Merge, Labeling and their interactions[*]

Samuel David Epstein, Hisatsugu Kitahara, T. Daniel Seely

Abstract: This paper reviews and discusses a series of papers by Epstein, Kitahara and Seely, related to Chomsky's (2013, 2014) 'labeling by minimal search' analysis. After providing a brief history of 'labels,' some empirically (and explanatorily) advantageous consequences of Chomsky's labeling by minimal search analysis are revealed, including that (i) it explains 'obligatory exit' in A-movement without any reference to Merge-over-Move, lexical arrays and subarrays, nor in fact to the construct 'phase' (motivated in Chomsky 2000), at least suggesting the possibility of their eliminability, and (ii) it explains 'obligatory halting' in key instances of criterial freezing (without appeal to the analytical apparatus proposed in either Epstein 1992 or Rizzi 2014). These results are consistent with the twin (yet often implicit) goals of: (i) reducing Merge to its simplest and most unified form (with no labels nor label projection, as (to our knowledge) first proposed in Collins 2002, Seely 2006) while (ii) concomitantly maximizing Merge's explanatory effects (postulating as few operations as possible beyond Merge). It is important to note that this research is entirely continuous with the 65 year old (scientific) enterprise of seeking to construct an explanatory theory of the format of descriptively adequate transformational and phrase structure rules (now unified under Merge) and to also explain the nature of the (apparent) constraints on transformational rule application, including when transformational application is obligatory ("obligatory exit") and when it is prohibited ("freezing"), and why.

1. Introduction

This paper provides a brief (and selective) history of the nature, motivation and use of labels within the generative tradition and then explores recent developments regarding labeling, focusing on Chomsky's (2013, 2014) labeling by minimal search analysis. More specifically, we review, and add some speculative extensions to Epstein, Kitahara, Seely (EKS) (2014, 2015, to appear).

[*]We thank the volume editors and conference organizers, Leah Bauke, Andreas Blümel, and Erich Groat for helpful comments. We are indebted to Noam Chomsky, Chris Collins, and Ezra Keshet for valuable discussion. We'd also thank the audience of the Labels and Roots Workshop at Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, Philipps-Universität Marburg, Germany 2014, and particularly, Stefanie Bode, Hagit Borer, Dennis Ott, and Marc Richards. Finally, many thanks to an anonymous reviewer for insightful comments and suggestions that helped improve the final version.

2. Labeling By Minimal Search

In this section we briefly review the recent labeling by minimal search analysis of Chomsky (2013, 2014), which provides the point of departure for EKS (2014, 2015, to appear). EKS adopts the labeling analysis of Chomsky and then provides further positive (and unnoticed) consequences of that analysis.

2.1 A short history of labeling

Before presenting the details of Chomsky's (2013, 2014) labeling analysis and our extensions of it, we first provide a selective history of the notion 'label' in generative grammar, from the PS rules of Standard Theory through X-bar theory, to Binary and Singulary Generalized Transformation, Internal and External Merge and finally to unified and simplest Merge.[1]

In part, this is a history of the simplification of a central aspect of syntactic[2] theory, namely that labels were explicitly represented in the syntactic objects that constitute the representational output of the structure building mechanism(s). But, over time, labels and label projection were eliminated from the syntax. The structure building mechanisms have changed over the course of the development of generative grammar and, with these changes, we find different notions of label and label projection. A number of researchers, including Collins (2002) and Seely (2006), argue for the elimination of labels and labeling entirely. In Chomsky's most recent work, however, the effects of labels, which are not explicitly represented in syntactic representations, are derived from the application of independently motivated, 3rd factor mechanisms (specifically minimal search), and with interesting empirical consequences.

2.1.1 PS rules in the Standard Theory

Recursive PS rules of the Standard Theory (see Chomsky 1957, 1965)

[1] This discussion of the history of labeling draws extensively from EKS (to appear).
[2] See Collins (this volume) for important related discussion.

provided a revolutionary solution to the cognitive paradox of discrete infinity: while the human brain is finite, the generative capacity of any I-language (representing an individual's knowledge of language) is infinite. A finite set of recursive PS rules (or a single recursive rule itself, see below) provided the means to generate an infinite number of mentally representable abstract structures and thus provided an explicit representation of human knowledge of syntactic structure and accounted for the fundamental "creative aspect of language use," while playing a central role in the (re)birth of the cognitive sciences and the development of computational-representational theories of mind.

A recursive structure-building mechanism of some type is necessary for any adequate theory of I-language. But of course, one central question[3] is: "Why do we find these particular (construction-specific or category-specific) rules, and not any of an infinite number of other PS rules, or other types of rules?" Why, for example, does a rule like (1) have the properties it has?

(1) VP → V NP

For instance, Why is the 'mother node' on the left labeled VP (and not some other category or, for that matter, some non-category)? And more generally still, why is there a label at all? Within Standard Theory, these questions were not asked; rather PS rules were axiomatic and any single phrasal category could be rewritten as any sequence of categories and thus the existence and categorial status of mother labels were pure stipulation, true by definition. So, for example, Standard Theory allowed a rule like (2)

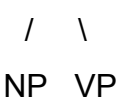
(2) S → NP VP

in which the mother node S is not a projection of (i.e. it is categorially unrelated to) its daughters.[4] This in turn raised the question: why do we find such headless

[3] These questions were not asked at the time. We explore them in hindsight; they largely emerge with the Minimalist Program, see Epstein and Seely 2006 for discussion.
[4] Note that there is an "unprojecting headed" rule like (i) in Chomsky 1981:
(i) S→NP INFL VP.

phrases as S, while the major lexical phrasal categories seem to have heads, e.g., V in VP, N in NP, etc.? The (only) answer available at the time was: it seems to be true by definition, hence by stipulation, i.e. we have no explanation.

Note that besides being stipulative, there is arguably a formal, or interpretive unclarity concerning the relationship between PS rules and the PS trees generated by applying rules. For example, rule (2) contains one and only one formal symbol "S". However in the PS tree (3) generated by applying the rule (2), there are two entities we call "S":

(3)


That is, in the tree representation (3), there appears the label "S" (appearing immediately above NP and VP), yet in addition, the entire tree itself is called 'an S'. This disparity, between the rule and the representation, has perhaps engendered confusion concerning the nature of PS generation vs. PS representation.

Furthermore, the mother nodes of (1) and (2) on the left, and phrasal category labels in general, involve, at least in one sense, 'look ahead.' Standard Theory appealed to 'top down' PS rule application, but as pointed out by Chomsky (1995b), attributing the insight to Jan Koster,[5] such PS rules are telic in that they indicate the categories generated by the syntax that *will* be relevant to the

---

Here we have not only S being projected from no head, but in addition a head (INFL) which fails to project.

[5] Chomsky (1995b) states "Some of the general constraints introduced to reduce the richness of descriptive apparatus also had problematic aspects. An example is Emonds's influential structure-preserving hypothesis (SPH) for substitution operations. As has been stressed particularly by Jan Koster, the SPH introduces an unwanted redundancy in that the target of movement is somehow 'there' before the operation takes place; that observation provides one motive for nonderivational theories that construct chains by computation of LF (or S-Structure) representations. The minimalist approach overcomes the redundancy by eliminating the SPH: with D-structure gone, it is unformulable, its consequences derived … by general properties of Merge and Attract/Move" (Chomsky 1995b, 318). See below for further comment on the shift from *Aspects* to early minimalism.

interpretive components, PF and LF.[6]

As discussed in detail in EKS (to appear), such look ahead is particularly evident given Chomsky (1965) *Aspects'* postulation of the empty Δ node, combined with substitution transformations:[7]

> "… suppose that (for uniformity of specification of transformational rules) we add the convention that in the categorical component, there is a rule *A* → Δ for each lexical category *A*, where Δ is a fixed 'dummy symbol.' The rules of the categorical component will now generate Phrase-markers of strings consisting of various occurrences of Δ (marking the positions of lexical categories) and grammatical formatives." *Aspects*, p. 122.

So, consider passive: Δ would appear in the (simplified) deep phrase marker associated with passive,[8] as in (4)

(4) $[_S [_{NP} \Delta]$ was arrested $[_{NP}$ the man]]

The object NP then raises via substitution to the pre-existing Δ subject NP generated in the Deep Structure (DS), yielding:

(5) $[_S [_{NP}$ the man] was arrested $[_{NP}$ the man]]
^_____|

In effect, Δ is an empty and non-branching maximal projection with a purely formal status, lacking in lexical (nominal) featural content, i.e. it is a projection of no head at all, raising one of the problems noted with respect to S in rule (2) above. The DS in (4) in fact 'preordains' the categorial structure of what the Surface Structure (SS) *will be*. If such structure preserving Δ substitution is employed, then the label of the NP subject of S is already present at DS, 'awaiting' the obligatory arrival of *the man*. This encoding of SS into DS threatens the concept of level itself, suggesting that levels are in some sense intertwined, or non-existent (as was later postulated in Chomsky 1993, Brody 1995, Chomsky

[6] Such labels are also relevant to Standard Theory's category-specific operations in the syntax.
[7] It should be noted that the Δ node played an important role in another crucial development in *Aspects*; namely, the separation of the lexicon from the syntactic component.
[8] We are simplifying what is meant by 'passive' in this context. At the time passive was analyzed as involving two transformations—one moving the deep subject into the *by*-phrase and the other moving the object into the vacated subject position.

2000, Epstein et al. 1998, Uriagereka 1999).

Overall, then, recursive PS rules of the sort found in Standard Theory provided an empirically motivated, profound answer to a paradox, and solved the fundamental cognitive problem of discrete infinity. But, the nature of labels and projection raised a number of important (and unanswered) questions.

2.1.2 X-bar theory: the elimination of PS rules

X-bar theory represented a major development in the history of phrase structure, and specifically for our purposes here, in the history of the notion phrasal label.[9] X-bar theory attempted to provide answers to (at least) some of the questions raised by PS rules. Rather than the stipulated, hence non-explanatory PS rules of Standard Theory, the X-bar format imposed clear restrictions on, and provided a uniform analysis of, 'humanly possible phrase structure representation,' eliminating PS rules,[10] and leaving only the general and uniform X-bar format as part of UG.

The three central tenets of X-bar theory are endocentricity, cross-categorial uniformity, and (in the most widely adopted version) ternary levels of projection. All phrases have a lexical head and they all have the same basic internal structure, as encoded in the X-bar template. Also, as compared with standard PS rules, another essential property of X-bar is a *third* level of projection, neither lexical nor a full phrase, namely, X-bar.

Endocentricity was assumed without exception: since some categories

[9] See Chomsky 1970, Jackendoff 1977, Stowell 1981, among others; see also Chametzky 2000 Chomsky 1995a, b, and Lasnik and Londahl 2013 for discussion.

[10] There seem to be two views of X-bar theory in the literature. One is that there are PS-rules, it's just that they are reduced to the absolute minimum (expressed in X-bar theoretic terms; thus, for example: XP → Spec X'; X' → X YP). The other view, which we assume, is that the X-bar template is a filter, the (implicit) assuming being that there something like Generate-alpha that produces the syntactic objects to be filtered. Just like Move-alpha, Generate-alpha can build any phrase structure it wants, and only those satisfying the X-bar template survive. In retrospect, some general structure building rule must have been assumed, but there was no point of discussing it, since no matter how structures are build, only those X-bar compliant structures survive. Thanks to a reviewer for requesting clarification on this point. As Epstein & Seely (2002) p. 6 note: "GB, as contrasted with the Standard Theory, is traditionally assume to be representational, characterized as a 'virtually rule-free system' (Chomsky 1986: 93). But 'virtually rule free' isn't 'rule free.' Indeed GB theory did have rules, including, for example, Move-alpha and whatever generated structures that could comply with (or violate) the X-bar schema."

seemed to be endocentric (the lexical categories VP, NP, PP, AP, *etc*.), it was assumed that all categories, lexical and functional alike, are endocentric, thereby expressing cross-phrasal uniformity.[11] "Headless" PS rules, like (2), S→ NP VP, are thus eliminated, reduced to the X-bar template and thus 'forced' to have a lexical head.

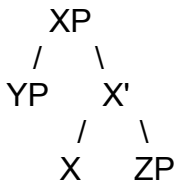
Another crucial innovation of X-bar theory, representing a profound step in the development of the strong minimalist thesis, is the elimination of linear order from the PS component; X-bar theory specified no linear order of elements within the syntactic structure. By contrast, Standard PS rules simultaneously defined two relations, dominance and precedence, and therefore the application of a single PS rule could not (in retrospect) be a primitive operation since two relations, not one, are instantaneously created. X-bar theory takes an important step in reducing the two relations to one, and it does so by eliminating linear order, which is a property of PF and (by hypothesis) not a property of LF. X-bar theory thus disentangled "dominance" (in hindsight, a misnomer, better characterized as 'set membership' in more recent work, and avoiding the 'confusion' noted above concerning the difference between a label vs. the entire category the label is the name of) and precedence. In addition, it sought to explain their existence, and the non-existence of all other relations, as required by, hence subservient to, the interfaces (dominance for semantics, precedence for phonology).[12]

What is the nature of a label in X-bar theory? Clearly, the mother is predetermined. Assuming binary branching, if α is non-maximal (i.e. a head or a non-maximal projection of a head), its mother will be the category of α. If α is

[11] Of course, the asymmetry could have been resolved by alternatively assuming no category is endocentric. In current work the asymmetry (some phrasal categories like VP are endocentric and some categories, like S, are not) is simply a consequence of recursive simplest Merge – with simplicity of the generative procedure, not uniformity of its representational output, being *the* or at least *an* explanatory criterion (see below for further comment; and see Narita (2011)).

[12] It should be noted that the elimination of word order from syntax did not happen all at once. For example, the head parameter made explicit reference to word order. The 1980's also saw the notion of directionality of government (Travis 1984). In Chomsky 2007, with the advancement of the primacy of CI hypothesis, however, it is clearly suggested that order is out of syntax; (or that an optimal syntax is "out of order"!). it is part of externalization. And we have Chomsky's revolutionary hypothesis, consistent with the primacy of CI, i.e. language is primarily 'for thought' and not 'for communication;' see Chomsky (2013). For important, and influential, work on the 'removal' of linear order from the syntax, see also Kayne (1994).

maximal, its mother will be the category of α's sister. Thus, with respect to the following tree representation (ignoring order)

(6)


since X and X' are non-maximal,[13] each will itself project. YP is maximal and hence its mother is the category of YP's sister (in this case X-bar). Projection from a head (i.e. endocentricity), and the syntactic representation of projection, are taken to be central concepts of X-bar theory, defining two core relations: Spec-head and head-complement.[14]

Notice however that Δ – the preordained landing site for movement of a maximal projection – as originally introduced in *Aspects* implicitly remains in the X-bar format. Under X-bar theory, the landing site of movement is often called "Spec", but "Spec" is in effect a cover-term for Δ as well. So, we could say that Δ was still assumed for movement under X-bar theory, i.e. X-bar consistency was a constraint also imposed on transformationally derived structures in which projection is determined by the X-bar schemata: *a moving category has no chance to project* — the mother of the mover 'landing in' Spec is by definition not a projection of the mover.[15]

X-bar theory represented an important advance but raised a new set of questions: specifically, why is there projection at all, and why should it satisfy X-bar theory? Why does the mover never project (if that is in fact true)? Why are phrases endocentric (if they, in fact, all are)? And why are phrasal labels represented in the narrow syntax; are they in fact required syntax-internally? – and has there been a continued confusion between the label vs the category

[13] See Muysken 1982 on relational definitions of maximal and minimal categories.
[14] As we will see below, the available relations changes dramatically under Chomsky's (1995a) "Bare Phrase Structure," particularly the Spec-head relation, which under recent analyses based on simplest Merge, does not exist.
[15] Note that in this discussion we do not consider head movement nor adjunction (neither of which is movement to Spec, and neither of which involves Δ), but see May 1985 for a theory of (segmental) adjunction seeking to render it X-bar consistent.

bearing the label, as discussed above, regarding (3)?

2.1.3 The initial transition from X-bar to Merge

Early minimalism brought major shifts in the architecture of the computational system for human language and initiated changes in the mechanics of structure building. Chomsky (1993), for example, eliminates DS (and paved the way for the elimination of syntax-internal levels entirely, including SS, see also Chomsky 1986). Chomsky (1993) also saw the re-introduction of Generalized Transformation (GT), a structure building operation the output of which is required to be consistent with X-bar schemata by definition. In the new theory, there are two distinct kinds of applications of GT. Binary GT takes two separate syntactic objects and combines them into a single object. Binary GT is thus the 'ancestor' of what would become External Merge. Singulary GT is the precursor of its most immediate descendant, Internal Merge, where one of the objects being made a member of a newly created set is initially contained within the other. In effect, the elegantly constrained X-bar theory, together with its stipulated (or axiomatic) properties including its prohibition on mover projection, was taken to be a UG filter on both DS-level representations and on transformationally derived output representations, another form of unification (of DS and transformationally derived SS PS representations)

2.1.4 The eMERGEnce of bare phrase structure

While X-bar theory represented a very significant step in the continued quest for explanation, it was of course not exempt from explanatory scrutiny. Why should X-bar theory hold? Why do we find these particular relations (endocentricity, trinary projection, mover non-projection, head-complement, and Spec-head--the latter falling under the general definition of "government," the cross-modular, unifying but quite complex, see Chomsky 1986, sole relation--as opposed to an infinite number of alternative phrase structure systems? Adhering to Minimalist method (see Chomsky (2007) "Approaching UG from Below"), we can ask: how "should" phrase structures be generated under minimalist

assumptions?

In "Bare Phrase Structure" (BPS), Chomsky (1995a: 396) provided an initial answer:

> "Given the numeration N, C_{HL} may select an item from N (reducing its index) or perform some permitted operation on the structure it has already formed. One such operation is necessary on conceptual grounds alone: an operation that forms larger units out of those already constructed, call it Merge. Applied to two objects α and β, Merge forms the new object γ. What is γ? γ must be constituted somehow from the two items α and β; ... The simplest object constructed from α and β is the set {α, β}, so we take γ to be at least this set, where α and β are constituents of γ. Does that suffice? Output conditions dictate otherwise; thus verbal and nominal elements are interpreted differently at LF and behave differently in the phonological component ... γ must therefore at least (and we assume at most) be of the form {δ, {α, β}}, where δ identifies the relevant properties of γ, call δ the label of γ." BPS p. 396

Merge was introduced as the central structure building operation of the narrow syntax (NS), necessary on conceptual grounds alone, and the simplest object γ constructed from α and β by Merge was taken to be the set {α, β}. However, as stated in the above excerpt, Chomsky (1995a) assumed that the set {α, β} was in fact descriptively inadequate; it was assumed that empirical adequacy demanded some departure from the simplest assumption (the standard scientific tension between explanation and 'empirical coverage'); that is, the set must be labeled as in e.g. {δ, {α, β}}, where δ identifies the relevant properties of the entire set, i.e. such identification is required given output conditions imposed by the interfaces; thus, for example, {n, {α, β}} is identified as a nominal object, while {v, {α, β}} is identified as a verbal object.

Interestingly, note that in the above passage from Chomsky (1995a) the argument for labels mentions only their necessity at the interfaces, and does not mention any reason for requiring them, as had always been assumed, in the NS.[16] We return to the status of labels in NS, momentarily.

[16] And, in fact, as argued in Seely (2006) since labels, as defined in Chomsky (1995a, b), are not syntactic objects/terms, they are inaccessible to syntactic operations and are thus "syntactically

Chomsky (1995a,b) did not discuss exactly how Merge operates to form the labeled sets {δ, {α, β}}, but he assumes that either α or β may project (in principle), but if the wrong choice is made, deviance results. Projection, then, is a defining suboperation of Merge, but it is 'free,' to project either the head of alpha or of beta, with the result subject to output conditions.

Notice that Chomsky (1995a,b) eliminated both the ∆/Spec node of Standard Theory and X-bar theory. Projection, however, is still present; projection invariably applies by definition and is thus stipulated. Merge was defined as Merge (X, Y) → {Z, {X, Y}}, where Z is either the head H(X) of X or the head H(Y) of Y, and under this definition, it was guaranteed that the label (= projected node) is either H(X) or H(Y), again by definition.

To sum up, (i) under 'top down' phrase structure grammar with ∆-substitution, a moving category has no chance to project by definition; a mover (in e.g. passive) arrives in a preordained NP position whose mother node S is also pre-determined, and categorically distinct from the mover (NP) (ii) under X-bar theory with Spec/∆ substitution, a moving category has no chance to project, again by definition, (iii) under GT with ∆ now internal to it and X-bar theory as an "everywhere" output constraint on GT application (Chomsky 1993), a moving category 'still' has no chance to project by definition, but (iv) under Merge (X, Y) → {Z, {X, Y}}, where Z is either H(X) or H(Y), either a hosting category a moving category can project.[17]

Though for the first time permitting mover-projection, it is important to note as does Seely (2006) that Merge appears at this stage to remain non-primitive in that it simultaneously creates two formal objects anew: (i) it creates a set {X, Y} that was not present in the input to Merge, and (ii) it creates a second set, {Z, {X,Y}} (where Z is identical to H(X) or H(Y); thus we would have {H(X), {X, Y} where H(X) is the label, or {H(Y), {X, Y}} where H(Y) is the label). Note that the

---

inert." Labels in Chomsky (1995a, b) are not terms (as 'term' is defined there) and hence (informally speaking) 'don't exist.' See Seely (2006) for detailed discussion.

[17] Certain previous analyses of mover-projection include projection of a moving (*wh*-phrase) maximal projection (Donati 2006) and the projection of a moving verb at LF (Epstein 1998). See also Hornstein and Uriagereka 2002.

second set, {Z, {X, Y}} expresses the relation 'Z and {X, Y} are sisters' (i.e. co-members of the set that contains them) representing the projection of Z as the label of {X, Y}.

As we will see momentarily, although the label Z was, as just noted, assumed to be necessary, at the same time a definition of syntactic object/term was also necessary which had the specific intent of (correctly) excluding labels from the class of syntactic objects/terms. In effect, the theory implicitly hypothesized the necessary presence but concomitant invisibility of labels.

2.1.5 Toward Simplest Merge: the elimination of labels and projection from the theory of syntactic mental representation

The strong minimalist thesis (SMT), presented by Chomsky (1993, 1995a, b) and elaborated by Chomsky (2000) and in subsequent work, takes the computational system for human language to be a "perfect system", meeting the interface conditions in a way satisfying third factor principles.[18] This is of course not an "assertion" but a hypothesis deemed worthy of exploration on a number of methodological grounds common to normal science.[19]

Under SMT, therefore, the combinatorial operation of the generative procedure assumes (by hypothesis) the simplest formulation in what comes to be called "simplest Merge", a set-formation device that takes X and Y, and forms {X, Y}.

(7) Merge (X, Y) = {X, Y}[20]

To the best of our knowledge, Collins (2002) was the first within the generative

[18] See Epstein (2007) for discussion of the idea that the theory is an "I(nternalist)-functional" (or physiological) one in the sense that the rules apply only "in order to" satisfy the interfaces. See also Chomsky (2007, 2008, 2013, 2014) for discussion of the idea that operations freely apply as long as they conform to the laws of nature. For detailed discussion of these ideas, see EKS forthcoming.

[19] See Chomsky (2014) regarding standard misguided criticism that "biological systems are 'messy'—so they cannot be perfect."

[20] More accurately, (7) should be understood as the simplest instantiation of the combinatorial operation under SMT; a number of properties of (7) follow from 3rd factors—e.g., the No Tampering Condition and the Inclusiveness Condition (see EKS forthcoming).

tradition to propose that labels be eliminated from the representation of syntactic objects and thus that the output of Merge (X, Y) is {X, Y} and not {Z, {X, Y}}. Taking Collins as his point of departure, Seely (2006) reanalyzes the matter derivationally, arguing that the rule simplest Merge (i.e. Merge (X, Y) creates {X, Y}) is motivated on minimalist grounds alone and simplest Merge entails the the elimination of (any type of) projection-suboperation within Merge, thereby entailing the (Collinsonian) postulation of the absence of syntactically represented labels:

> "It is important to stress that, viewed derivationally, it is not labels and projection that are eliminated in and of themselves, RATHER WHAT IS ACTUALLY ELIMINATED ARE TWO SUBOPERATIONS OF THE "COMPLEX" OPERATION MERGE. It is a consequence of adopting the "simplest" version of Merge, namely, [Merge (x, y) = {x, y}], that there are no phrasal labels nor projections, i.e. it is a consequence of the simplification of Merge that phrases are represented as in [{x, y}], and not represented as in [{z, {x, y}}]. I'll argue that this simplification of Merge is motivated on Minimalist grounds. The absence of labels is an immediate consequence of a well-motivated simplification of a fundamental, and arguably necessary, structure building (derivational) operation, namely Merge as in [Merge (x, y) = {x, y}]. In short, the question I am asking is: If indeed [{x, y}] is the "right" type of representation, what is the nature of the generative procedure from which the relevant properties of these representations could be deduced?" Seely 2006, p. 193.

Seely (2006) argues that if Merge creates the only relations, then since labels (as in Chomsky 1995a, b) are in fact not merged, they are in no relation to anything; i.e. Seely seeks to deduce their absence from independently motivated proposals.

Summarizing to this point, we've traced the development of labels and projection, from their original postulation in Standard Theory PS, to their hypothesized elimination. What now of Chomsky's most recent analyses?

3. Label elimination in Chomsky's (2013, 2014) labeling-by-minimal search analysis

This section reviews Chomsky's labeling by minimal search analysis first tracing the basic ideas of the analysis (section 3.1), and then (section 3.2) briefly

outlining the conceptual and empirical advantages of it, as presented in Chomsky 2013, 2014. This sets the stage for our review of our extensions of the analysis (in sections 4 and 5, which review EKS 2014 and EKS 2015 respectively).

3.1 What are labels for Chomsky (2013, 2014)?

Chomsky (2013, 2014) develops an important new analysis of 'labeling,' and provides conceptual and empirical advantages of it. The basic, intuitive idea is that 'labeling' is nothing other than 3rd factor minimal search finding relevant object-identification information within the (representationally unlabeled) set that constitutes the output of (simplest) Merge. Label projection is not part of the definition of the structure building operation Merge, and labels are as a result, not explicitly generated, qua labels, in the output representation of the syntactic object that Merge produces. Just what are 'labels', then, and what role do they play in Chomsky's analysis?

First, for Chomsky (2013, 2014) Merge is maintained in its simplest form, as advocated by Seely (2006) (as just discussed) namely, Merge (X, Y) = {X, Y}. Consonant with minimalist methodology and in particular the strong minimalist thesis, Merge is deconstructed to only what is virtually conceptually necessary. Merge takes two (and only) two objects and puts them into the set {X, Y}, thereby creating the relation 'member of' for X and Y. As discussed, linear ordering is relegated to the phonology and as in Collins (2002) and Seely (2006) there are no syntactically encoded labels[21] since there is no projection as a defining property of the operation Merge. What then are 'labels' under Chomsky's system? This is an important question since, as we've seen above, labels have an empirical motivation in that they provide the information that categorically distinguishes one phrasal category from another, differentiating a VP from an NP, for example; and this identification is necessary, according to Chomsky, for proper interface interpretation.[22]

[21] More specifically, there is no *representationally dedicated symbol serving as the label*, and only the label, of a syntactic object.

[22] Traditionally, labels were assumed to be required in the syntax; what were considered syntactic operations appealed to labels in the syntax. Consider, for example, that *one* and *do so* were

For Chomsky (2013, 2014), labeling is the process of finding the relevant information within the set, {X, Y}, which identifies the categorial status of the entire set generated by simplest Merge. Labeling is "just minimal search, presumably appropriating a third factor principle, as in Agree and other operations" (Chomsky 2013). Again, there are no labels resulting from the application of a projection-suboperation internal to Merge, representationally dedicated to labeling the set. Rather there is just minimal search. 'Labeling' then is simply the name given to the result of an independently motivated minimal search procedure, itself 3rd factor and hence not stipulated.[23]

Note further that Merge applies freely. Unlike earlier stages of minimalism, where operations applied 'in order to' satisfy output conditions, Merge is completely optional. It applies or it doesn't, and if it does, it applies for no other reason than because 'it can.' Of course, Merge is subject to 3rd factor principles (i.e. laws of nature), but, these are not constraints built into Merge, rather Merge can't help but conform to them. Thus, Merge (X, Y) leaves X, Y unchanged as a result of the 3rd factor No Tampering Condition (NTC), and by the (3rd factor) inclusiveness condition, "no new objects are added in the course of computation

argued to substitute for an N' and a V' respectively. Under current assumptions, it is hard to implement "one-substitution" or "do-so replacement" as transformational rules. So, by hypothesis such proforms are simply created by Merge (selecting lexical items). Then, the contrasts exhibited above come down to the question: what is wrong with e.g. merging "that" with "do so" (forming "I like this theory more than that one/*do so") instead of "one" (forming "that one") – presumably, a CI-interface problem, concerning the semantic interpretations of certain proforms. Labels were also required for C-selection and their absence in Chomsky's recent analysis raises interesting questions about how empirical phenomena motivating C-selection are explained. Collins (2002), for instance, articulates the role of labels in various subsystems of GB theory and then attempts to derive their properties without labels (see also Seely 2006). See Hornstein and Nunes (2008) for additional important discussion.

[23] A reviewer raises a series of important issues regarding Chomsky's labeling by minimal search analysis that we too find somewhat unclear and that require further research. One questions has to do with the timing of labeling. We assume, with Chomsky (2013), that labeling (= minimal search) takes place as part of Transfer "Since the same labeling is required at CI and for the processes of externalization (though not at SM, which has no relevant structure), it must take place at the phase level, as part of the Transfer operation." Another question has to do with why labels are required. Again, we follow Chomsky (2013; p. 43) in assuming that "For a syntactic object SO to be interpreted, some information is necessary about it: what kind of object is it?" The interfaces must be able to distinguish, say, 'walk' as an event vs. an 'entity.' As for exactly how this information is used by which interfaces, further research is required.

apart from arrangements of lexical properties" (Chomsky 1993, 1995b).[24]

To see how Chomsky's new labeling analysis works, let's consider two central cases. Suppose first that the syntactic object (SO) is {H, XP}, H a head and XP not a head. Then minimal search will select H as the label, allowing the object {H, XP} to be identified as 'an H' at the interfaces.[25]

It is interesting to note that all phrases of the form {H, XP} are endocentric, not in the sense of representationally projecting the head H (as in X-bar theory), but rather in the sense that properties of the head H serve as the identifiers of the entire object {H, XP}. As with X-bar theory, then, VP = {V, ...}, NP = {N, …}, etc. will have a 'nucleus' where the head of the phrase (in effect) matches the category label of the phrase. Within X-bar theory this followed by stipulation. For Chomsky's labeling analysis, it follows naturally from 3rd factor minimal search and thus endocentricity relative to {H, XP} is deduced, and without the postulation of an X-bar level of projection.[26]

By contrast, suppose SO is {XP, YP}, neither a head (recall PS rule (2)

[24] In early minimalism, all syntactic filters were eliminated. Only naturalized interface filters, called bare output conditions, survived. As for syntactic constraints, they were expected to reduce to principles of economy, now understood as third factor principles. But it didn't go that smooth. So, what we had in early minimalism were: operations (e.g. Merge, Move), third factor principles, and bare output conditions. But operations in early minimalism were very complex. If you look at the definition of Move, for example, it has various sub-clauses beginning with "only if..." In other words, all the syntactic constraints on movement were "stipulated" as part of the defining properties of Move. But in subsequent work, there was some success in reducing those defining properties of Move to the third factor principles, and we now have the simplest formulation of Merge for both Merge and Move. Under the framework of Chomsky (2013, 2014), what we have are: Merge (putting aside Agree), third factor principles (labeling by minimal search, NTC, inclusiveness), and bare output conditions. It is important to note that Chomsky (2013, 2014) adopts simplest Merge; the Merge-internal constraints of the form "Merge applies only if..." are all gone. In this system, "operations can be free, with the outcome evaluated at the phase level for transfer and interpretation at the interfaces" (Chomsky 2014). If this overview is correct, then the definition of Merge/Move has changed, but free application of Merge/Move has remained constant. The shift has been taken place from freely applying *complex* Merge/Move to freely applying *simplest* (unified) Merge. For more detailed discussion, see EKS forthcoming. See also Boeckx 2010 and Ott 2010 for relevant discussion.

[25] Minimal search 'looks into' the set {H, XP} and finds two objects (the members of the set), H and XP. Only H, a lexical item that bears linguistic features, qualifies as a label since XP, a set, does not directly bear linguistic features that could provide object identification information; in short, only an item that directly bear linguistic features can serve as a label.

[26] As discussed in detail in EKS 2015, Rizzi's (2014) analysis of 'halting' – discussed below – crucially relies on the postulation of an X-bar category, which is not compatible with Chomsky's labeling by minimal search analysis, an adaptation from which Rizzi (2014) seeks to deduce halting.

above, S → NP VP). Here minimal search is ambiguous; search finds the sets XP, YP, neither of which is a head[27]; it then searches further, finding both the head X of XP and the head Y of YP. Overall, search does not find a *unique* element (that can provide the needed object-identification information). It is assumed that this ambiguity is intolerable; left as is (an option available under free Simplest Merge), Full Interpretation (FI) is violated at the interface levels. How can this XP-YP problem for labeling be solved? Chomsky (2013) suggests, and explores the empirical consequences of, two strategies: (A) modify SO so that there is only one visible head, and (B) X and Y are identical in a relevant respect, providing the same label, which can be taken as the label of the SO.[28] These strategies, in turn, have important empirical consequences, to which we return below.

To summarize so far, Chomsky (2013)'s analysis assumes that

(i) Labels are required, but only at the interfaces.

(ii) Labeling is just minimal search.

(iii) There must be a single element that serves as the 'identifier' of a syntactic object, ambiguity of identification is not tolerated.

For Chomsky, labels do not 'exist' in and hence can't be referred to in NS.[29, 30] In this system, minimal search identifies syntactic objects by looking at

[27] We assume that only a head (i.e. a lexical item) can provide object-identification information, a set, like XP, YP, cannot. This asymmetry arguably follows from the fact that a head directly bears features, whereas a set does not.

[28] The basic idea is that in {XP, YP}, minimal search equally finds the head X, Y; if these heads share some prominent feature, and specifically if the prominent feature(s) match via Agree (as in, say, phi-agreement or Q-agreement), then that shared feature counts as the label of the set {XP, YP}. Thus, in {NP, {T, {NP, {v, VP}}}}, which is of the form {XP, YP}, minimal search finds the *phi* features shared by the two relevant heads N and T (in finite clauses) after Agree, and thus phi is taken as the label of the set.

[29] More specifically, the labeling algorithm LA (i.e. minimal search) does not engage internal to the syntax and hence there are no 'labels' in NS. LA takes place at Transfer since the same labels are required by the CI and SM interfaces. As notes in Chomsky 2014, "since the same labeling is required at CI and for the processes of externalization (though not at SM, which has no relevant structure), it must take place at the phase level, as part of the Transfer operation."

[30] As we noted in footnote 21, given the absence of labels in NS, *one* or *do-so* substitution have two distinct problems: first if we have just (unified) Merge, then there is no substitution operation. If we add a substitution operation, arguably a departure from SMT, there may still be problem if syntactic objects are not identified in the course of a derivation (e.g. nominal for one substitution or verbal for do-so substitution).

features/properties of lexical items, so it needs neither postulation of labels (as a separate category) nor implementation of a label identification algorithm that is independent from minimal search. Interestingly, this eliminates what might be seen as the last vestige of 'construction specificity.' Internal to the syntax, there are syntactic objects, namely sets, but the label identification algorithm (i.e. minimal search) applies only at (the transition to) the interfaces. The interfaces need to know the identity of an object and it is minimal search that allows inspection of a set's members to find information relevant to determining the identity of the set. Furthermore, the interfaces can't tolerate 'conflicting information' concerning identity, information such as 'this is simultaneously an N and a V,' rather the interfaces require unique identification information.[31]

Chomsky's recent work then explores certain empirical consequences of this labeling by minimal search analysis; for example, it explains obligatory exit from intermediate positions of A-bar movement.

3.2 Deriving obligatory exit in A-bar movement, Chomsky 2013

Chomsky's (2013) labeling by minimal search analysis provides an interesting new account of obligatory exit from intermediate positions in successive cyclic A'-movement. Consider *wh*-movement as in (8), (where *t* stands for a full copy of an internally merged element):

(8) [$_{\beta}$ In which Texas city did they think [$_{\alpha}$ *t* [C [$_{TP}$ the man was assassinated *t*]]?

Suppose that the *wh*-phrase *in which Texas city* (hereon, *wh*-PP) is internally merged to the Spec(ifier) of the embedded C and stays there. Then, the embedded clause α is of the form of {XP, YP}, where XP is the *wh*-PP and YP is {C, TP}, as in (9), which is intended to depict the first 'half' of the derivation of (8), and where C in (9) is (to be) selected by *think*, not *wonder*:

[31] Note that this represents a shift from Chomsky 2007, 2008 where in the face of ambiguous information, either option could be chosen, any 'bad' results filtered by the interface. See Chomsky's discussion of, eg, Donati on 'who you like' as in 'who you like is irrelevant' where D projects, or 'I wonder who you like' where CP projects.

(9) … (think) [$_{\alpha}$ [$_{XP}$ in which Texas city] [$_{YP}$ C [$_{TP}$ the man was assassinated *t*]]

This is the {XP, YP} situation reviewed above. The heads X, Y are found by minimal search, resulting, potentially, in ambiguous object identification. Note that if Y (= C) bears no Q feature, then X, Y will not share any prominent agreeing feature, and object identification (i.e. label) failure will in fact result. Thus, if XP remains in this intermediate position, minimal search cannot find a label for α, since there is no prominent feature (e.g. phi or Q), shared by X (the head of *the wh*-PP) and Y (the head of {C, TP}).

What happens if the *wh*-PP raises to a higher position, as in (8)? In this case, the lower copy of the *wh*-PP is "invisible" inside α. Consequently, minimal search "sees" only C and TP, (= {H, XP}) which is therefore labeled C; i.e. minimal search *can* find a unique "visible" head, namely C as the label of α. Notice that the matrix clause β of (8) is also of the form of {XP, YP}, but there is an agreeing feature shared by X and Y, namely the Q feature of the *wh*-PP and the Q feature presumably borne by the interrogative-mood marked C of the matrix ("direct question"); hence, Q can be the label of β.

We see, then, that Chomsky's analysis nicely accounts for the "obligatory exit" from an intermediate position of *wh*-movement. Importantly, exit is not in fact obligatory, but rather, the *wh*-PP is free to remain in the intermediate position, but doing so results in labeling failure (FI violation) at the interfaces. Chomsky's (2013) analysis makes no appeal to a mismatch of features between *think,* which selects a [–Q] C, and the [+Q] *wh*-PP occupying the Spec of this [–Q] C (as in traditional analyses); no appeal is made to an explicit Spec-head relation, defined via "m-command" or "government" or the notion "maximal sister to an X-bar projection" in CP, nor to any of the technical devices that have been non-explanatorily invoked like co-superscripting of Spec and head in CP or to an S-structure level of representation (as in important prior analyses such as Rizzi's (1997) *wh*-criterion or Lasnik and Saito's (1984, 1992) S-structure condition which blocks a [+Q] *wh*-phrase from occupying the Spec of a [-Q] C in English, at

SS). Nor is EPP (obligatory Spec-T or Spec-C) appealed to. No such non-explanatory descriptive technicalia are invoked, nor is the central principle appealed to (minimal search) specifically linguistic, but is rather attributed to third factor.

4. Extensions of the labeling by minimal search analysis: EKS 2014, Obligatory Exit in A-movement:

EKS (2014) argues that Chomsky's labeling by minimal search analysis not only accounts for obligatory exit from intermediate position in successive cyclic A-bar movement, but that it *also* provides an elegant account of obligatory exit from intermediate position in A-movement. The central ideas are as follows.

Consider a typical instance of successive cyclic A-movement, as in (10)

(10) a. a man is likely [$_{TP}$ *t* to be *t* in the room]
b. *there is likely [$_{TP}$ a man to be *t* in the room]

Assume that in (10a) the DP *a man* has moved through the intermediate A-position, Spec of the embedded TP, on its way to the matrix subject position. Such movement is clearly acceptable (analogous to *wh*-movement in (8), *In which Texas city did they think [$_{\alpha}$ t [C [$_{TP}$ the man was assassinated]]]*?). But, what happens if this DP moves to the intermediate position and then stays there, as in (10b), the analog of (9)? Here we would have an {XP, YP} set, namely {DP, TP}, with no chance of finding a label. Specifically, we would have

(11) … likely {$_{\alpha}$ {the, man}, {T, vP}}

Since T is infinitival in (11), it will not bear the phi features necessary for the shared prominent agreeing (phi) feature option of labeling the {XP, YP} structure of α; it is not the case that the 'head' of {the, man} (X of XP) and infinitival *to* (Y of

YP) bear phi features.[32] Thus α will not have a label and a violation of Full Interpretation (FI) at the interfaces will result.

As discussed in EKS 2014, Chomsky's labeling by minimal search analysis naturally extends to these A-movement cases, which were the central motivation for Chomsky's 2000 Merge-over-Move analysis, including the postulate 'phase.' (as well as lexical arrays and subarrays). Merge-over-Move is no longer statable as PS generation and transformational rule application have been unified under simplest Merge. If this labeling analysis of (10) is viable, it suggests the possibility of eliminating the concept phase, at least to the extent that it was based on the analysis of such examples (the conceptual motivation for phases, locality and 'chunking' in general, remain). See EKS (2014) for a detailed review of the history of and motivation for Chomsky's (2000) phase based Merge-over-Move analysis of cases like (10).

5. EKS on Criterial Freezing, the 'halting problem'

Having just discussed 'obligatory exit' phenomena, we now discuss 'obligatory halting' phenomena, cases where a DP has moved to a position with respect to which all of the DP's features are checked. With *wh*-movement, when such a position is reached, the DP is 'frozen' in that position, further movement results in ungrammaticality. Just as we have argued above regarding 'obligatory exit,' we are in fact assuming Merge is freely applied, and if it fails to apply in 'obligatory exit' cases, as just argued for A movement, then FI is violated at CI due to labelessness. So 'obligatory exit' is somewhat of a descriptive misnomer.

[32] We assume that in XP (= 'the man'), the element that counts as X bears phi features. There are open questions about the technical details of this assumption. If *the* is the head of the DP {the, man}, then this D must bear phi (at the point of minimal search); but it's the N that is assumed to inherently bear phi (see Carstens (2010) for important discussion); hence there could be N to D raising to get phi features on D. See Chomsky 2008 for relevant discussion, pp. 25-26. In terms of Chomsky's 2014, assumptions, in case of {the, man}, there must be at least three items involved: D, n, and R, where the lexicon contains R, and nominal and verbal are determined in NS. If we interpret Chomsky's (2007) proposals in terms of Chomsky's (2014) assumptions, then, Merge constructs {n, {D, R}}, D inherits unvalued phi from n, R moves to Spec-D, and finally D moves to n. As a result, we have {<D, n>, {R, {D, R}}. This will give the right order "the man," and the label <D, n> contains the valued phi features.

In this section we will use the term “Obligatory Halt” or “Freezing” in the same way – i.e. we are in fact assuming that Merge application is free, and we seek to explain the anomaly resulting from Merge application to a frozen position, as an interface anomaly. Thus Merge is in fact free to apply or not, and we seek to reanalyze what have been assumed to be constraints on the application of movement as freely applied (unconstrained) movement which sometimes yields interface anomaly(ies), as when move is applied to a category occupying a so-called frozen position.

To begin, consider (12)

(12) a. You wonder [$_{\alpha}$ [$_{XP}$ which dog] [$_{YP}$ C_Q [$_{TP}$ John likes *t*]]].

b. * Which dog do you wonder [$_{\alpha}$ *t* [C_Q [$_{TP}$ John likes *t*]]]?

Historically, such cases have been accounted for with constraints on movement itself. That is, movement out of a criterial/frozen position is prohibited by a constraint barring the syntactic application of movement For example, Epstein (1992) derives such freezing as an arguably deducible (“last resort”) effect of Chomsky’s Strong Minimalist Thesis (SMT) encapsulated as “computationally efficient satisfaction of bare output conditions.” In short, if there is no need to move, then you can’t move, and if in the syntax we have ‘already’ generated what would (or will) be a legitimate LF representation, then syntactic movement is barred, by economy of derivation.

More recently, Rizzi (2014) attempts to explain such freezing phenomena also as resulting from the inapplicability of syntactic movement to certain syntactic objects, by appeal to a particular (re-)formulation of the independently motivated hypothesis that X’ projections are invisible for movement, “the so-called X’ Invisibility hypothesis”. Once a phrase moves to a criterial position, it is argued that given certain modifications of Chomsky’s (2013) labeling by minimal search analysis, movement halting can be explained since a phrase moved to a criterial position becomes “an X’ projection”, hence invisible to syntactic movement. Thus, Criterial freezing is explained by appeal to a constraint on syntactic rule application.

The core of Rizzi's (2014) analysis can be illustrated as follows.[33] Consider (13), where the *wh*-phrase *which dog* has raised to a 'criterial position' (namely, Spec of CP headed by a complementizer C_Q bearing a Q-feature):

(13) $[_Q\ [_Q$ which dog$]\ [_Q\ C_Q\ [_{TP}$ John likes *t* $]]]$

For Rizzi, the result of such movement is the creation of an X-bar, namely the Q-bar "which dog," which is then frozen by X-bar invisbility. That is, (13) represents an {XP, YP} configuration of the form {{which dog} {C, TP}}. The entire structure in (13) is labeled by Q-agreement (i.e., the head of [which dog] Q and the head of [C_Q TP] each bear Q). Thus the entire representation in (13) is labeled Q as indicated. Importantly this makes "which dog" a Q-bar 'intermediate between the head (= Q^{min}) and the entire structure (13) (= Q^{max}). Since "which dog" becomes an X-bar in the course of the derivation, then by X-bar invisibility, it cannot move further; "only maximal objects with a given label can be moved" (Rizzi 2014). Rizzi's analysis thus seeks to (elegantly) deduce freezing from Chomsky's {XP, YP} label-identification algorithm coupled with the independently proposed X-bar Invisibility hypothesis. As EKS (2015) point out, however, there may be a number of conceptual and empirical disadvantages of Rizzi's analysis. For one thing, it requires obligatory and immediate labeling in the syntax; in the example above, for instance, further movement of *which dog* is blocked only if we 'know' in the syntax that *which dog* is obligatorily and immediately an X-bar and hence can't be moved (given X-bar invisibility). But such obligatory and immediate labeling is in fact inconsistent with Chomsky 2013, 2014, which allows complete non-labeling in NS. Note further that Rizzi's analysis has X-bar invisibility as a constraint on Move, but not on Merge, which runs contrary to the attempted unification of Move/Merge.

There is also a potential empirical problem with this analysis as it says nothing about a case like

[33] For detailed discussion of Rizzi's (2014) analysis, and of potential conceptual and empirical disadvantages of it, see EKS 2015.

(14) * I wonder John likes this dog

In the interface-based reanalysis of "freezing" (under what is in fact freely applied Merge) we present below, such cases and freezing are unified, in a way they cannot be under freezing-specific constraints on syntactic movement. For EKS (2015), following Chomsky 2013, 2014, movement is completely free. That is, the *wh*-phrase in a case like (13) can in fact syntactically exit (via application of Merge) a criterial position. However, it's argued that if it does so either label failure or an interpretive problem at the interface will result. In short, EKS seeks to eliminate a syntactic constraint (X-bar invisibilty, which they argue is not formulable in a way consistent with current theory) and reassign its empirical effects to independently motivated interpretive constraints.

Recall from our review of Chomsky that simplest Merge applies freely, subject only to 3rd factor. Merge is an operation that constructs larger units out of those already constructed, and simplest Merge is a 3rd factor compliant instantiation of Merge. Thus, relative to (12a), repeated in (15a) here

(15) a. You wonder [$_{\alpha}$ [$_{XP}$ which dog] [$_{YP}$ C_Q [$_{TP}$ John likes *t*]]].[34]

b. * Which dog do you wonder [$_{\alpha}$ *t* [C_Q [$_{TP}$ John likes *t*]]]?

nothing prohibits the (bottom-up) applications of Merge that would produce (12b), repeated in (15b). In informal terms, the *wh*-phrase is free to move from the intermediate Spec of CP to the higher Spec of CP position. Such Merge does not run afoul of any 3rd factor principle (like NTC nor Inclusiveness). In fact, any constraint on the application of Merge that is not a 3rd factor constraint would represent a departure from the SMT, and hence would require substantial empirical support. As Chomsky (1998) states, "one plausible element of optimal design is that here are no constraints on application of operations."

[34] Note that subscripts (XP, YP) are used here only for ease of reference, indicating the {XP, YP} situation for labeling.

For EKS, freezing of the sort in (12)/(15) is not the result of a constraint on the application of Merge. Rather, EKS argues that the contrast results from independently motivated morpho-phonological and CI requirements for properly interpreting clauses, specifically clauses whose labels are identified as either the interrogative complementizer C_Q (yes/no questions) or the Q-feature shared by the two heads C_Q and WH_Q (wh-questions)—i.e. the shared prominent (Q) feature option of Chomsky's labeling analysis. EKS argue that 'obligatory syntactic halt' in *wh* criterial position is the only way to satisfy these requirements. In short, *wh*-movement from *wh* criterial position (freely applied simplest Merge) is allowed to apply in NS, but if it does, independently motivated morpho-phonological and/or CI requirements are violated.

As pointed out above, EKS adopts the assumption that every syntactic object must be labeled at CI (Chomsky 2013).[35] EKS then proposed the following (minimum) assumptions concerning C_Q: (i) there is only one C_Q in the (English) lexicon, appearing in both *yes*/*no*- and *wh*-interrogatives (Chomsky 1995b),[36] (ii) a CP with the label C_Q, unaccompanied by a "*wh*-specifier," is interpreted as a *yes/no*-question at CI, while (iii) a CP with the label Q, when Q is shared by the two heads C_Q and WH_Q (the latter being the head-feature of a wh-phrase in "Spec-C"), is interpreted as a *wh*-question at CI (Chomsky 2013).

Thus, in a typical *yes/no*-question structure, such as (16)

(16) $[_{\alpha}\ C_Q\ [_{TP}$ John likes this dog]]

the label of the CP will be C_Q by minimal search. Thus, the CP will be interpreted as a *yes/no*-question. However, as a language particular property of English, it is

[35] In Chomksy 2013, 2014, labeling cannot be required in NS, e.g. merger of T and vP must be allowed, yet vP (= {XP, YP}) has no label. By contrast in order to affect syntactic freezing under X-bar invisibility, Rizzi must obligatorily generate labels, and projection types, including D-bar, which to some extent, presupposes obligatory and immediate labeling in NS.

[36] Why would C_Q appear in both *yes*/*no*-questions and *wh*-questions? This might be explained under analyses of *wh*-questions in which they are interpreted as a family of *yes/no*-questions. So, for example, *what did you buy* is interpreted as something like: "Did you buy a car? Answer me yes or no; Did you buy a pen? Answer me yes or no," etc. We are indebted to Ezra Keshet for this idea and for valuable discussion of issues relevant here.

assumed that in order to *actually be* interpreted as a (direct) *yes/no* matrix interrogative, either T-to-C inversion or rising (question) sentential prosody is required.[37] Thus, (16) will have to 'surface' as *Does John like this* dog? or *John likes this DOG?* (with question intonation).[38] Again, this morpho-phonological requirement is a language particular property of English. Now, consider the following case in which (16) is embedded:

(17) * You wonder [$_{\alpha}$ C_Q [$_{TP}$ John likes this dog]].

In (17), there is a C_Q unaccompanied by a "*wh*-specifier." α is then labeled C_Q and hence interpreted as a *yes/no*-question at CI. But in (17) there is a morpho-phonological problem with this state of affairs: in embedded clauses in English, T-to-C is simply unavailable as is rising intonation in English. Thus, α in (17) though required to be interpreted as *yes*/*no*-question in fact cannot be interpreted as a *yes/no*-question. That is, with Chomsky (2014), we assume that when embedded, a *yes*/*no*-question, interpreted in concert with the structure above it, yields a composed representation that is "gibberish, crashing at CI" (Chomsky 2014, see also Chomsky 1995b). Leaving aside whether it is "crashing," i.e. some yet-to-be-proposed unvalued feature appears at an interface, one possibility regarding its status as gibberish is as follows: the CP headed by C_Q is itself interpreted as a *yes/no*-question and so would be interpreted as: "Answer me this: Does John like this dog?" that is, a performative request made of the speaker's interlocutor, for a specific kind of information. As such, embedding it, as in "I wonder John left" yields an interpretation like: "I wonder 'Answer me this, Did John leave?" This is anomalous to the extent that one cannot wonder a request for information. Given this analysis, (16) violates only the English

[37] Presumably, one or the other is needed as an overt indicator of the otherwise undetectable presence of C_Q, as Chomsky (personal communication) notes.

[38] A reviewer points out that there may be a difference between the syntax of yes-no questions with subject-aux inversion vs. the syntax of yes-no questions with rising intonation, noting "that the latter does not license NPIs, unlike the former: Does John have any money? vs. *John has any money? (rising intonation)." Thus, "it might not be the case that both kinds of yes-no questions have a +q C." We leave this interesting issue open here.

morpho-phonological requirement (if neither T to C raising nor rising intonation is applied) while (17) violates the morpho-phonological requirement and is gibberish at CI.

As EKS (2015) points out, this morpho-phonological, CI analysis of (16) and (17) naturally extends to the classic criterial freezing cases considered above and repeated here:[39]

(18) a. You wonder [$_{\alpha}$ [which dog] [C_Q [$_{TP}$ John likes *t*]]].

b. *Which dog do you wonder [$_{\alpha}$ *t* [C_Q [$_{TP}$ John likes *t*]]]?

The converse, however, does not hold—that is, analyses of freezing cases like (18b), including Rizzi's (2014) do not extend to (16), which lacks a *wh*-phrase of any kind thereby exempting it from a freezing analysis, entailing that (18b) and (17) cannot be unified. Under the labeling analysis of Chomsky (2013), in (18a), at CI, the label of α is the Q-feature, shared by the two heads, namely C_Q and the operator WH_Q, and this label Q, accompanied by a "*wh*-specifier," is interpreted as a *wh*-question (an indirect one in (18a)) at CI. In (18b), however, minimal search fails to identify the Q-feature (shared by the two heads C_Q and WH_Q) as the label of α, because the operator WH_Q (= *t*) in α is "invisible" to minimal search. That is, Chomsky (2013) takes WH_Q to be inside α if and only if every occurrence of WH_Q is a term of α. Thus, after *wh*-movement into the matrix clause, the copy of WH_Q in α is "invisible" to minimal search when it searches α for its label-identification (see EKS 2012 for further empirical support of this analysis). Therefore, the analysis proposed here asserts that the embedded clause α in (18b) cannot be interpreted as a *wh*-question, because *which dog* in the "specifier" of the embedded C_Q is "invisible" to minimal search. It instead predicts that the label of α is the C_Q (recall α appears to minimal search as [C_Q TP]), and although selection is thereby satisfied, as *wonder* does select C_Q, α

[39] Epstein (1992) and Rizzi (2014) discusses other cases of "freezing", beyond the core case examined here. Determining the predictive content of the analysis proposed here, regarding all such freezing cases requires further research.

cannot be interpreted as a *wh*-question. So what interpretation does (18b) receive?

EKS (2015) argues that α in (18b) receives a *yes/no*-question interpretation. Recall that a CP with the label C_Q, unaccompanied by a "wh-specifier," is interpreted as a yes/no-question at CI. The hypothesized problems with (18b) are then that T-to-C is unavailable as is rising intonation in English embedded clauses, and when embedded, the larger construction resulting from the embedding, containing a *yes/no*-question as a term, is gibberish (and perhaps crashing) at CI. In short, semantically anomaly at CI results from interpreting an SO that in part means " wonder a (performative) request" (by contrast, of course, interpreting a structure that denotes "wondering if a proposition is true or false" is semantically nonanomalous.

Summarizing, we made the following assumptions concerning English C_Q:

(i) There is only one C_Q in the (English) lexicon, appearing in both *yes/no*- and *wh*-interrogatives.
(ii) Every syntactic object (SO) must be labeled at CI.
(iii) An SO, the label of which is identified as the head C_Q, unaccompanied by a "*wh*-specifier," is interpreted as a *yes/no*-question.
(iv) An SO, the label of which is identified as the Q-feature, shared by the two heads C_Q and WH_Q, is interpreted as a *wh*-question.
(v) English *yes/no*-questions require T-to-C inversion or rising (question) sentential prosody, available only in matrix clauses, and when embedded, the resulting structure cannot be felicitously interpreted; such structures are gibberish (and perhaps crash) at CI.

(i)-(v) are all independently motivated, and to explain both apparent "obligatory syntactic halt" in *wh* criterial position, and cases like "I wonder John left" nothing more seems to be needed. We argued that there is no need to invoke an NS-

specific halting constraint; the "halting" effect, observed in (18b), naturally follows from the independently needed morpho-phonological, CI analysis.[40, 41]

[40] In Japanese, unlike English, raising from *wh* criterial position appears to be permissible. Consider (i) (from Takahashi 1993):

(i) Nani-o Taroo-wa [Hanako-ga *t* katta ka] siritagatteiru no
what-ACC Taroo-TOP Hanako-NOM bought Q want-to-know Q
'What does Taroo want to know whether Hanako bought?'

Given that (i) converges and is interpretable at CI, we suggest that the interrogative complementizer C_Q and the counterpart of "whether" are homophonous in Japanese; they are pronounced as *ka.* Thus, in (i), *ka* is not an interrogative complementizer C_Q; rather, it is the Japanese counterpart of "whether" as the translation indicates.

[41] The labeling analysis, developed here, sheds new light on partial *wh*-movement. Consider the following German data (from Sabel 2000):

(i) a. [$_\beta$ Was [C_Q meinst du [$_\alpha$ wen [C Peter Hans *t* vorgestellt hat]]]?
WH think you$_{nom}$ who$_{acc}$ P$_{nom}$ H$_{dat}$ introduced has
'Who do you think Peter has introduced to Hans?'
b. [$_\beta$ Was [C_Q meinst du [$_\alpha$ wem [C Peter *t* die Leute vorgestellt hat]]]?
WH think you$_{nom}$ who$_{dat}$ P$_{nom}$ the people$_{acc}$ introduced has
'To whom do you think Peter has introduced the people?'

It is generally assumed that *was* is not a *wh*-phrase; it is a *wh*-expletive that functions as a scope marker; and the *wh*-phrase *wen/wem* "who$_{acc}$/who$_{dat}$" is interpreted at the matrix CP, thanks to this *wh*-expletive, even though the *wh*-phrase is located in the embedded CP. From the labeling point of view, however, if the *wh*-phrase headed by the WH_Q remained in α and appeared there at CI, a labeling failure would result, contrary to fact. So, what is going on? One possibility is that even though the WH_Q (or the phrase containing it) can remain, violating FI at CI, in (ia,b) the WH_Q (or the phrase containing it) can choose an option of moving out, allowing α to be labeled. Pursuing this possibility, what is left behind by such movement may in fact be only the pronominal material of the *wh*-phrase, including *phi* and Case; it is no longer the *wh*-phrase headed by the WH_Q. One possible implementation of this might be to apply Obata and Epstein's (2011) "Feature splitting Internal Merge" hypothesis.

In this regard, Dutch provides an interesting case. Instead of *wie* "who", the pronominal element *die* can appear in α, as in (ii) (from Boef 2013):

(ii) a. Ze vroeg wie jij denkt [$_\alpha$ wie het gedaan heeft]
she asked who you think who it done has
'She asked who you think has done it.'
b. Ze vroeg wie jij denkt [$_\alpha$ die het gedaan heeft]
she asked who you think DEM it done has
'She asked who you think has done it.'

If the structure of this A' pronoun is analyzed as "WH_Q + pronominal material," then the WH_Q (or a phrase containing it) moves out of α to form the label Q of the matrix clause, leaving its pronominal content behind, and such non-Wh, Non-Q pronominal content gets pronounced as *die* in Dutch, leaving the door open for a way to circumvent a labeling failure in the embedded clause. See Obata 2014 for detailed discussion.

6. SUMMARY

In this paper we have reviewed a number of recent papers, EKS 2014, EKS (2015) and (to appear), tracing first the history of labels from PS rules of Standard Theory to Chomsky's recent labeling by minimal search analysis. Labels have gone from being stipulated (and thus non-explanatory) constructs of PS rules to being nothing other than the result of 3rd factor minimal search. Chomsky's recent labeling by minimal search analysis was explored and the extensions of proposed in EKS 2014 and EKS (2015) were presented. We've seen that Chomsky's analysis accounts for 'obligatory exit' in successive cyclic A-bar and in A movement. In both instances, the mover continues out of an intermediate position to avoid label failure. We've also seen that 'obligatory halt,' i.e. freezing, can be accounted for, by interface problems—once in a 'criterial position', further movement would induce e.g. CI anomaly (and hence 'gibberish'). Most importantly, all of these positive empirical results are obtained appealing only to simplest Merge operating within 'natural laws' (3rd factor principles such as minimal search, NTC, and Inclusiveness).

References

Boeckx, Cedric. 2010. A tale of two minimalisms: reflections on the plausibility of crash-proof syntax, and its free-merge alternative. In *Exploring crash-proof grammars,* ed. By Michael T. Putnam, 89-104. John Benjamins, Amsterdam.

Boeckx, Cedric. 2011. Approaching Parameters from Below. In Cedric Boeckx and Anna-Maria Di Sciullo (eds.), *The Biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty*. Oxford: Oxford University Press.

Boef, Eefje. 2013. Partial *wh*-movement revisited: A microcomparative perpsective. Paper presented at Toward a theory of syntactic variation, Bilbao, The Basque Country, June 5-7, 2013.

Brody, Michael 1995, *Lexico-Logical Form: A Radically Minimalist Theory,* Linguistic Inquiry Monograph Twenty-Seven, *MIT Press.*

Carstens, Vicki, 2010. Implications of grammatical gender for the theory of uninterpretable features. In *Exploring Crash Proof Grammars*. M. Putnam (ed), 31-57. Amsterdam: John Benjamins.

Chametzky, Robert 2000, Phrase Structure: From GB to Minimalism, Wiley-

Blackwell.
Chomsky, Noam. 1957. Syntactic Structures. Mouton.
Chomsky, Noam. 1965. Aspects of the Theory of Syntax.
Chomsky, Noam. 1970. Remarks on nominalization. In Roderick Jacobs and
Peter Rosenbaum, eds., Readings in English transformational grammar, 184-221. Waltham, MA: Blaisdell.
Chomsky, Noam. 1981. Lectures in Government and Binding. (Studies in generative grammar 9.) Dordrecht: Foris.
Chomsky, Noam. 1991. "Some Notes on Economy of Derivation and Representation." In Principles and Parameters in Comparative Grammar. Current Studies in Linguistics, Series no. 20, edited by Robert Freidin, 417-54. Cambridge, MA and London: The MIT Press.
Chomsky, Noam. 1993. “A Minimalist Program for Linguistic Theory,” *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, ed. by Kenneth Hale and Samuel Jay Keyser, 1–52, MIT Press, Cambridge, MA. [Reprinted in *The Minimalist Program*, Noam Chomsky, 1995b, 167–217, MIT, Press, Cambridge, MA.]
Chomsky, Noam. 1995a. Bare phrase structure. In *Government and binding theory and the minimalist program: Principles and parameters in syntactic theory,* ed. by Gert Webelhuth, 385-439, Blackwell, Oxford.
Chomsky, Noam. 1995b. *The minimalist program.* MIT Press, Cambridge, MA.
Chomsky, N. 1998. Some observations on economy in generative grammar. In *Is the best good enough?* ed. by P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, and D. Pesetsky, 115-127. Cambridge, Mass.: MIT Press.
Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. by Roger Martin, David Michaels, and Juan Uriagereka, 89-155. Cambridge, Mass.: MIT Press.
Chomsky, Noam. 2007. Approaching UG from below. In *Interfaces + Recursion = Language?* ed. by Uli Sauerland and Hans-Martin Gärtner, 1-29, Mouton de Gruyter, Berlin.
Chomsky, Noam. 2008. On phases. In *Foundational issues in linguistic theory: Essays in Honor of Jean-Roger Vergnaud*, ed. by Robert Freidin, Carlos P. Otero and Maria Luisa Zubizarreta, 133-166, MIT Press, Cambridge, MA.
Chomsky, Noam. 2013. Problems of projection. *Lingua* 130, 33-49.
Chomsky, Noam. 2014. Problems of projection: extensions. Unpublished manuscript, MIT.
Collins, Chris. 2002. Eliminating labels. In *Derivation and Explanation in the Minimalist Program,* ed. by Samuel D. Epstein and T. Daniel Seely, 42-64, Blackwell, Oxford.
Collins, Chris. (this volume). Merge(X,Y) = {X,Y}
Donati, C., 2006. On wh-head movement. In: Cheng, L.L.-S., Corver, N. (Eds.), Wh-Movement: Moving on. MIT, Cambridge, MA, pp. 21--46.
Epstein, Samuel D. 1992. Derivational Constraints on A’-Chain Formation. *Linguistic Inquiry* 23, 235-259.

Epstein, S. D. 1998. Overt Scope Marking and Covert Verb Second. *Linguistic Inquiry* 29:181-229.
Epstein, Samuel D. 2007. On i(internalist)-functional explanation in minimalism. *Linguistic Analysis* 33, 20-53.
Epstein, S. D., E. Groat, R. Kawashima, and H. Kitahara. 1998. *A derivational approach to syntactic relations.* Oxford: Oxford University Press.
Epstein, S. D., Kitahara, H., Obata, M., and Seely, T. D. 2013. Economy of Derivation and Representation. In Den Dikken, M. (ed.), *The Cambridge Handbook of Generative Syntax*. Cambridge: Cambridge University Press.
Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely. 2012b. Structure building that can't be! In *Ways of structure building,* ed. by Myriam Uribe-etxebarria and Vidal Valmala, 253-270, Oxford University Press, Oxford.
Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely. 2014. Labeling by minimal search: Implications for successive cyclic A-movement and the elimination of the postulate "phase". *Linguistic Inquiry* 45, 463-481.
Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely, to appear. From *Aspects* 'daughterless mothers' (aka delta nodes) to POP's 'motherless'-sets (aka non-projection): a selective history of the evolution of Simplest Merge. In Aspects volume.
Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely. 2015. *What do we wonder is not syntactic? In Routledge volume.
Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely. Forthcoming. In McGilvary's CUP volume.
Epstein, Samuel D. and T. Daniel Seely. 2006. *Derivations in minimalism.* Cambridge University Press, Cambridge.
Hornstein, N. and J. Nunes. 2008. Adjunction, Labeling, and Bare Phrase Structure. *Biolinguistics* 2:57-86.
Hornstein & Uriagereka 2002 Reprojections. In Epstein, Samuel David and T. Daniel Seely, eds. Derivation and Explanation in the Minimalist Program, Blackwell Publishing.
Jackendoff, Ray 1977 *X-Bar Syntax: A Study of Phrase Structure*. Cambridge, MA: MIT Press.
Kayne, Richard 1994 The Antisymmetry of Syntax. MIT press.
Kitahara, Hisatsugu. 1997. *Elementary operations and optimal derivations.* MIT Press.
Lasnik, Howard and T. Lohndal (2013) "Brief overview of the history of generative syntax, in den Dikken (ed) *The Cambridge Handbook of Generative Syntax*," Cambridge University Press 2013.
Lasnik, Howard and Mamoru Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15, 235-289.
Lasnik, Howard and Mamoru Saito. 1992. *Move α: conditions on its application and output.* MIT Press, Cambridge, MA.
May, Robert 1985 Logical Form: its structure and derivation. MIT Press.
Muysken, P. 1982 "Parametrizing the notion 'Head,'" J. of Linguistic Research 2, 57-75.
Narita, Hiroki. 2011. Phasing in Full Interpretation. PhD Thesis. Harvard.

Obata, Miki. 2014. Labeled vs. Unlabeled Syntactic Objects: A View from Partial Wh-Movement in German.

Obata, M., M. Baptista, and S. D. Epstein. Forthcoming. Can Crosslinguistically Variant Grammars be Formally Identical?: Third Factor Underspecification and the Possible Elimination of Parameters of UG. Lingua.

Obata, M. and Epstein, S. 2011. Feature-Splitting Internal Merge: Improper Movement, Intervention, and the A/A′ Distinction. *Syntax* 14:122-147.

Ott, Dennis. 2010. Grammaticality, interfaces, and UG. In *Exploring crash-proof grammars,* ed. By Michael T. Putnam, 89-104. John Benjamins, Amsterdam.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar: Handbook of generative syntax,* ed. by Lilliane Haegeman, 281-337. Kluwer, Dordrecht.

Rizzi, Luigi. 2014. Cartography, criteria and labeling. Unpublished manuscript, University of Geneva.

Richards, Marc. 2008. Two Kinds of Variation in a Minimalist System. In *Varieties of Competition.* Fabian Heck, Gereon M¨uller & Jochen Trommer (eds.) Linguistische Arbeits Berichte 87, Universit¨at Leipzig. 133-162.

Sabel, Joachim. 2000. Partial *wh*-movement and the typology of *wh*-questions. In *Wh-Scope Marking,* ed. by Uli Lutz, Gereon Müller, and Arnim von Stechow, 409-446. John Benjamins Publishing Company.

Seely, T. Daniel. 2006. Merge, derivational c-command, and subcategorization in a label-free syntax. In *Minimalist Essays,* ed. by Cedric Boeckx, 182-217, John Benjamins, Amsterdam.

Stowell, Tim 1981, Origins of Phrase Structure, MIT diss.

Takahashi, Daiko. 1993. Movement of wh-phrases in Japanese. *Natural Language and Linguistic Theory* 11, 655-678.

Travis, Lisa. 1984. Parameters and effects of word order variation, MIT diss.

Uriagereka, J. 1999. Multiple spell-out. In S. Epstein and N. Hornstein, eds., *Working Minimalism*, Cambridge, MA: MIT Press.