the edge of the pied-piped phrase, it is plausible they are then close enough for predication to take place. Thus pied piping is unrestricted by the application of Merge, but pied-piped constituents must be interpretable where they land, and that may depend on further movements.

The account of pied-piping in a free Merge approach advocated here contrasts with the trigger-based account of Cable (2010). He argues that all A'-moved constituents in Tlingit are complements of an overt Q-head. That Q-head is the goal C probes for and that moves QP to Spec CP. The Q-particle *sá* in Tlingit, as illustrated in (25a,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 (25c) (p.570)

25a) 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 inserted head is null, as in (26) (see also Safir, 2010).

26a) Whose father's cousin's uncle did you meet at the party?

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

As Cable points out, the penultimately merged quantificational head also derives the pied-piping property of the A-A'-distinction on the assumption that QP is the goal of a probing feature in C that triggers the movement. Cable (2010: 574 fn.12) assumes there is a special constraint that blocks *sá* in situ, but if Q is inserted too early, e.g., in a thematic position, it would block Case and thematic assignment (and if moved, it would be the trace of a QP, so movement does not change this). Thus it would appear that Q would have to be inserted in the course of a derivation (by penultimate Merge) for this independent reason. The key motivation for introducing Q in his account, however, is that only QP satisfies the criterial property of C. In Tlingit, moreover, the insulating head is required even if the fronted

constituent is not a nominal, and for those cases, insulation is not independently motivated by A'-opacity as proposed here.

We have already seen that English does not require the insulating head to make any semantic contribution, and insofar as Cable is right about what the Q-head of Tlingit contributes, most of the effects of insulation in languages like Tlingit would be overdetermined by the necessary overt Q, and thus masked. Moreover, within a free Merge approach, it is not necessary to appeal to a head that forms a phrase that can be probed for and pied-piped, and so there is no motivation to extend Cable's account of Q as semantically contentful to the insulating head in other languages. Nor is there any need to assume that 'A'-binders' are different from 'A-binders' in any semantic sense, at least not in the general case. For what is at stake in this essay, no semantically contentful insulating head is needed to explain pied-piping nor is it necessary for the reduction of the A- vs. A'-distinction to an epiphenomenon.

3.5 More on improper movement and EPP

Where must insulation apply when a subject position is moved? 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. 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.[8]

27a) Who is Gladys glad (*that) ate the fish?

b) [$_{CP}$ [$_{QP}$ *Q who*] [$_{TP}$ *Gladys* T [$_{vP}$ [$_{QP}$ Q *who*] [$_{vP}$ *Gladys* be glad

[$_{CP}$ [$_{DP}$ *who*] [(that) [$_{TP}$ [$_{DP}$ *who*] ate the fish]]]]]]]

Notice, however, that this derivation would still be banned by both the Caselessness of *Gladys* and Double Case assignment (13). *Who* would be assigned Nominative in passing through Spec TP position of the subordinate clause to [wh-DP] and then Nominative again in matrix vP-adjoined position unless the insulating head is inserted in the matrix vP phase.

Where there is no thematic subject, hence no Caselessness issue, Double Case assignment still rules out a case of improper movement that would generate (28a) by the derivation in (28b).

28a)*Who is important that ate the fish?

b) [$_{CP}$ [$_{DP}$ *who*] [$_{TP}$ *who* T [$_{vP}$ [$_{DP}$ *who*] [$_{vP}$ be important

[$_{CP}$ [$_{DP}$ *who*] [(that) [$_{TP}$ [$_{DP}$ *who*] ate the fish]]]]]]]

Unless an insulating head is inserted to block it, *who* adjoined to vP would be assigned NOM a second time.[9]

Yet another improper movement derivation that would result in (28a) is ruled out by the same mechanism. Suppose that *who* is subextracted from Spec CP after insulation has taken place, as in (29).

29) [$_{CP}$ [$_{DP}$ *who*] [$_{TP}$ [$_{DP}$ *who*] T [$_{vP}$ [$_{DP}$ *who*] [$_{vP}$ be important

[$_{CP}$ [$_{QP}$ Q *who*] [(that) [$_{TP}$ [$_{DP}$ *who*] ate the fish]]]]]]]

This derivation in (29) would strand the insulator in subordinate Spec CP, but since it is silent in English, the pronunciation would come out the same as (28a); This result is still ruled out by Double Case Assignment applying in the matrix clause. However, there is reason to assume that heads cannot be stranded from their complements in intermediate positions in general, as Postal (1972) pointed out for examples like (30).[10]

30a)*Who did John think to (that) Mary should talk?

b) *Who(m)* did John think [$_{CP}$ *to* ~~*who(m)*~~ (that) [$_{TP}$ Mary should talk *to who(m)*]

Although this derivation might also be ruled out by Double Case Assignment, there appears to be a broader ban on stranding prepositions in intermediate positions. The ban in question, whatever it is, does not appear to be 'A'-specific', which is important if we are to avoid obliquely appealing to the A-/A'-distinction. Since the Spec TP position is always a position that is moved to, we would expect, if a PP is permitted there, then the P could not be stranded. As discussed in the next section, some PPs fill the Spec TP position, but when they do, P cannot be stranded by further extraction unless the PP is itself embedded,[11] ((31e) is based on an example from Kayne, 1984).

31a) The cat thinks (that) under the bed seems to be a nice place to hide.

b) the bed under which the cat thinks is a nice place to hide

c)*Which bed does that cat think (that) under is a nice place to hide?

d)?Who did the lawyers say that to talk to would be a problem?

e)?Who did you wonder which picture of to sell?

The difference between (31d,e) and (31c) indicates that it is stranding of P when it is not part of a larger fronted constituent that the ban applies to. This stranding ban thus does not distinguish A- vs. A'-constructions.

However this ban is accounted for, it will readily extend to the complement of the insulating head if that head is also a P, as suggested, and if so, it rules out an otherwise possible derivation that would undermine my approach (as pointed out by a reviewer).

32a) $[_{vP}$ 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)*[$[_{DP}$ wh-DO]] [T $[_{vP}$ $[_{HP}$ H $[_{DP}$ wh-DO]] $[_{vP}$ EA [v [V $[_{DP}$ wh-DO]]]]]]]

The derivation illustrates the progress of wh-extraction from the DO position and up to (32d) it proceeds exactly as in (15), but in (32e), the DP complement of the insulating head H has moved to Spec TP, satisfying EPP and remaining an uninsulated DP. The wh-DP would avoid double Case assignment because it is insulated in the c-command domain of T and the EA would remain a candidate for Nominative assignment. Only the ban against stranding a preposition when it is the head of a PP that is not in its first-merged position rules this derivation out. Since the ban does not distinguish between so-called A- or A'-positions, no new stipulation is necessary to block (32e).

At this point, it should be clear that the IMB arises from a conspiracy of principles and non-A-/A'-specific restrictions and has no status in the theory. It would therefore be less surprising if a slightly different conspiracy could predict that improper movement is possible. Indeed there is plausibly such a context in English. The generalization that A'-landing sites are not assigned Case has been derived by the necessity to insulate a DP from double Case assignment when it moves from a position where it has already been assigned Case. A reviewer points out that *wager* class verbs seem to require A'-movement to get Case (see Kayne, 1984 and Pesetsky, 1982, for example, and Moulton, 2009, for more recent discussion.).

33a) *No one would wager this candidate to be a fine judge of character.

b) this candidate, who(m) no one would wager to be a fine judge of character…

c) [$_{CP}$ [whom] C [no one would wager [$_{CP}$ ~~whom~~ C [$_{TP}$ ~~whom~~ to be…]]]]

Following Kayne's (1984) analysis, the wh-phrase only is assigned Accusative Case as it passes through Spec CP of the clausal complement of *wager*, as in (33c), assuming that *wager* cannot take a TP complement or it would allow an ECM complement without wh-movement. On the account proposed here, all that must be assumed is that the null C of *wager* class verbs prevents Case or PRO licensing in the complement Spec TP. If so, a wh-

phrase originating in Spec TP (where it has moved to satisfy EPP) could raise uninsulated to Spec CP where it could be assigned Accusative Case. Movement beyond the Spec CP, e.g., to vP adjoined position, would then require insulation to avoid double Case assignment from matrix T. The issue then arises as to why a non-wh-DP could not also raise to Spec CP of the *wager* clausal complement and be assigned Accusative there. Notice, however, that, in trigger theories, it is a stipulation induced by a spurious feature that distinguishes wh-DPs from non-wh-DPs, since the local Spec CP is not the final landing site for the wh-DP. In Chomsky's (2013), approach, the inability of any DP to remain in intermediate Spec CP is instead derived by a failure of labeling unless the two elements agree. If intermediate C does not agree, then movement will be required to resolve the labeling quandary. This works out as long as there is a destination where the DP that moves contributes to agreement with a higher C and stops moving. This will never happen for movement of a DP to Spec CP in the *wager* case, so only DPs that will eventually satisfy a higher criterial position can pass through the local Spec CP. Beyond the stipulations, common to all approaches, that the subject of the CP complement of *wager* cannot get Case or be licensed as PRO, nothing further needs to be said about such derivations. However, and this brings us to improper movement, it is also possible to have passives of *wager*-class complements, though acceptability with different members of the class varies.

34a) *We said the witch to be responsible for the recent influx of mosquitoes.

b) The witch was said to be responsible for the recent influx of mosquitoes.

Such examples could, in fact, be generated by so-called improper movement, because in this context, the Spec CP will not be Casemarked or insulated and can have a destination where it can resolve a labeling quandary, that is, it can move to Spec TP, satisfy EPP, and agree with T in Spec TP. Thus the *wager*-class pattern is predicted by the distribution of insulation.

3.5.1 EPP

I have suggested that EPP will not tolerate an insulated DP in Spec TP position. A reviewer suggests that if this is achieved by specifying that EPP requires not just something in Spec TP, but specifically a DP, then I might be coding the A-A'-distinction into a separate stipulation that keeps A'-binders out of an A-position. For a true explanation, independently necessary properties of EPP must be shown to determine its force..

Landau (2007) offers a potential explanation of the inability of insulated DPs to satisfy EPP that does not rely on restricting EPP satisfaction to DPs. He argues that EPP is a PF selection relation between a head with an EPP feature and the head of its specifier, such that the head of its specifier must not be null.

35) EPP: In $[_{HP}$ ZP $[_{H'}$ $H_{[P]}$ …]], Z must be pronounced.

A head is null if it contains a morpheme that has no phonology or no morpheme at all, but it does not count as null if it is a pronounceable lexical item or a copy of one (see, fn.10). As it applies to EPP on T, Landau's theory predicts, quite straightforwardly, that the null head which insulates DPs in English cannot satisfy EPP in Spec TP position.[12]

Notice, moreover, that the EPP can be satisfied by non-DPs in English particular constructions, such as the 'honorary NPs' described in Safir (1983).[13]

36) (When) does under the bed seem to be a good place to hide?

Subject-Auxiliary inversion is used in (36) as a test for conventional (Spec TP) subjecthood (before inversion) and the PP *under the bed* is raised from the subject position of an infinitival complement of *seem* just as a DP would be. Unless this is made possible by the covert DP structure dominating *under the bed*, it would appear that certain constructions allow EPP to be satisfied by non-DPs, but in this case, the non-DP does not have a null P head. If non-DPs in English are permitted to satisfy EPP, it is not clear why this possibility is so limited, and I have no explanation to offer here, but this restriction, whatever it is, may also account why 'pied-piping' is not generally possible for English 'A-movement'.[14]

Landau assumes, however, that EPP extends to all specifier positions, and if that were so, then an insulated DP could never appear in any specifier position, including Spec CP. I do not assume the Spec CP when C is interrogative is regulated by EPP. Indeed Landau's best evidence for his version of the EPP concerns Spec TP, which is also historically the first and the best motivated EPP effect, since it does not depend on any feature of T (or any criterial relation). Nonetheless, a skeptic of my approach might point out that even if I have eliminated the objection that EPP stipulates DP to derive the A-A'-distinction, saying only Spec TP is susceptible to EPP might be sneaking in a new stipulation that is not independently motivated. Let us consider whether this is so.

The extension of the EPP to all other functional projections typically involves arbitrary stipulation that other functional projections have an EPP feature where movement is needed. For example, if the highest Spec CP position is typically filled to satisfy a criterial requirement, no EPP feature is needed to attract something to fill the specifier position. The assumption that the EPP applies to a functional projection thus needs to be supported with evidence that an expletive element can satisfy EPP in that position, as in the case of English existential *there* in English and other non-pro-drop languages.[15] No other Spec position besides Spec TP seems to have such a requirement, except perhaps the initial position in Germanic V2 inversion structures.[16]

Indeed, the bulk of the evidence that Landau presents to defend his theory supports the EPP as it applies to Spec TP. In particular, they are subject-object contrasts where a missing head is not permitted in Spec TP position, but it is permitted in a direct object. The phenomena drawn from the literature on subject-object asymmetries that he captures under this generalization include restrictions on bare noun subjects in Italian and French, bare negative polarity items in Romance, and sentential subjects (which must have the C-position non-null). However, independent evidence for EPP generally applying to Spec CP for any

sort of C is not convincing (but see fn.16),[17] as pointed out for the stranding of remnants in Spec CP in the last subsection.

This discussion of the motivation for the EPP has been occasioned by the role it plays in blocking improper movement. If the EPP were only motivated to capture the IMB, then the proposed derivation of the A-/A'-distinction would not meet its explanatory burden. It has been demonstrated that the properties of the EPP that give rise to the IMB are independently motivated, including phonological sensitivity to a missing head, showing that EPP does not stipulate the subject must be a DP, and showing that the limited application of EPP (mostly) to Spec TP is also independently motivated and independently detectable by diagnostic tests.[18] Thus the difference between how insulated and uninsulated DPs satisfy EPP does not arise as a stipulation distinguishing A- vs. A'-positions.

3.6 Summary of this section

The PNC is a property of Merge that makes possible penultimate Merge of a head that takes a DP as a complement. I have now introduced three different reasons why insulation might be motivated to occur and they are listed in (37).

37) Insulation after movement

a) prevents wh-DP adjoined to vP from blocking assignment of NOM to Spec vP,

b) prevents double assignment of Case to wh-DP which is already Case-assigned, and

c) may be needed for criterial satisfaction in some positions in some languages (e.g., Tlingit).

It is possible that these motivations may be neutralized in some languages, and if so, the A-/A'-distinction should disappear empirically, as discussed in sections 7 and 8.[19] The timing of insulation, when it is necessary for a result acceptable at the PF interface, must always follow initial Case and theta-assignment and it must at least precede the close of a 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. In this section, I have shown that EPP insures that movement to Spec TP is not compatible with insulation of a DP with a null head. Thus positing a null insulating head, as it interacts with an independently motivated formulation of EPP, derives A'-opacity and the IMB.

4.0 Antecedents and the A-/A'-distinction

The existence of penultimate Merge allows for insulation to take place freely in keeping with free Merge, but the interface conditions constrain its timing and distribution. So far, I have shown how insulation derives A'-opacity for Case and agreement, the pied-piping distinction (largely due to independent factors) and the IMB. The key difference has been that an insulated DP is a non-DP. I now turn to the effects of PNC that derive some of the binding asymmetries embodied in the A-/A'-distinction, including the binding of anaphors, weak crossover, and the licensing of parasitic gaps. Three factors will figure prominently in the explanation to follow: (A) The relation between copies is a symmetric indistinctness relation (at least as generated in syntax), (B) the relation between separate thematic positions is an asymmetric dependent identity relation and (C) an insulated DP is not a proper antecedent for a DP, but might be a proper antecedent for another insulated DP.

4.1 Anaphor binding

The landing site for so-called A-movement can bind an anaphor, but this is not the case for the landing site of so-called A'-binding, as was illustrated in (6), repeated below.

6a)**Which boys* did *each other's* brothers trust?

b) *The boys* seem to *each other's* brothers *t* to be greedy.

Given that *which boys* in (6a) is an insulated DP by the time it arrives in matrix CP, the DP does not c-command *each other*, thanks to the projection that immediately dominates the insulating head and the DP. In (6b), *the boys* has arrived in matrix Spec TP uninsulated and can antecede *each other*, which it c-commands within the same phase.

There is one further scenario in the derivation (6a) to rule out, however, namely, one where *which boys*, binds *each other* at the edge of vP. If we reconsider the derivation of (6a) in schematic form, repeated in (12), there is a point before insulation, represented in (12b), where the wh-DO c-commands EA, yet binding does not succeed.

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***

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. The phase edge is complete at the moment a head is merged to vP (e.g., when T is merged). Recall, however, that if insulation of wh-DP does not take place as in (12c) before the phase closes, then PNC would block it afterward – it is too low for penultimate Merge. Thus at the moment the vP phase is complete, interpretation can only treat the wh-phrase as an HP, not a DP, and the DP within the HP does not c-command the EA and thus cannot bind the EA if it is an anaphor.

The reasoning applied so far does not rule out the possibility that a PP could antecede a DP as an anaphor binder. This is hard to test because it is difficult to find plausible examples where it is the whole PP that is the binder, but (38c) is close to such a case.

38a) In those villages are found people in those villages who are from other villages.

b) *In those villages are found people (in/from) each other.

By the usual assumptions about c-commanding antecedents, *those villages* cannot antecede *each other*, but it does not follow that *in those villages* could not bind *each other* to mean essentially that if one goes to those villages, people who are from (within) others of those

villages can be found. It seems that *each other* is never a PP (e.g., *They sent books *(to) each other*), and no PP can antecede it.

The same reasoning extended to insulated phrases predicts that a wh-phrase is only eligible for anaphora binding in its highest position before insulation. For example, the lowest DP trace/copy of *which students* c-commands *each other's* in (39) but not in (40).

39a) *Which students* did John say spoke to *each other's* mothers?

b) [H [*which students*]] [$_{TP}$ did John [$_{vP}$ [H-*which students*] [$_{vP}$ say

[$_{CP}$ [H-*which students*] [$_{TP}$ [*which students*] spoke to *each other's* mothers]]]]]]

40a)**Which students* did [*each other's* mothers] say spoke to Bill?

b) [H [*Which students*]] [$_{CP}$ did [*each other's* mothers] [$_{vP}$ [H [*which students*]] [$_{vP}$ say

[$_{CP}$ [H [*which students*]] [$_{TP}$ [*which students*] spoke to Bill]]]]]

In (39b), there is no insulation in the lower Spec TP since the wh-phrase must be uninsulated for Case and EPP, but as soon as the wh-phrase moves up to the lower CP, the moved subordinate subject must be insulated before it is 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 (40).

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. In the best case, following our explanatory standards for full derivation of the A-/A'-distinction, the independently determined distribution and timing of insulation, itself made possible as a

consequence of the PNC, should completely derive the existence and distribution of weak crossover (WCO) effects, insofar as WCO is found only where purported A'-binding is found. It turns out, however, that mismatch between insulated DPs and the pronouns they cannot be co-construed with has more to do with a different consequence of the PNC, one relating to reconstruction, but this section sets the stage for that last piece, presented in section 5.

Reinhart (1983) argued that pronouns are only bound by c-commanding antecedents and that WCO arises when a pronoun is not c-commanded by the quantifier (or quantifier extraction site) it 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 (41), where the anchor, ├, indicates the antecedent and the hook, ┘, indicates the element that depends for its value on the antecedent.

41a) Percy praised himself
├──────────┘

b) Alice praised her brother
├──────────┘

The bound pronoun reading in (41b) is also captured by the linking notation. Developing and revising Higginbotham's account, Safir (2004:52) proposes the Independence Principle.

42) **Independence Principle** (INP)

α cannot depend on β if α is (embedded in) a nominal γ and γ c-commands β.

Thus the Higginbotham-Safir account is based on free 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, depending on what other assumptions are made.

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:[20]

43) **Quantifier Dependency Condition** (QDC) (simplified from Safir, 2004)

If X is not the trace of OP, then X cannot depend directly on OP, but only on its trace.

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, 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 that the operator originates in, since only that site is a DP.[21]

Assuming, then, that an insulated DP cannot directly bind a pronoun, it would appear that the bound pronoun must depend on the extraction site. The relations that must hold for a pronoun to be interpreted as a bound variable are illustrated in (44a,b) for strong crossover and for WCO in (45a,b) (where qv=quantifier variable).

44a)*Who did *he* say *t* that Mary loved *qv*

*

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

45a)*_Who_ does [[_his_ mother] love qv]?

b) _Who_ [_qv_ loves [_his_ mother]]?

Insulation leading to binding mismatch is responsible for feeding (44a) and (45a) to the Independence Principle which in turn rules out both dependency patterns. In (44a), the bound reading is excluded because *he* depends on *qv* and *he* c-commands *qv*, while in (45a) [*his* mother] c-commands *qv* and *his* depends on *qv*.

The mismatch approach proposed here recalls Sauerland's (2004) proposal which is based on a semantic type mismatch between wh-phrases. According to Sauerland, wh-phrases assign individuals to properties (i.e., they are choice functions), and so there is a type mismatch with 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 – that is, it could be that the insulation account need not have the burden of explaining the distribution of 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 true resumptive pronouns, where no movement is involved, do not induce WCO, as is well known (see McCloskey, 1990 and Safir, 2004 and references cited there).

46) 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.

Nothing in the insulation account prevents a DP operator from directly binding a pronoun if the operator is not insulated and a DP operator directly generated in Spec CP would not need to be insulated. If the restrictive operator *who* is a choice function, it cannot bind either pronoun in the resumption structure unless the pronouns are shifted to be variables of properties to which individuals are assigned. However, shifting the pronouns would crucially undermine the supposed mismatch for WCO and would further predict that the pronouns would have trace-like structure even where they cannot be extracted (which leads to bad reconstruction effect predictions). The insulation approach is not committed to a choice function analysis, but 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. The full story is not so simple, however.

The QDC, which is necessary to create the right pattern of dependency so that the INP can make the right prediction, does not follow from anything else outside the A-/A'-distinction. More important, it does not follow from Binding Mismatch. As a reviewer points out, Safir (2004, 2014) allows a pronoun to be bound by a DP that does not c-command it. Thus the DP portion of the insulated DP can be an antecedent of a pronoun, even if the whole insulated DP cannot. The same issue arises for secondary WCO examples like those pointed out by Safir (1986), Postal (1993), and further discussed in Safir (2004).

47a) *Whose* father gave the books to *his* mother?

b)*To *whose* mother did *his* father give the books?

In (47a), *whose father* is in the EA position c-commanding *his* before moving to Spec TP and eventually to Spec CP, so neither *his* nor *his mother* is ever in a position where it c-commands *whose father*. No WCO effect is predicted. However, there is a point in the derivation of (47b) where *his father* c-commands *whose mother*. If WCO is sensitive to the existence of any such position, then WCO is expected. However, if the pronoun in (47b) can

be directly dependent on *whose* in its final position, then *his* does not c-command it there and, contrary to fact, there should be no secondary WCO effect. The QDC insures that only the lowest copy of *to whose mother* can be the one containing the antecedent of *his*, and that relationship is then banned by the INP.

What is needed is an independent reason to assume that any copy of *whose* c-commanded by *his father* in (47b) is enough to induce WCO. Recall that we have assumed that the copy relation is, at least initially, one of indistinctness, not an antecedent-dependent relation. Thus it is necessary to assume that if a pronoun P depends on any occurrence (copy) of x, then P depends on every occurrence of x, including the one that *his father* in (47b) c-commands. This is enough to insure that the INP will apply to induce a WCO effect without appeal to the QDC, which is now no longer necessary.

Two further issues now arise.

First, why is there no WCO in the resumptive pronoun construction in (46)? The answer has to do with the difference between copies and pronouns. Pronouns, following Safir (2014), are either native born, that is, they are inserted as pronouns with features that must turn out to match their antecedents, or they are 'D-bound', that is, a form that must be bound by a c-commanding antecedent at some point in a derivation (his 'one true anaphor'). If D-bound is bound within the phase where it is introduced, it will have anaphoric shape, but if it is unbound within that phase, then morphology gives them default pronominal form, consistent with the features that they were inserted with. Thus neither natural born pronouns nor D-bound in pronominal shape are copies and their anaphoric relations do not come about as a result of a relation between copies, but rather by being identity dependent on an antecedent that is not an occurrence (copy) of it. Therefore, in true resumption structures (where the pronoun cannot be shown to be a spell-out of a copy, see, for example, Sichel, 2014), the pronoun is not automatically dependent on copies that are variables of the operator

that binds it. It does not then matter if some other pronoun is bound by the same operator, since there is no copy relation that could create an INP configuration (where one pronoun bound by the operator (indirectly) c-commands and depends on the other. Thus in (46), the pronoun in *after you meet him* is not dependent on *he*. As Demirdache (1991) once suggested, both pronouns bear the same dependency relation to the externally merged *who*, but not to each other.

The second issue concerns the A-/A'-distinction more directly, namely, why doesn't A-movement show WCO effects for examples like (48).

48) *Every boy* seems to *his* mother ~~*every boy*~~ to be a genius.

If *every boy* originates in a position c-commanded by *to his mother*, then the INP should rule it out because *to his mother* c-commands *every boy* in embedded clause subject position. However, the right prediction would be made if the lowest full copy of *every boy* is not accessible for semantic interpretation. Put another way, if there is no reconstruction effect for 'A-movement', but there is for 'A'-movement', then this difference will follow from that one. Our attempt to derive the A-/A'-distinction altogether for WCO thus hinges on an independent account of this reconstruction asymmetry, one that is also a consequence of the PNC, but that discussion is reserved for section 5.

4.3 Parasitic gaps

A condition on antecedent matching might also be part of the explanation why only so-called A'-constructions license parasitic gaps (pgs). As Chomsky (1986) argues, pgs are traces of operators that have moved within the constituent hosting the pg, be it an adjunct clause or a clause within 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 pg must be an insulated operator, and if the licensing operator must bind the parasitic one, this might only be possible if they satisfy antecedent matching,

that is, both are insulated DPs. The structure in (49b) shows that the highest [H [*which mayor*]] c-commands the adjunct (adjoined here to TP for presentational purposes). The bolding in in (49b) shows that the categories of the null adjunct operator and the fronted wh-phrase match. The dependencies involved for (49a) are illustrated in (49c).

49a) 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*]]]]]]]

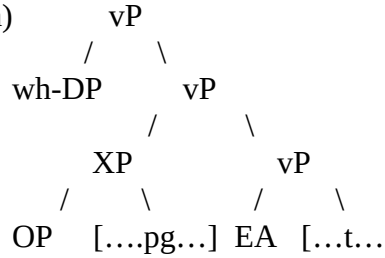
c) [*which mayor* [did John praise *t*]] [*OP* before PRO talking to *pg*]],…

(*OP* depends on *which mayor*)

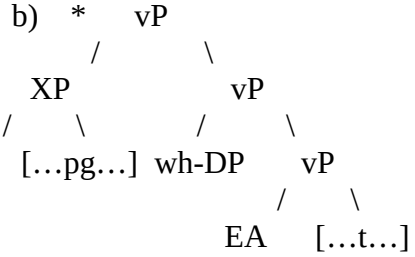
Not every movement that requires insulation supports pgs, but uninsulated movement never supports pgs, and it is this latter contrast that this theory accounts for. DP movement cannot license pgs because the DP does not match, and thus cannot antecede, the insulated operator that would have to be present in the pg clause. Once again, insulation makes the key difference.[22]

Our approach is inconsistent with Nissenbaum (2000), a potential competing account. Nissenbaum argues that only a wh-phrase adjoined to vP can be a possible binder for the operator inside the adjunct clause, where the adjunct clause, also adjoined to vP, is c-commanded by the wh-phrase.

50a)



b)



While Nissenbaum also appeals to a version of tucking in, two assumptions that are crucial to his account do not survive in current syntactic theory.

51a) Covert movement is extrinsically ordered after overt movement.

b) The distinction between internal and external merge is primitive.

Nissenbaum assumes (51a) to insure that if an overt clausal adjunct is adjoined to vP, then a covert movement will, on his assumptions, necessarily tuck in under it with the result that no pg construction can be licensed. Since Bobaljik (2002) and in much work afterward, it is generally assumed that all movement takes place before the semantic component and so there is no additional LF-component of covert movement. The overt/covert ordering distinction in (51a) cannot be expressed in current minimalist derivations. The second assumption is that movement has a different status in the theory than any other instance of Merge, since his Non-Extension Condition, which is what requires tucking in, only applies to instances of internal Merge, not to instances of external Merge.

However, even on Nissenbaum's assumptions, if an adjunct or wh-phrase were tucked in to vP after the EA is added, it would disrupt thematic assignment, which is why wh-adjuncts to vP do not tuck in during overt derivations. If theta-assignment to the EA is not at issue, however, as in a passive construction, the Non-extension Condition would never permit the wh-phrase to adjoin high enough to bind an operator in the adjunct clause (it would have to tuck in). This predicts straightforwardly, and falsely, that passive constructions like (52) cannot support parasitic gaps.

52) *What* was John given *t* before he asked for *pg*?

Thus even if we were to add conditions to the grammar that apply to internal Merge as opposed to external Merge and add ordering distinctions between overt and covert movement, this account does not deliver the right result.

This said, the theory proposed here does not derive the overt/covert distinction without further assumptions either. If all movement is in syntax and only insulated DP matching under c-command is what conditions pg-licensing, then the overt/covert distinction

that empirically exists is not accounted for – covert movement should license pgs. The only way to capture the overt/covert distinction would require deriving that the binder for the parasitic gap cannot be the lowest occurrence of the wh-phrase. One way to do this, outlined in Safir (2010), is to assume that when internal Merge applies to a phrase, it is always the highest copy that is pronounced, but if only the quantifier head is moved, then only the lowest occurrence is pronounced (as in in situ constructions, where the head moves to establish scope). If quantifier heads are not adequate to bind pgs the way full insulated DPs are, then the distinction might be captured as the indirect result of how occurrences are pronounced. Such a theory faces serious difficulties if it can be shown that covert movement is phrasal, a fact that Nissenbaum claims to establish (see also Pesetsky, 2000) and specifically for cases where parasitic gaps are licensed by in situ wh-phrases. Unless additional independently motivated conditions predict that certain instances of covert phrasal movement is pronounced low, we are left with no account of the full range of pg distribution.

The account of parasitic gap licensing based on insulation also does not derive the inability of wh-movement from a subject position licensing c-commanding the adjunct to license pgs.

53a) [$_{CP}$ [H [*who*]] [$_{TP}$ Sarah [$_{vP}$[$_{vP}$ insulted *t*] [after *OP* [PRO meeting *pg*]]]]]?

b)*[$_{CP}$ [H *who*]] [$_{TP}$ *t* [$_{vP}$[$_{vP}$ *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 subject extraction (there is no danger of an additional Case or thematic assignment from a higher phase head), but if insulation is optional, in (53b), where it is not forced, pgs are licensed contrary to fact.

Thus the matching-insulated-operator account of pgs offered here accounts for why insulated operators are capable of supporting pgs and DP antecedents are not, but it is not a complete account of the restrictions on the distribution of pgs. Moreover, it fails our standard

of explanation unless HP operator matching is independently motivated (and I have nothing to offer here to show that it is). The bottom line, though, is that the distinction brought about by insulation permitted by PNC successfully accounts for why so-called A-positions do not license pgs. Further, parasitic gap distribution does not appear to be better accounted for by any competing theory with parsimonious syntactic assumptions.

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 (54), where parasitic gaps are supported even though the WCO effect is absent.

54a) *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* …

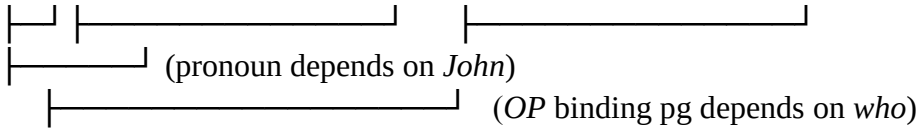
From the perspective of my proposal, the absence of WCO suggests that the operator (OP) should be a DP and thus a matching antecedent for pronouns, but support for parasitic gaps suggests that OP is not behaving as a DP. This analysis of weakest crossover might suggest that an internal contradiction 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 the relation of a pronoun to a copy of the operator that c-commands, which is blocked by the INP.

Safir (2004, with antecedents cited there), argues 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 (54d), as illustrated in (55), there is a nominal, the proper name *John*, which is not restricted by the clause over which the OP presides.

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



Who is insulated, an HP, as the analysis requires, and the pronoun *his* cannot be directly dependent on the operator, or else it is also dependent on the lowest copy of the operator, which *his mother* c-commands, which would violate INP. However, the pronoun *his* in (54d) does not have to depend on the operator or its trace (lowest copy), but rather depends directly on *John*. The same analysis can easily be extended to (54a) and (54c). For topicalization, a null operator analysis would produce the same result, and I will assume such an analysis here.

Thus weakest crossover arises due to the presence of independently available antecedents other than the insulated operator that are available in certain constructions. No appeal to differences between A- vs. A'-binding or different varieties of quantification is necessary. The analysis remains uncompromised.[23]

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 (56) is licensed because the pronoun *his* can be bound by *any man* in the position where it first entered the derivation.

56) [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 (57a), with the relevant copy relations presented in (57b).

57a)*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 (57a) contrast with (58a), in that the coconstrual of *he* and *Dean* is considered better, or fully acceptable, in (58a) (e.g., as argued for analogous cases in Freidin, 1986 and Lebeaux, 1990).[24]

58a) 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 (57a) 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 (58a) (after *that*) 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 (59) as a possible source for (57a).

59a) [$_{CP}$ [$_{DP}$ 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]]

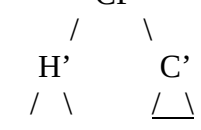
In (59b), a complement clause has been late attached (penultimate merged) to the D *which*. If this derivation is possible, then the reconstruction asymmetry should disappear, as late attachment can apply equally to complements and adjuncts – that is, (57a) would incorrectly be predicted to be acceptable. 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.

60a) [$_{CP}$ [H [which]] [did *he* expect [which] would incriminate Erlichman]]

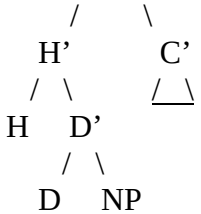
b)*[$_{CP}$ [[H [which]] [evidence that *Dean's* accomplices were guilty]] [did *he* expect [which] would incriminate Erlichman]]

Thus (60b) cannot be generated. In order to generate (57a), the complement to *which* must be introduced low, as in (57b), if it is to enter the derivation as a complement at all, as schematically illustrated in (61a,b), where late attachment applies to (61a), yielding (61b).[25]

61a)



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 (62).

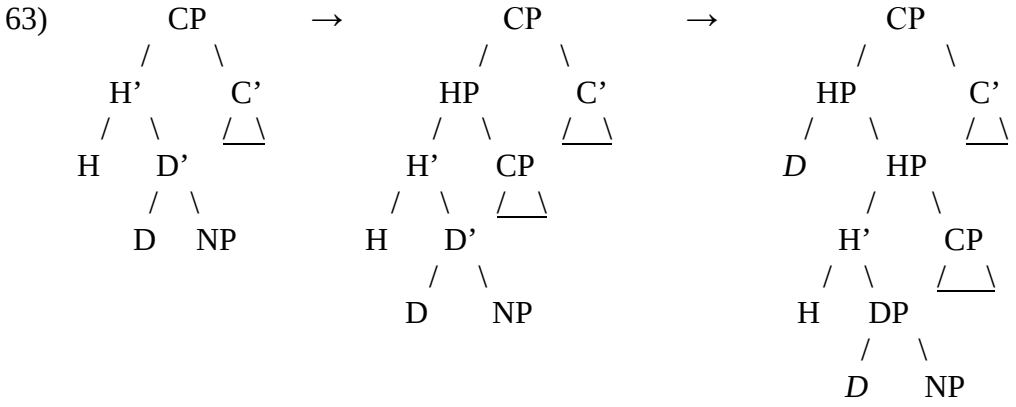
62a) [$_{CP}$ [H [which evidence]] [did *he* expect [which evidence] would incriminate Erlichman]]

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.

A reviewer points out, however, that the late merge analysis for adjunct restrictive clauses is not consistent with the scope of the determiner for such clauses; *that implicated Dean's accomplices* in (62b) is not in the scope of what delimits *which evidence*, which suggests it should have to be interpreted as an appositive. An analysis consistent with this interpretive requirement would be to assume that the determiner, or the wh-head in this case, must adjoin above both H and the late attached clause coda after late attachment of the clause takes place, as illustrated in (63).

63)



Head movement of D adjoined to HP is treated as resulting in an HP because it is not an instance of external merge – D has already been assigned a complement, namely NP, but D now has scope over the complement clause coda.[26]

Sauerland (1998) argues for a further distinction between A- vs. A'-structures, as in (64a,b). 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.

64a) A-movement *optionally* leaves a trace.

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

65a) 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 ((65a) 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 consist only of a D head (e.g., [*the*] in (65b)). The D head moves to the position of the top of the A-chain, and then [*evidence that John is guilty*] is introduced by late attachment (penultimate Merge).

66a) [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 (60) 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. 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.

Recall now that the absence of WCO effects for A-movement depended on the assumption that copies left by A-movement would somehow be invisible to the INP, while wh-complement copies left by A'-movement should be visible so that WCO can be derived by INP. In a raising structure, the complement (*woman*) of the raised D (*every*) can be late-attached and the trace of QR (applying to the newly formed DP in syntax but pronounced low) will be a full DP that can antecede a pronoun.

67) *Every woman* seems to *her* mother to be smart.

This is possible because there is no HP layer insulating the D head, so at the moment that the D is immediately dominated by TP, it is eligible to be a site for penultimate Merge. If *every* were insulated by H before *woman* is moved to CP, it will not only fail EPP, but it will block late attachment, which will be in a site lower than the penultimate node (it is below the TP node that immediately dominates HP). Thus so-called A-movement does not have to reconstruct a trace in a position where INP would be triggered. This completes our account of why WCO only holds of so-called A'-constructions, since only A'-constructions are not susceptible to the late attachment of complements of D. Such a derivation would violate PNC.

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 and rules out a derivation that penultimate merge would otherwise allow.

68a) [$_{vP}$ he believe D] (merge bare D as object)

b) [$_{vP}$ D [$_{vP}$ he believe D] (move D to Spec vP)

c) [$_{vP}$ [$_{DP}$ D NP] [vP he believe D] (late merge the NP restrictor containing the clause)

d) [$_{vP}$ [$_{HP}$ H [$_{DP}$ D NP]] [$_{vP}$ he believe D] (insulate)

On this derivation, the NP would not get Case where it is penultimately merged because it will then be insulated, assuming, as we have all along, that case is assigned at the end of the phase.[27]

To summarize this section, the PNC 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), which also explains why so-called A-chains do not induce WCO. 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 independently motivated factors, such as phase theory, EPP, Case assignment, and criterial satisfaction.

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.

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 (69) and neutralizes WCO, as in (70) (both (69) and (70) are from Saito, 2003; 485).

69) ?[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.'

70a)?*[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 (71) and WCO is still in force (72).

71) *[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*'

72) ?*[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 force insulation where the construction (descriptively) shows A'-properties, prevent it where it only shows A-properties, or allow both options 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 of an object across a subject to a topic or focus position would not require insulation and 'A'-effects' would be neutralized.

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.[28] 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.

Miyagawa's αP in a position just above TP may be all that is needed to adapt his approach to mine if we simply assume that α is the source of Nominative Case, not T. Consider the derivation in (73) after the DO has adjoined to vP, the option to insulate is not taken and T has been merged, as in (73a). No Case is assigned by T so the vP-adjoined DO is not blocking anything. The next step in the derivation, movement of EA to Spec TP (73b) puts the EA in a position where it will be assigned Nominative when α is merged.

73a) [T [$_{vP}$ [$_{DP}$ wh-DO] [$_{vP}$ EA [v [V [$_{DP}$ wh-DO]]]]]]

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

c) [α [$_{TP}$ EA [T [$_{vP}$ [$_{DP}$ wh-DO] [$_{vP}$ EA [v [V [$_{DP}$ wh-DO]]]]]]]]

d) [$_{αP}$ [$_{DP}$ wh-DO] [α [$_{TP}$ EA [T [$_{vP}$ [$_{DP}$ wh-DO] [$_{vP}$ EA [v [V [$_{DP}$ wh-DO]]]]]]]]]

Now Spec αP could be formed by moving the Accusative-marked wh-DO to Spec αP, as in (73d). Because the source of Nominative is higher, the configuration that gives rise to A'-opacity does not arise – indeed this derivation is possible whether the DO is a wh-phrase or not. Since the DO arrives in Spec αP as a DP, it is an appropriate 'A-binder' for an anaphor that is the possessor of the EA. It is not necessary to assume that the either Spec TP or Spec αP is an EPP position. If the next step in the derivation after (73a) moves wh-DO into Spec TP, then double Case assignment would crash the derivation when α is merged. If αP is not an EPP position, then it can be left empty, in which case the 'unscrambled' order emerges with the Nominative EA c-commanding the Accusative DO.

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 double Case assignment by a higher clause, as I did with respect to adjectival clausal complements in (27). An uninsulated DP in Spec CP, a position higher than αP,

would be susceptible to Case assignment by the higher v. The assumptions that insure double Case assignment may be delicate based on how Accusative is assigned in vP (e.g., to all DPs below v in the v phase instead of to just the highest?), or for cases where the higher v cannot assign a Case (i.e., how Nominative is assigned for an intransitive) but even if an additional stipulation is required to insure insulation after scrambling leaves the clause of origin, the streamlined assumptions of the free Merge account proposed here still seem appealing by comparison with what is stipulated in other accounts of Japanese scrambling.

I am not claiming that my adaptation of 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 PNC-based insulation theory to succeed would be appealing.

7.0 Languages where movement can be 'improper'

The IMB is derived by the interaction of two factors, namely, that EPP cannot be satisfied by a null-headed constituent in Spec TP and conspiracies that insure that insulation must have occurred before movement to Spec TP. Insulation protects a DP from Double Case assignment and/or it enables the EA to get Case that it might otherwise fail to be assigned, in favor of an uninsulated DP externally merged to vP. However, if A'-opacity is not required

for these reasons, then a wh-phrase need never be insulated and it should be an adequate specifier to satisfy the EPP as Landau defines it. 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 (74) (all Kilega data from Carstens, 2005).

74) 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.'

Though languages may differ as to what the thematic conditions related to EPP are, it is important to understand that this is orthogonal to the question of whether or not insulated DPs are permitted. The EPP still conditions null heads, but does not specify syntactic category type, at least not as a part of the explanation offered here. However, this leaves an open question about how the subject, 'elephants', is licensed by Case assignment.

As in many other Bantu languages, questions in Kilega can be formed in situ, as in (75a) or by fronting as in (75b), 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.

75a) Bábo bíkulu b-á-kás-íl-é mwámí bɪ´kɪ´ 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) Bikɪ´ bi-a´ -te´nd-ɪ´le´ ba´na bi-a´ -gu´ l-ɪ´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.[29] 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 (75c).

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 satisfy EPP because the DP is uninsulated, but 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. 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. This is much less attractive with respect to EPP because it would appear that a stipulation about the syntactic category of what satisfies EPP would be required and it would not address the Case requirement on the EA that does not move to Spec TP. For this reason the latter approach is rejected here. 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, and none of the adjustments to bring about these derivations refer directly to any A- vs. A'-distinction.[30]

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.[31]

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. 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 (76b) (van Urk, 105), a topicalized PP becomes nominal.

76a) Bòl à-thɛt 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.'

Van Urk shows that preposition incorporation is possible for PP complements and argues that only those PPs that can undergo the incorporation allow for the P-complement DP to move to preverbal position. He assumes that 'only nominals may move to spec FP and be assigned Case there'. Since a structural Case cannot be stacked over an oblique Case (following Richards, 2012), if the stripping does not happen, Case assignment in Spec CP will fail and it the stipulation necessary is that Case assignment in Spec CP must not fail. This stipulation would work in my account as well, but double Case assignment would still have to be avoided. This is achieved in van Urk's theory by essentially reassigning any structural Case as Absolutive if it is in Spec CP. If such reassigning is permitted, then Double Case Assignment will not force insulation. So-called Genitive DPs, DPs that would be Genitive in non-subject contexts (e.g., as prepositional objects) have a different shape in Spec TP. Urk (2015: 86) appeals to late head insertion as a Last Resort to license subjects with Genitive Case that are moved to Spec TP to satisfy EPP. Van Urk's null preposition insertion would fail EPP, given the approach to EPP taken here. I would argue instead, following an

alternative that van Urk himself suggests (p.92), that the 'stripped' Genitive is simply Nominative assigned by T which moves to Spec TP to satisfy EPP. Reassignment of Case would apply if the subject moves to Spec CP (for subject extractions), so again, no insulation is necessary. On this account, no appeal to Last Resort is necessary either. Thus there is no A'-opacity that insulation is required to compensate for, which in turn predicts that uninsulated movement can pass intervening subjects, DPs in Spec CP can bind anaphors and neutralize WCO, etc. No appeal to edge features is necessary, nor does any assumption about EPP identify an A- or A'-position.

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.[32]

8.0 Competing explanation 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,[33] 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.

77) 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.

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. 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)[34] 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 obligatory features 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 semantic 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 discussed here. Van Urk's also adopts the Takahashi and Hulsey approach in its essentials, but assumes late attachment without ever explaining why it should be allowed to apply and violate Extension. This difference may be the most crucial distinction that divides feature-based accounts from insulation-based ones.

The insulation account only introduces one new and independently motivated 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 falls short by introducing an additional assumption which is precisely the one that the insulation approach uses to derive everything else.

9.0 Summary and conclusion

The key theoretical assumption that underlies my explanation of the A- vs. A'-distinction is the reformulation of Chomsky's Extension as the PNC. The PNC permits more sites for Merge than Extension by allowing penultimate merge, but it still limits derivational moves to the crest of the derivation and it is independently motivated by a variety of phenomena, including late attachment of adjuncts, superiority, antecedent-contained deletion and head movement, which do not bear on the purported A- vs. A'-distinction. Since the PNC defines how Merge can apply, penultimate Merge is possible anywhere unless other factors rule it out.

The distinctions we set out to explain are listed in Table 1, repeated here, and in addition, there is the IMB.

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. Must reconstruct
A'	-/%-	-	+	+	-	+	+	Mostly yes
A	+	+	-	-	+	-	-	no

In every case, the explanation stems from the PNC, which means that penultimate Merge is a possible instance of Merge.

The insulating head motivated by A'-opacity effects is typically null and Landau's version of the EPP forbids EPP to be satisfied by constituents with null heads, so the IMB follows: No DP that is insulated after adjunction to vP can satisfy EPP. Thus an insulated DP cannot move from DP-adjoined position to Spec TP.

A wh-DP adjoined to vP is a candidate for Case assignment unless it is insulated, which prevents the insulated DP from being doubly Case assigned. The head insulating a wh-DPx can only enter the derivation through penultimate Merge if there is a wh-DPx copy that Case assignment can apply to, or else no copy of wh-DPx will satisfy the Case Filter (or whatever does that work), as in extraction from DO position. From this conspiracy, it follows that Case is not assigned to 'A'-movement' landing sites because A'-moved constituents are not DPs. The uninsulated EA gets Nominative as if the insulated wh-DP were not there and T can agree with the EA. The properties 1-3 in Table A are now derived. Since PPs cannot bind anaphors, insulated DPs cannot bind anaphors, which derives 5 in Table 1.

While the existence of pied-piping is a consequence of Free Merge, the trick is to restrict its distribution. What is in the pied-piped constituent may determine whether or not it satisfies criterial conditions in its final landing site. Thus only constituents that embed a wh-phrase will satisfy interrogative C, whatever the mechanism that permits the criterial head to detect the presence of the wh (movement to the edge of the pied-piped phrase as proposed

here for English, or a semantically contentful head that must c-command wh-elements as in Tlingit). The source of the restrictions on the content of PPs that satisfy EPP are less than clear, but some kinds of PPs and 'honorary NPs' can satisfy EPP, though insulated DPs are PPs that do not satisfy it because the insulating head is null. As Landau shows, the EPP as it applies to rule out subjects with null heads is independently motivated and is thus no threat to our reduction of the A-/A'-distinction.

WCO is induced for a pronoun when dependency on an operator means depending also on all of its copies such that one of those copies is c-commanded by (the constituent containing) the pronoun, thus violating INP. This sort of dependency would also hold for copies of any movement and is thus not enough to predict the absence of WCO triggered by A-movement. The reason so-called A-movement does not result in WCO is independently derived by the way that PNC restricts the distribution of reconstruction effects, that is, the explanation of 8 in Table 1 derives the distinction in 7.

The distribution of parasitic gaps, distinction 6 in Table 1, is partially captured by the like-binds-like assumption, and it appears that no competing account of parasitic gaps, including Nissenbaum's, has fewer problems. However, the like-binds-like assumption as it applies to insulated DPs is so far not independently motivated beyond the parasitic gap portion of the A-/A'-distinction that it derives, and if not, it falls short of the explanatory standard we aspire to. The issue remains open.

Finally, the distribution of reconstruction effects, 8 in Table 1, is derived from the PNC as it interacts with insulation. Adjunction to an insulated DP cannot be a complement to anything, since it is not the sister to a head. Thus only adjuncts can be late-attached to insulated DPs. If D is generated without a complement and moves to a higher position without being insulated, penultimate Merge can insert the complement of D. The late attachment account, built on a proposal by Sauerland (2004), simply follows from what the

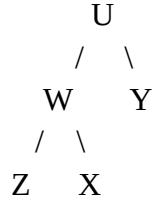
PNC does and does not allow, and what is at stake is whether or not there is a copy in a position where it could induce WCO or Principle C effects. Since the complement of D does not have to originate where D is first merged, so-called A-movements do not have to leave a copy. So-called A'-movements cannot avoid leaving copies that are complements because the PNC does not allow for late attachment to add a complement to an insulated D. This derives 8 of Table 1 and in so doing derives 7.

Although issues remain, including questions surrounding languages where A'-opacity plays a smaller role, we have eliminated the A-/A'-distinction from linguistic theory, while predicting the properties of the two descriptively different classes of constructions. Insofar as the PNC interacts with independently motivated constraints to predict the nature and distribution of A-/A'-distinctions, the theory proposed here meets the criteria for an explanation set out in (1a,b).

Appendix: The PNC and keeping track of copies

The calculation that is necessary to determine whether or not the PNC has been obeyed is requires determining for any instance of Merge, merge^{i} that the undominated node X immediately dominate a node that X did not immediately dominate before merge^{i}. When merge^{i} is penultimate, how do we know that the node immediately dominating what has just been merged is in fact a new node, e.g., when the node W is formed by penultimate Merge of Z with X in (1)? If the relation is adjunction (i.e, Z is adjoined to X), whether one regards adjunction as pair merge or set merge, the syntactic category label of that node would be the same as the node it adjoins to (the label of X), as in the case of late attachment.

1)



If the PNC is to recognize the DP node it dominates as 'new', as in the case of W in (1), then the category label of that node is insufficient for the task. In this brief appendix, it will be demonstrated that an independently necessary revision of assumptions about how copies are identified also happens to facilitate how the PNC calculates that an instance of Merge is possible.

The revision in question concerns the operation SELECT and numeration indices, originally introduced by Chomsky (1995). Apart from facilitating economy calculations and enabling deterministic derivations, which we have set aside as theoretical motivations, given our assumption of Free Merge, the abiding motivation for numerations is that they permit copies resulting from instances of internal Merge to be identified. Numeration indices distinguish between multiple selections of the same lexical item (LI) from the lexicon from multiple occurrences of the same selected LI. Occurrences are instances where the same LI is used more than once in a derivation, but where that LI is chosen from the lexicon only once. Numeration indices, which were introduced to make this distinction possible, are assigned to each instance of a selection of a LI for use in a derivation (by the operation SELECT), and if the same LI is selected more than once, then it will have different indices for each time it is selected from the lexicon. To insure that the indices for each choice of a LI are distinct, all the LIs to be used in a derivation are chosen by SELECT at once, e.g., each choice is given a distinct number from a sequence. Thus if the word *the* is used twice in a derivation, we can distinguish non-copies, e.g., the use of the^1 from the use of the^2 , as opposed to the repeated use of the^1, which would mean that the two incidences of the^1 are copies. Non-terminal constituents that contain all the same numeration indices on the lexical items inserted into their terminal nodes are copies of one another.

There are several problems with this algorithm. First, all the LIs to be used in a sentence of whatever length must be chosen at once. This seems to be a very clumsy device

insofar as the entire derivation must be prefigured in an intuitively non-local way, that is, in a space where there is no locality restriction (because structures have not been built yet).

A second much more serious problem is that the computation that determines that two constituents are occurrences of the same constituent (copies) is potentially computationally burdensome and non-local. Consider, for example, what would be required to know if the first of the two bracketed constituents is, in fact, a copy of the second bracketed constituent.

2) [Which picture of a president that John considers to be unfit for any national office], did Mary see [which picture of a president that John considers to be unfit for any national office].

In order to know that these two constituents are the same, an algorithm must list all of the numeration indices in each constituent and then make sure that they match by pairing each index in one constituent with the position of its counterpart in the other. At the Sensory-Motor (SM) interface, this might not be too burdensome if the calculation is simply an evaluation of strings without regard to the structure that dominates them (even though the strings are of potentially infinite length). However, in the Conceptual-Intentional (CI) interface, semantic calculations that determine which copy is an operator and which copy serves as a variable must know which constituents are copies, even if this is determined at the end of every phase from the point of the last extraction (internal Merge). Otherwise, the interpretation of (1), which must essentially be of the form in (2), cannot be calculated.

3) For which x, x a picture of a president that John considers to be unfit for any national office, Mary saw x.

To be certain that the lower constituent is indeed a copy, the semantic component, which does not compute strings, would have to search from the top of the constituent to the bottom of it to determine that it is indeed a copy, assuring that every numeration index in the higher constituent matches every one in the lower one. In so doing, the search would have to reach

below phase boundaries and islands to locate all of the numeration indices in both constituents. This requirement seems to miss an obvious generalization, namely, that when the whole constituent moves, all of the stuff under it is the same. Notice that this problem is not solved by the No Tampering Condition, which assures that movement does not change the contents of the constituent moved. At the CI interface, there is nothing that connects the two constituents but the calculation that determines they are the same. This is a problem that indices left on traces of movement did not have because those indices related constituents rather than all the LIs that they dominate.

Third, in theories that include subarrays that limit what can be used in any one phase, use of LIs with the same numeration indices in different subarrays needs to be avoided, not least because elements with the same numeration index could end up in a precedence relation where they precede themselves, and following the reasoning of Kayne (1994) or Nunes (2004), no LI can precede itself in spell-out. Spell-out, however, can only see syntactic objects in within a phase. If lexical arrays, subsets of the numeration, are not policed to avoid lexical choices with the same numeration index introduced in a higher phase, then lexical choices with the same numeration indices could co-occur in different phases and Spell-out would be unaware of the coincidence. This is why the sets of LIs in a subarray can only be partitioned from the larger set already SELECTed because all of those LIs will already have numeration indices. Thus SELECTion must be followed by linguistic array selection followed by Merge. This seems like a lot of machinery.

However, if we are to recognize copies when they have to be detected, what is the alternative to numeration indices that would avoid these pitfalls? Notice that if copies are only detectable within a phase, then it could be assumed that coincidences of form for two constituents that have all the same lexical items simply cannot be copies if they are not in the same phase. If that is our default assumption, then we only need to recognize copies when

they are in the same phase. Put another way, only the terms introduced by Merge within a phase need to be evaluated as to whether or not they are copies that could lead to linearization paradoxes and permit appropriate operator-variable interpretations. Suppose we limit the indices that permit us to identify copies in the same way we limit everything else, that is, only terms within the same phase are indexed distinctly unless a term that already has an index in the phase is used. Thus copies would only be recognized within a phase. Now let us assume that SELECT is not in the grammar, hence no array selection from what is SELECTed, and indeed that there are no numeration indices. This is now too radical, because now we have no way to identify copies, so we need an alternative.

Here is a way to do it: Suppose we build a derivation merging X and Y, in which case X and Y get distinct *term indices*, X^1 and Y^2 in $[X^1\ Y^2]$. Notice that the new constituent formed by this instance of Merge has no term index because it is not a combined term in a Merge operation. Then $[X^1\ Y^2]$ is merged with Z, in which case Z and $[X^1\ Y^2]$ get term indices as Z^3 and $[X^1\ Y^2]^4$. Now suppose that $[X^1\ Y^2]^4$ is internally merged to $[Z^3\ [X^1\ Y^2]^4]$, in which case only one new term index is introduced, namely, that for the second term in (4), which is $[Z^3\ [X^1\ Y^2]^4]^5$.

4) $[[X^1\ Y^2]^4\ [Z^3\ [X^1\ Y^2]^4]^5]$.

We now introduce the following definition:

5) Two constituents are occurrences (copies) if and only if they bear the same term index.

If Z^3 is a phase head, then both copies of $[X^1\ Y^2]^4$ are visible to spell-out and precedence relations are resolved accordingly, e.g., the lower occurrence of $[X^1\ Y^2]^4$ is not pronounced, or, in the semantics, it can be interpreted as a variable because the constituents with the superscript '4' within the phase are copies.

The next step is to move beyond the phase. In the next phase, under standard assumptions about phases, only the constituents in the Z^3 phase edge are visible to the next phase head. However, since we have already calculated what the copies are in the Z^3 phase, we do not need that information in the next phase. So let us assume that all the term indices assigned within the preceding phase (Z^3) are erased, including those constituents on the phase edge. Since they have no term indices, the edge-dwellers cannot be copies of anything within the next phase unless they are terms in a Merge operation that takes place in the new phase.

For example, suppose that we have merged another head, B, above the Z phase, in which case term indices are introduced for B and the whole Z phase, as in (6a). Then assume that some other non-terminal constituent is merged above that, as in (6b), above that, e.g., what is descriptively a Spec B, of the form [C D] (leaving the indices for C and D blank, on the assumption that [C D] is a phase, e.g., a DP phase). Next let's introduce a new phase head A by external merge as in (6c) and then apply internal merge to [X Y], as in (6d).

6a) $[B^1\ [[X\ Y]\ [Z\ \ldots]]]^2$

b) $[C\ D]^3\ [B^1\ [[X\ Y]\ [Z\ \ldots]]]^2]^4$

c) $A^5\ [[C\ D]^3\ [B^1\ [[X\ Y]\ [Z\ \ldots]]]^2]^4]^6$

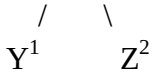
d) $[X\ Y]^7\ [A^5\ [[C\ D]^3\ [B^1\ [[X\ Y]^7\ [Z\ \ldots]]]^2]^4]^6]^8$

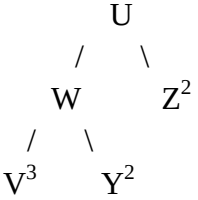
Now [X Y] receives a term index because it is a term in an instance of Merge and as a result that index is on both instances of [X Y], identifying the two [X Y] constituents as copies within the A phase. When the A phase ends, all the term indices are erased and we begin again. At every point, the identification of copies is limited to the phase and does not violate minimal search by investigating the internal contents of the copies.

The term index system eliminates SELECT, selection of lexical arrays and numeration indices at the cost of introducing term indices. All the problems mentioned are solved, but term indices violate Chomsky's Inclusiveness Condition, which forbids any

addition of extra information to the derivation after Merge applies. Numeration indices avoid violating Inclusiveness by a sort of trick – they are technical derivation-tracking information introduced 'before' the derivation is built, but they violate the spirit of Inclusiveness, which is to avoid adding diacritics to the derivation that are not elements already introduced in the lexicon. Of course numeration indices are not lexical either, but they are defined, theory internally, as preceding the derivation, when in fact, like the SELECT operation, they are intuitively part of it. For example, there is no reason for SELECT to apply unless a derivation is being built that uses the output of SELECT. Compare Chomsky's (2013, 2014) Labeling Algorithm which, though part of the derivation, plausibly does not violate inclusion; Non-terminal nodes are identified as of one type or another based on the algorithm, but the features (presumably) and labels originate in the lexicon. One might hope that the effects achieved by term indices might be collapsed into the Labeling Algorithm, since both rely on minimal search, but that will not be attempted here.

Returning now to the reason for this appendix, the introduction of the term index system makes it possible to compute compliance with the PNC by extremely local search, that is, the undominated node U after merge^{i} will immediately dominate a node that it did not immediately dominate before merge^{i}. This is straightforward for a case of external merge that creates U as a new node, thus anything it immediately dominates it does so for the first time.

7a)


b)


In (7b), we have an instance of external penultimate Merge applying to (7a). The node W created by Merge in (7a) is not a term in the sense intended here, namely, it is not the third member of a set formed by Merge, but the name of the set (assuming some labeling algorithm), so W has no term index. This is as it should be, since W, which dominates [V Y] is not a copy of Y. Internal penultimate merge would have the same consequence for W. Thus W is a node distinct from Y^2 because it does not have the same term index – it has no term index at all.

To conclude, there is always a clear way to determine if an instance of Merge has satisfied PNC and it only requires comparing the immediate dominance relations of the undominated node before and after an instance of Merge applies. The system that permits this determination does not require any addition to the theory that is not independently motivated.

References

Alexiadou, Artemis and Elena Anagnostopoulou. 1998. Parametrizing AGR: Word order, V movement and EPP-checking. *Natural Language and Linguistic Theory* 16.3:491-539.

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.

Bresnan, Joan. 1977. Variables in the theory of transformations. In Peter Culicover, Thomas Wasow and Adrian Akmajian, eds., *Formal syntax*, 157-196. New York: Academic Press.

Cable, Seth. 2010. Against the existence of pied-piping: Evidence from Tlingit. *Linguistic Inquiry* 41: 563-594.

Carstens, Vicki. 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.

Demirdache, Hamida. 1991. Resumptive chains in restrictive relatives, appositives and dislocation structures. Ph.D. dissertation. Cambridge, MIT.

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.

Grohmann, Kleanthes. 2011. Anti-locality: Too-close relations in grammar. In, Cedric Boeckx, ed., *The Oxford handbook of linguistic minimalism*, 260-290. Oxford: Oxford University Press.

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.

Jaeggli, Osvaldo and Ken Safir. 1989. The null subject parameter and parametric theory. In Osvaldo Jaeggli and Kenneth J. Safir, eds., *The null subject parameter*, 1-44. Dordrecht: Kluwer.

Kayne, Richard. 1984. *Connectedness and binary branching*. Dordrecht: Foris Publications.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge: MIT Press.

Koopman, Hilda and Dominique Sportiche. 1983. Variables and the Bijection Principle. *The Linguistic Review* 2: 139-160.

Landau, Idan. 2007. EPP Extensions. *Linguistic Inquiry* 38: 485-523.

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.

Legate, Julie Anne. *Voice and v: Lessons from Acehnese*. Cambridge: MIT Press.

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. 1990.Resumptive pronouns, A'-binding and levels of representation in Irish, in Randall Hendrich, ed., *The syntax and semantics of Celtic languages*, 199-248. San Diego: Academic 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.

McFadden, Thomas. 2004. The position of morpohligical Case in the derivation: A study on the syntax-morphology interface. Ph.D. dissertation, University of Pennsylvania.

Miyagawa, Shigeto. 2010. *Why agree? Why move?* Cambridge: MIT Press.

Moulton, Keir. 2009. Clausal complementation and the *wager*-class. In Anisa Schardl, Martin Walkow and Muhammad Abdurrahman, *Proceedings of the 38th annual North East Linguistics Society*, vol.2, 165-178, Amherst: GLSA.

Nissenbaum, Jon. 2000. Covert movement and parasitic gaps. In M. Hirotani, A. Coetzee, N. Hall and J.-Y Kim, eds. NELS 30. Amherst: U. of Massachusetts GSLA.

Nunes, Jairo. 2004. *Linearization of chains and sideward movement.* 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.

Pesetsky, David. 2000. *Phrasal movement and its kin*. Cambridge: MIT Press.

Postal, Paul. 1972. On some rules that are not successive cyclic. *Linguistic Inquiry* 3: 211-222.

Postal, Paul. 1993. Remarks on weak crossover effects. *Linguistic Inquiry* 24: 539-556.

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, Norvin. 2012. Lardil "case stacking" and the timing of case assignment. *Syntax* 16:42-76.

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. 1986. Relative clauses in a theory of binding and levels. *Linguistic Inquiry* 17: 663-389.

Safir, Ken. 1991. Evaluative predicates and the representation of implicit arguments. In Robert Freidin, ed., *Principles and parameters in Comparative grammar*, 99-131. Cambridge: MIT Press.

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. 2014. One true anaphor. *Linguistic Inquiry*: 91-124.

Safir, Ken and Mamadou Bassene. 2016. The Morphosyntax of Verb Stems in Jóola Eegimaa: Stem-internal movement and compositional interpretation. *Natural Language and Linguistic Theory* 1-59.

Saito, Mamoru. 2003. A derivational approach to the interpretation of scrambling chains. *Lingua* 113:481-518.

Sauerland, Uli. 1998. The meaning of chains. Ph.D. dissertation. MIT, Cambridge, MA.

Sauerland, Uli. 2000. Two structures for English restrictive relative clauses. In Saito et. al., eds., *Proceedings of the Nanzan GLOW*, 351-366. Nanzan University: Nagoya, Japan.

Sauerland, Uli. 2004. The interpretation of traces. *Natural Language Semantics* 12: 63-124.

Schneider-Zioga. Patricia. 2007. Anti-agreement, anti-locality, and minimality: The syntax of dislocated subjects. *Natural Language and Linguistic Theory* 25:403-446.

Selvanathan, Nagarajan. 2017. An account of *tough* predicates and *Tough movement*. Rutgers University, ms.

Sichel, Ivy. 2014. Resumptive pronouns and competition. *Linguistic Inquiry* 45: 655-693.

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.

Wurmbrand, Susi. 2006. Licensing Case. *Journal of Germanic Linguistics* 18:175-236.

Ken Safir
Deparment of Linguistics – Rutgers University
18 Seminary Place
New Brunswick, NJ 08901
safir@linguistics.rutgers.edu

---

* I would like to thank the many audiences that responded to versions of this work beginning with GLOW in Asia XIII at Beijing Language and Culture University (2010) and audiences at U. of London, Queen Mary, U. of Leipzig, and ZAS Berlin, as well as eight classes of graduate students, particularly Carlo Linares, Matt Barros, Livia Souza, Nick Winter, Naga Selvanathan, Lydia Newkirk, Ang Li, Hazel Mitchley, Ümit Atlamaz and Kunio Kinjo. I

would also like to thank Mark Baker and Vicki Carstens in particular for insightful discussions. Two LI reviewers made substantial contributions with constructive criticism that took me years to do justice to, if indeed justice has been done. This work was also supported by NSF BCS 0919086 and 1324404.

[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.

[2] Some languages appear to have 'case stacking', where more than one morphological casemarker occurs on the same nominal. It may be that (13) only holds of DPs where at least one of the Cases is structural (e.g., Nominative or Accusative but not Dative) but assumptions about morphological Case may need to be more complex. See Richards (2012) for a recent discussion based on data from Lardil. If Case can be assigned to an insulated DP in some languages, it might avoid violating (13), but might perhaps require morphological reduction in PF in most cases. Some other assumptions might have to be adjusted, but I leave the matter open.

[3] While doubt has been cast on whether the Case Filter should be in the theory (e.g., McFadden, 2004), if whatever principle is appealed to to derive its effects requires that a DP must be in a particular kind of relation to be licensed, then the consequence for the theory proposed here is probably harmless, but only if the one-one requirement on Case assignment is matched by this hypothetical relation. I am not assuming, however, that the distribution of PRO is determined by lack of Case. Instead I assume that PRO must be licensed as well – leaving unexplained why it is silent – and that the licensor of PRO has the same potential as abstract Case to trigger (13).

[4] 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, (19b,c) are not really distinct in syntax, but I are provided for convenience of presentation only.

[5] 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. This introduces questions about how it is known that a node is new, a matter addressed in the Appendix to this paper.

[6] I abstract away from the literature that contests whether Spec vP or a higher Spec VoiceP that is the criterial position for the EA. If Spec VoiceP is the criterial position, then Voice rather than v must be the phase head. See Legate (2014) for recent discussion. Clearly this approach is inconsistent with 'criterial freezing', which is not assumed here, a matter not addressed for reasons of space.

[7] Thus wh-movement could not proceed by tucking in under the external argument, for example, because it would then have to be interpreted as an additional external argument. Another possibility pointed out by a reviewer would be one where (12b) is followed by adjunction of the EA higher in vP above the wh-DO. This is plausibly ruled out by antilocality (as the reviewer suggests), e.g., in any phase, copies must always c-command or be adjoined to an intervening head to be successfully interpreted. For discussion of antilocality effects, see, e.g., Grohmann, (2011).

[8] Extraction when *that* is present is a typical *that-e* violation. See Kayne (1984:3), 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.

[9] If there are languages where the Case Filter does not apply, that is, languages where Case licensing is not necessary for some reason, then a rather different pattern is predicted, as discussed in section 7.

[10] Landau (2007:505-506) points out that examples like (30a) could arise by another derivation involving distributed deletion. If only one of a set of copies can survive at PF, as is generally assumed, what is prevent a derivation of (30a) as in i., where the PP remains intact as it moves, but the copy of P in matrix CP is not deleted and instead the copy in subordinate Spec CP is retained.

i. $[_{CP}\ [_{PP}$ ~~*to*~~ *who*$]\ [_{TP}$ John T $[_{vP}\ [_{PP}$ ~~*to who*~~$]\ [_{vP}$ think

$[_{CP}\ [_{PP}$ *to* ~~*who*~~$]\ [$ (that) $[_{TP}$ Mary should talk $[_{PP}$ ~~to who~~$]]\ ]]\ ]]]$

Landau argues that such cases are blocked by a version of EPP that applies uniquely at PF. The relation is one of selection between the head H that bears EPP requirement and the head of its Spec, X, such that EPP requires that X be non-null. If the matrix C bearing the interrogative feature also bears the EPP feature, then the head of the PP cannot be null. However, earlier in his paper he assumes traces can satisfy EPP. If he did not say so, then *Who did you say left?* would not be possible because the subject of the lower clause is a copy after movement to Spec CP within the CP phase and would cause an EPP violation if the subject copy did not count for EPP at the end of the phase. Thus the EPP must be met at the point where the Spec is merged, before copy deletion, but if that is so, then his account of (30a) is lost. Although I adopt Landau's proposal about EPP with respect to T, I do not accept it for Spec CP, at least for languages where Spec CP cannot support an expletive. The stranding ban may be derived by the proposal for the pronunciation of highest copies in section 4.3 drawn from Safir (2010), though problems remain.

[11] As mentioned in section 2, I am not attempting to explain the differences in English between what wh-movement can strand and what passive can strand (see, for example, Safir

(1991), where this distinction is used as a diagnostic for the A-/A'-distinction in the analysis of the *worth* construction).

i. The boys have not yet been spoken to.

ii. *The boys have not yet been given a book to.

iii. Which boys have the teachers not yet given a book to?

Although it is a more English particular contrast, it still deserves explanation and I do not have an account of it.

[12] Reviewers point out that Landau's theory does not explain how bare plurals and generics satisfy EPP, especially if they are associated with a null D. Intuitively, the distinction in English appears to concern whether or not there is phonological material in the extended projection of NP, on the common assumption the P is not in the extended projection of NP, at least not in English. This would make the appropriate distinction for Landau's EPP for English, but it remains to be seen how general such a solution would be (e.g., something different would have to be said about bare nouns and NPIs in Romance, that cannot satisfy EPP in this way on Landau's account) or how it would be best executed, a matter not examined here.

[13] Other candidates for PP subjects, as in locative constructions in English, are much less fully felicitous with subject-aux inversion, although if the PP is not in Spec TP, then it is mysterious as to what satisfies the EPP in subject position, unless it is an unpronounced copy of the postverbal subject. I have no proposal to make here.

i. Under the banyan tree was sitting a newly famous guru.

ii. ?*(When) was under the banyan tree sitting a newly famous guru?

iii. When under the banyan tree was there sitting a newly famous guru?

[14] Potential instances of 'pied-piping' for A-movement to satisfy EPP include examples pointed out by Bresnan (1977).

i. In these villages are found the finest examples of native cuisine.

ii. ? (How often) Are in these villages found the finest examples of native cuisine?

Some may judge ii. to have similar status to ii. in fn.11, but if not, attraction of a PP to Spec TP instead of the direct object of a passive sentence could be seen as a form of pied-piping, though not a movement that is required because something in the constituent moved has a property required for criterial satisfaction, as in the movement of a larger phrase that must contain a wh-phrase to meet criterial satisfaction for interrogative or relative clause C.

[15] I am assuming that EPP is satisfied by pro or phi-features on T in pro-drop languages (see, e.g., Alexiadou and Anagnostopoulu, 1998) and that PRO somehow satisfies EPP, a common assumption that is particularly unnatural in Landau's system. See Landau (2007:497 fn.10)

[16] Landau (2007:503-505) argues that determiner doubling in split topicalization structures is required by EPP, and although his analysis has more complications than are worth reviewing here, the point is that the topic position will have a semantically unnecessary determiner required in the initial position for V2 when the split topic has stranded a determiner in the position extracted from. This is the sort of independent motivation for an EPP effect that my proposal requires. Simpler evidence for German is that the initial position can be filled by expletive *es*. By contrast, Spec CP position for an interrogative complementizer does not appear to be regulated by EPP, as it can be empty in yes-no questions (no expletive inserted) and insulated DPs must land there. For whatever reason, both interrogative C and 'topic C' attract V-to-C

There is some question as to whether the Spec TP position in German is an EPP position. Wurmbrand (2006) argues that EPP does not hold of Spec TP in German, but the matter is complicated by the contrast in i. vs. ii.

i. Es ist möglich, dass getanzt wurde.

It is possible that danced was

ii. *Es ist möglich getanzt zu werden.

It is possible to be danced.

Safir (1985) and Jaeggli and Safir (1989) point out that unlike pro, PRO cannot be expletive, but ii. is not ruled out by this generalization if EPP does not require a PRO subject for the infinitive in ii. It is possible that EPP holds for only tenseless T in German, though this seems like an odd result.

[17] For null operator constructions, such as complex adjectivals, it is hard to see how EPP is satisfied. Unless all relative clause structures involve raising of the relative clause nucleus out of Spec CP position, such that the trace of raising out of CP satisfies EPP, then relative clauses must have null operators. See Sauerland (2000) for an argument that not all relative clauses should be generated by the raising analysis. Relatives formed by late attachment cannot possibly be raising structures.

[18] There are predictions of this account that are not pursued here. It is predicted that in languages like Tlingit, for example, that constituents insulated by Q could pass through Spec TP positions (on their way to Spec CP) if no external argument is present, a prediction that I have not tested. If this is not possible, then some further investigation would need to determine if this is a true challenge to the theory here, or a result that arises from independent factors. See section 7 where other conditions that could lead to grammatical improper movement are explored.

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

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

[21] 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.

[22] 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.*

[23] The smuggling analysis of *tough*-movement proposed by Hicks (2009) is not 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. A smuggling is discussed and ruled out in section 3.5. Even if a smuggling analysis were correct for *tough*-constructions, this form of analysis does not appear to extend to (54b-d). Recent work by Selvanathan (2017) suggests that complex adjectivals are formed by the same form of predicate formation found in clefts, and thus *tough* constructions are not so anomalous.

[24] The generalization about sentential complements is disputed by Safir (1999), but Safir ultimately accepts a complement adjunct distinction based on other evidence. Illustration of

the effect with contrasts between sentential complements and adjuncts based on the more commonly assumed Principle C effect generalization is offered here for convenience of presentation.

[25] In diagrams (61a,b) and those subsequent to them in the text, the non-terminal node labels are descriptive only. Specifically, X' describes a non-terminal of type X that immediately dominates the head X in the position where X is assigned its complement. XP describes any projection of X that immediately dominates X'.

[26] Note that if this operation were to apply after a DP mergest to T' to form TP, the head of the non-terminal in Spec TP would still be identified as null H, so such an operation would not change susceptibility to EPP.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] Another possibility would be to try a version of the analysis proposed for Japanese, but then the issue arises as to why long distance movement in Japanese behaves differently from long distance movement in Kilega.

[32] A reviewer points out that the PNC might be consistent with van Urk's account, but if so, there is much in van Urk's account of the Dinka facts that may be redundantly derived by insulation.

[33] 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.

[34] 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.