Á. Gallego & D. Ott (eds.). 2015. *50 Years Later: Reflections on Chomsky's* Aspects. Cambridge, MA: MITWPL.
Available for purchase at http://mitwpl.mit.edu

FROM *ASPECTS*' 'DAUGHTERLESS MOTHERS' (AKA DELTA NODES) TO *POP*'S 'MOTHERLESS SETS' (AKA NON-PROJECTION): A SELECTIVE HISTORY OF THE EVOLUTION OF SIMPLEST MERGE*

SAMUEL D. EPSTEIN
The University of Michigan
HISATSUGU KITAHARA
Keio University
T. DANIEL SEELY
Eastern Michigan University

1 Introduction

This squib briefly and very selectively explores the theory of human phrase structure generation and representation from *Aspects*[1] to "Problems of Projection" (POP). Generally, our goal is to reveal the continuity of Chomsky's always central goal: explanation via simplification. Our primary focus concerns labeling, projection, and the evolution of Simplest Merge in the quest for third factor reduction of Universal Grammar (UG).

Our central more specific point is that Merge, in its simplest form, Merge (X, Y) → {X, Y}, as recently developed in POP, eliminates both the mother node (i.e. projection) of standard Phrase Structure (PS) rules and representations, as well as the empty symbol delta (i.e. Δ) postulated in *Aspects* (and in "A Minimalist Program for Linguistic Theory" (MPLT) as internal to Generalized Transformation (GT), see below). As a result, the POP system attains far greater depth of explanation, while, as Chomsky notes, concomitantly minimizing UG, thereby reducing the set of currently (and perhaps perennially) scientifically unanswerable questions regarding the evolution of UG.[2]

* For helpful discussion of certain of the ideas presented here, we thank Noam Chomsky and Chris Collins.
[1] We use the abbreviation *Aspects* to refer to Chomsky 1965.
[2] See Hauser et al. 2014.

2 A Brief Review of Certain "Aspects" of *Aspects*

Standard theory recursive PS rules of the sort postulated in *Aspects* (and earlier) provided a revolutionary solution to the long-standing paradox of discrete infinity: while the human brain is finite, the generative capacity of an I-language is infinite. Taking this as the core 'creative' property of human knowledge of syntax, it was Chomsky's postulation of recursive PS rules that formally solved what had previously seemed paradoxical, namely "infinite use of finite means" (Humboldt's term, as Chomsky notes). More generally Chomsky resurrected abandoned 17th century ideas regarding the mind, and centrally contributed to the modern day birth of the cognitive sciences, by noting the legitimacy of the postulation of abstract (non-material) concepts in normal (including of course, even physical) modern science, hence its permissibility, in fact necessity, in (previously, and still by many condemned) mentalistic theories of the brain.

Recursive PS rules provided an explicit representation of knowledge of linguistic structure – "the basic principle" in current terms – and accounted for the "creative aspect of language use." A recursive structure-building mechanism is necessary for any adequate theory of I-language. But, given the commitment to explanation, important questions emerged regarding the phrase structure component of *Aspects*, some of which we are fully appreciating only now, some 50 years later. There are two questions we focus on here, involving the nature of the mother node (i.e. projection) and the nature of the empty symbol delta.[3]

One central explanatory question that arises with *Aspect's* conception of phrase structure and PS rules more generally is: "Why do we find that humans develop these particular (constructions-specific or category-specific) rules, and not any of an infinite number of other PS rules, or other types of rules?" Consider a PS rule like (1).

(1) VP → V NP

Why exactly is there a VP on the left side of the arrow? Why more specifically is the mother labeled VP (and not, say, NP, or some other category or, for that matter, some non-category)? Is the label of the higher category stipulated, or does it follow from some deeper principle of generative systems, or perhaps from an even more general principle not specific to the human language faculty? We might even ask, Why is there a label at all? Indeed Collins (2002) proposed the elimination of labels, and Seely (2006) sought to deduce labelessness from the derivational theory of relations (Epstein et al. 1998, Epstein and Seely 2006). Since labels are not merged, they are in no relation, hence are equivalently absent from PS representations.

In *Aspects*, PS rules were essentially unconstrained, anything could be rewritten as anything and thus the existence and categorial status of mother labels were stipulated. Thus, for example, there was a "headless" rule like (2).[4]

(2) S → NP VP

[3] Our primary concern is the cases where delta undergoes transformational substitution, as opposed to lexical insertion (see *Aspects* and below).

[4] Note that there is an "unprojecting headed" rule like (i) in *Lectures on Government and Binding (LGB)*.

(i) S→NP INFL VP

In this case too the mother is categorially unrelated to its daughters.

In (2), the mother node S is not a projection of (i.e. it is categorially unrelated to) its daughters. So, why do we find such headless phrases as S, while the major lexical categories seem to have heads, e.g., V in VP, and N in NP?

Another issue that arose involves a particular innovation of *Aspects*, namely, the postulation of the empty symbol delta Δ:[5]

> "[S]uppose that (for uniformity of specification of transformational rules) we add the convention that in the categorical component, there is a rule $A \rightarrow \Delta$ 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."[6] (p. 122)

Thus, delta would appear in the (simplified) deep phrase marker associated with passive,[7]

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

Raising of the object NP would involve the substitution of that object NP for the delta subject NP, yielding:

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

In effect, the delta is an empty and non-branching maximal projection; in this case an NP, and it has a purely formal status.

Note that both the mother (the S in (3)) and delta nodes of *Aspects* involve, at least in one sense, 'look ahead;' they are telic, anticipating transformation that *will* take place. And this in turn results, in part, from the 'top down' conception of phrase structure and from the particular model of grammar assumed in *Aspects* (in which (3) is postulated since S → NP VP and in which semantic interpretation is at the Deep Structure (DS) level (see Katz & Postal 1964)).

Aspects appealed to 'top down' PS rules of the sort in (1) and (2). But as pointed out by Chomsky (1995b), attributing the insight to Jan Koster,[8] PS rules of this sort, where the mother node is projected, involve, at least in one sense, 'look ahead.' The mother node, the label VP of, say, VP → V NP, is telic in the sense that it indicates the categories generated by the syntax that

---

[5] According to Harris (1993), the delta node was developed from an earlier idea of Katz and Postal (1964).

[6] It should be noted that the delta node played an important role in another crucial development in *Aspects*; namely, the separation of the lexicon from the syntactic component.

[7] 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.

[8] Chomsky (1995b: 318) 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." See below for further comment on the shift from *Aspects* to early minimalism.

will be relevant to the interpretive components, PF and LF. Put another way, the DS in (3) in fact encodes the categorial structure of what the Surface Structure (SS) *will* be. That is, if (structure preserving) delta substitution is required, then the NP subject of S is already present at DS, 'awaiting' the obligatory arrival of *the man*. This encoding of SS in 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 2000, Epstein et al. 1998, Uriagereka 1999).

Since in *Aspects* the initial level of representation, namely DS, fed into the semantic component, and since objects (presumably) required a label for semantic interpretation, it was necessary that labels be encoded as soon as possible, and hence encoded in the structures that served as input to semantics. Furthermore, since SS fed the Phonological Component, and since PF also (presumably) required labels, then the mother nodes needed to be present also at the last stage of a derivation.

Projections were needed, then, at both the initial (DS) and final (SS) levels, and were represented throughout the derivation. PS rules, projecting labels, – including NP delta projected from no lexical head N[9], and the S-mother of NP delta, which is not a projection of NP delta, as in (3) above – served precisely the job of providing "sentential level" projection from the outset of the derivation (the theory being a theory of sentence-structure).

The delta is similarly 'anticipatory.' Recall a case like (3), in which a subject NP is already present *structurally* even before it is filled *lexically*, and this subject NP, since the mother S is already present, has no option of itself projecting; the mother S is already predetermined.

So, recursive PS rules of *Aspects* provided an empirically motivated, profound answer to a paradox solving the problem of discrete infinity. But, the nature of projection and of the empty symbol delta employed in *Aspects* raised a number of important questions, involving: projection, the nature of the lexicon, the relation between the lexicon and syntax, delta as a lexical item, the intertwining of levels, 'look ahead' the inviolability of S → NP VP, and semantic interpretation at DS.

3 X-bar theory: the elimination of PS rules

X-bar theory represented a major development in the history of phrase structure.[10] Rather than unconstrained, stipulated, and ultimately non-explanatory PS rules, the X-bar template imposed tight restrictions on what counts as 'humanly possible phrase structure representation.' X-bar theory sought to eliminate PS rules, leaving only the general X-bar-format as part of UG.

X-bar theory pushed endocentricity to its logical conclusion: since some categories seemed to be endocentric, like for example the lexical categories mentioned above (VP, NP, etc.), then, to eliminate unexplained asymmetries, it was assumed that all categories, including functional categories, are endocentric.[11] Thus, a PS rule like (2), is excluded; it must be reduced to the X-

[9] In *Aspects*, a rule $A \rightarrow \Delta$, is proposed for each lexical category. Given this proposal, it follows that prior to lexical insertion, every lexical category, e.g. N, V, etc., is a delta. Thus, phrases like, say, NP are still headed, by N, in the sense that NP immediately dominates a delta N.

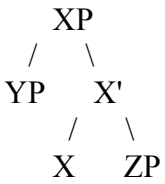
[10] See Chomsky 1970, Jackendoff 1977, Stowell 1981, among others; see also Chametzky 2000, Chomsky 1995a, Chomsky 1995b, and Lasnik and Londahl 2013 for discussion.

[11] Of course, the asymmetry could have been resolved by alternatively assuming no category is endocentric. In current work the asymmetry is simply a consequence of recursive simplest Merge – with simplicity of the generative

bar template and thus must be 'headed.'

Another major development within X-bar theory was that linear order was removed from the PS component; X-bar projections represented structure but not the linear order of elements within the structure. 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 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. Thus, the theory came to express that "Syntax is not word order!" since word order is phonological. This disentangling of "dominance" (in hindsight, a misleading misnomer) and precedence, along with explaining their existence as subservient to the interfaces (dominance for semantics, precedence for phonology) was a profound step in the development of the standard strong minimalist thesis.[12]

In X-bar theory, the mother is predetermined. Assuming binary branching, if X is non-maximal, its mother will be the category of X. If X is maximal, its mother will be the category of X's sister. Consider the following tree representation (ignoring order)

(5)


In (5), X and X' are non-maximal[13] and hence themselves project. YP is maximal and hence its mother is the category of YP's sister (in this case X). The same holds for ZP. 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).

What about the delta introduced in *Aspects*? It too implicitly remains in the X-bar schema. Under X-bar theory, the landing site of movement is often called SPEC, but SPEC is in effect a cover-term for delta as well. So, we could say delta was still assumed for movement under X-bar theory, i.e. X-bar was a constraint on transformationally derived structures in which projection is determined by X-bar schemata.[14] So, the moving category has no chance to project — the mother of the mover 'landing in' SPEC is by definition not the mover.

X-bar theory thus raised a new set of questions – again, following the Chomskyan theme of a continued quest for yet deeper explanation. Namely, why should there be projection, and why should it conform to the X-bar template?

---

procedure, not uniformity of its representational output, being the or at least an explanatory criterion (see below for further comment).

[12] It should be noted that the removal 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; it is part of externalization. And we have an evolutionary hypothesis, consistent with the primacy of CI, i.e. language is primarily 'for thought' and not 'for communication;' see Chomsky (2013).

[13] See Muysken 1982 on relational definitions of maximal and minimal categories.

[14] 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 delta).

4 A Minimalist Program for Linguistic Theory: the Initial Transition from X-bar to Merge

Early minimalism brought major shifts in the architecture of the computational system for human language. In MPLT, Chomsky (1993) eliminates DS, and explicitly re-introduced delta into the theory, but it appears only internal to GT – not as a symbol appearing in/at a level of representation, as *Aspects* had delta in DS.[15]

So, GT adds delta to a targeted category, a merging category substitutes delta, and the resulting structure is required to be consistent with X-bar schemata by definition. Thus:

> "GT is a substitution operation. It targets K and substitutes K^1 for Ø [delta] in K. But Ø is not drawn from the lexicon; therefore, it must have been inserted by GT itself. GT, then, targets K, adds Ø, and substitutes K^1 for Ø, forming K*, which must satisfy X-bar theory … we never see Ø; it is subliminal, like the 'first half' of the raising of the NP to subject position." (MPLT, p. 22)

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 the more recent Internal Merge, where one of the objects being joined together is initially contained within the other. In effect, the elegantly constrained X-bar theory, together with its stipulated (or axiomatic) properties (endocentricity, head-complement, and spec-head relations) was taken to be a UG filter on transformational output representations.

5 Emergence: Bare Phrase Structure[16]

X-bar theory represented a major step in the continued quest for explanation. But X-bar theory was not exempt from explanatory scrutiny. The question emerged: Why should X-bar theory hold? Why do we find these particular relations (endocentricity, head-to-complement, and spec-head), as opposed to an infinite number of alternative phrase structure systems? Stated in another way, and adhering to Minimalist method (see Chomsky 2007), we can ask: how "should" phrase structures be generated under minimalist assumptions? In *Bare Phrase Structure*, Chomsky's (1995a: 396) answer was:

> "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

[15] It should be noted, however, that there is an interesting similarity between *Aspects* and MPLT with respect to the delta node: each application of Generalized Transformation taking X and Y, and forming Z may be regarded as a transformational mapping, and an empty dummy element delta exists only during this mapping.

[16] The first part of the exposition of this subsection closely follows that of Epstein, Kitahara, Seely (2013), see also Epstein, Kitahara, Seely (to appear).

> 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."

Merge was introduced as the central structure building operation of the narrow syntax,[17] 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, Chomsky (1995a, b) assumed the set {α, β} was too simple; 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 γ. Interestingly, note that in the above passage the argument for labels mentions only their necessity at the interfaces, and does not mention any reason for requiring them in the NS.[18] We return to the status of labels in NS, momentarily.

So, the output of Merge is a labeled set γ={δ, {α, β}}.[19] In *BPS,* Chomsky (1995a:397-398) asked what exactly the label of γ is:

> "If constituents α, β of γ have been formed in the course of computation, one of the two must project, say α. At the LF interface, γ (if maximal) is interpreted as a phrase of the type α (e.g. a nominal phrase if its head κ is nominal), and it behaves in the same manner in the course of computation. It is natural, then, to take the label of γ to be not α itself but rather κ, the head of the constituent that projects, a decision that also leads to technical simplification. Assuming so, we take γ = {κ, {α, β}}, where κ is the head of α and its label as well."

Under this definition, the label of γ is the head of one of its immediate constituents; i.e. one of its members, assuming two-membered set theory. If α projects, then the object γ constructed from α and β by Merge is {H(α), {α, β}}, where H(α) is the head of α (see also Chomsky 1995b). Additionally, the notion "term" (essentially the definition of "syntactic constituent") is defined as follows: (i) K is a term of K; and (ii) if L is a term of K, then the members of the members of L

---

[17] Ideally, there was a *single* operation but at the time there was appeal to both External Merge (X, Y are separate objects) and Internal Merge (Y contained in X). As we will see in some detail below, Chomsky (1995a,b) recognized that the simplest formulation of Merge is Merge (X, Y) -> {X, Y}, but he took it to be *too* simple, and proposed Merge (X, Y) -> {Z, {X, Y}}. Merge (External Merge) and Move (Internal Merge) are essentially the same. The crucial differences between Merge and Move are: Merge is a binary operation, while Move is singulary operation, and the latter singulary operation is morphologically driven, and all sorts of conditions were assumed to constrain how it is morphologically driven, as we'll see.

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

[19] For Chomsky (1995a,b), if (i) Merge is the only structure-building operation forming SOs, and (ii) each SO must be of the form γ={δ, {α, β}} for the interface systems, then (iii) it must be Merge that creates a labeled set for the interface systems. We don't have anything other than Merge in NS.

are terms of K (Chomsky 1995a:399).

Chomsky (1995a,b) did not discuss exactly how Merge operates to form such labeled sets, but he assumes that either α or β may project (in principle), but if the wrong choice is made, deviance would result.[20]

Notice, Chomsky (1995a,b) eliminated both delta and X-bar theory, but projection is still present; projection applies by definition. Merge was defined as Merge (X, Y) → {Z, {X, Y}}, where Z is either the head of X or the head of Y. Under this definition, it was guaranteed that the label (= projected node) is either the head of X or the head of Y, again by definition.

To sum up, (i) under phrase structure grammar with delta-substitution, a moving category has no chance to project by definition, (ii) under X-bar theory with SPEC/delta substitution, a moving category has no chance to project, again by definition, (iii) under GT with delta + X-bar theory (Chomsky 1993), a moving category has no chance to project by definition, (iv) under Merge (X, Y) → {Z, {X, Y}}, where Z is either the head of X or the head of Y, either a hosting category or a moving category projects.[21]

6 Simplest Merge: the Elimination of Labels and Projection from the Theory of Mental Representation

The strong minimalist thesis (SMT), presented by Chomsky (1993, 1995b) 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.[22] This is of course not an "assertion" but a hypothesis deemed worthy of exploration on a number of methodological grounds common to normal science.[23]

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}.

(6) Merge (X, Y) → {X, Y}

To the best of our knowledge, Collins (2002) was the first within the generative 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 the point of departure, Seely (2006) reanalyzes the matter derivationally, arguing that simplest merge (i.e. Merge (x, y) = {x, y}) is motivated on minimalist grounds alone and simplest merge entails the elimination of

[20] We're assuming that it's deviance at CI, e.g. "chased the dog" labeled and hence interpreted as NP, rather than VP.

[21] 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).

[22] See Epstein (2007) for detailed discussion of the idea that the theory is an "I(nternalist)-functional" one in the sense that the rules apply only "in order to" satisfy the interfaces. With the framework of Chomsky (2007, 2008, 2013, 2014) the idea is that operations freely apply as long as they conform to the laws of nature. There may be many functions, but there is no "the function." Convergent derivations have the property of yielding interface interpretable representations that have been derived in a computationally efficient (meaning "laws-of-nature" compliant) way. In short, Merge freely applies as long as the applications conform to the laws of nature.

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

syntactically represented labels and the elimination of (any type of) projection:

> "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: 193)

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

For Chomsky (2013), Merge, defined in the simplest form, also applies freely as long as it conforms to third factor principles (as it must) such as the proposed inclusiveness condition, "no new objects are added in the course of computation apart from arrangements of lexical properties" (Chomsky 1995b), and the no-tampering condition (NTC), "Merge of X and Y leaves the two SOs unchanged" (Chomsky 2008).[24]

[24] Given considerations of space, we can't comprehensively consider a host of interesting questions regarding the (re-)development of 'free' application of rules, and just what 'free' means, and the extent to which there was a shift between constrained rules vs. free application (see Boeckx (2010) and Ott (2010) for important discussion). Given the SMT, in early minimalism there was computationally efficient satisfaction of the interfaces. The activity condition, Greed, Last Resort, and Criterial Freezing – (see Epstein 1992 in early MIN = Der constraints on Abar chain formation: once a wh phrase is in spec of +wh C, it is frozen – moving it would "undo' satisfaction of a filter seemed to be constraints on the application of rules. Free Merge seemed to represent a partial return to the 'free' application of move in GB. But, on one view, SMT itself hasn't changed at all, in the following sense. In early minimalism, we eliminated all syntactic filters. Only naturalized interface filters, called bare output conditions, survived. As for syntactic constraints, we hope that they all reduce to principles of economy, now understood as third factor principles. 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 POP+, 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 POP+ adopts simplest Merge; the clauses beginning with "only if..." are all gone. In other words, there is no constraint on Merge (other than the laws of nature), meaning it is an operation that can apply freely. If so, the shift from early minimalism to POP+ is the success of reducing the additional defining properties of operations to third factor principles (and their interaction with Merge). When we had complex operations like Move, the system seemed deterministic, because the applications of Move are well constrained, but that was an illusion. It was always free to apply rules, but those rules themselves had necessary conditions as their defining properties, so the domain of free applications were limited. Say, if Merge is defined to operate on C and D only. Then, even if Merge is free, by definition, Merge cannot operate on M and D. Merge is defined to operate on C and D, but not on M and D, period. But this is part of the definition of Merge. So, it is still true Merge freely applies, but it cannot do anything beyond what it is designed to do. To sum up, when Merge/Move were defined with

In the absence of any stipulated constraints, there are just two applications of Merge. Suppose X is merged to Y (introducing the asymmetry only for expository purposes). Then, either X originates external to Y, call it External Merge; or X originates internal to Y, call it Internal Merge. Given NTC, Internal Merge yields two copies of X: one external to Y and the other within Y. There is no need to stipulate a rule of formation of copies (or remerge), and Chomsky"s (1993) copy theory of movement follows from "just IM applying in the optimal way, satisfying NTC" (Chomsky 2007). It would require stipulation to bar either type of application of Merge. Hence Chomsky explains the existence of what had been thought to be a very curious property, namely displacement and shows that it follows from simplest unifying assumptions regarding the computational procedure (despite creating well-known parsing and communicative problems).

Notice, Merge does not encode a label.[25] There is no labeled categorial node above X and Y present in the proposed NS mental representation. The categorial status of the set {X, Y} is representationally unidentified.[26] That is, projection (or labeling) is no longer stipulated in the system as a defining property of Merge. Recall that standard PS rules simultaneously created dominance and precedence and so could not be primitive (as discussed above), likewise we could say that Merge with labels defining both mothers and sisters is not a possible primitive operation (though, in standard PS of course, sisters cannot be defined independent of mothers, nor can mother be defined independent of sister) Eliminating linear order, we create a set. This set is identified by minimal search, and at CI it then satisfies Full Interpretation (a bare output condition); a "higher node" or the "label" of an object formed by Merge is not represented in NS but is identified/computed by application of third factor minimal search — an (ideal) expectation under SMT.

Chomsky (2013) takes labeling to be the process of finding the relevant information of {X, Y} generated by Merge. He proposes that such labeling is "just minimal search, presumably appropriating a third factor principle, as in Agree and other operations." So, labeling is not syntactically represented. No new categories are created in the course of a derivation (which, in fact, reduces to Inclusiveness). 'Labeling' is simply the name given to the independently motivated minimal search procedure, itself 3rd factor and hence not stipulated. POP then eliminates labels and projection, replacing it with a labeling algorithm that is an instance of the general principle of Minimal Computation, hence gaining yet greater depth of explanation.[27]

To understand how the labeling analysis, outlined in POP, works, let us examine the following two cases. Suppose SO = {H, XP}, H a head and XP not a head. Then minimal search will select H as the label, and the usual procedures of interpretation at the interface can proceed. By contrast, suppose SO = {XP, YP}, neither a head (recall PS rule (2) above). Here minimal

---

necessary conditions, we found Merge/Move rigorous/deterministic, but that was an illusion. The system has been equipped with free applications of Merge. As stated in PoP+ "Another reason has been the lingering idea, carried over from earlier work, that each operation has to be motivated by satisfying some demand. But there is no reason to retain this condition. Operations can be free, with the outcome evaluated at the phase level for transfer and interpretation at the interfaces."

[25] Note, if Merge does not encode a label, then label-identification cannot be a prerequisite for merge application within NS, as discussed in POP (see in particular p. 43, footnote 30).

[26] See Collins (2002) and Seely (2006), among others, for earlier views of label-less syntax.

[27] Note that eliminating labels from syntactic representations in fact comports with the standard view, in the following (counterintuitive) sense: (assuming binarity) we take an X and a Y and form one new object as a result of combining them. In current terms, this new object is the set containing X and Y and nothing else. Thus there is no 'room' for a *fourth* newly created object (i.e. a label) being created by combinatorial operations, again, assuming, as is standard that combinatorial operations create one and only one new object not present in the structural description.

search is ambiguous, finding both the head X of XP and the head Y of YP. This ambiguity is intolerable; left as is, labeling fails and Full Interpretation (FI) is violated at the interface levels. To resolve this so-called XP-YP problem for labeling, Chomsky (2013) suggests the following 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]

7 Conclusions: from *Aspects* to POP … and Beyond

The labeling analysis proposed in POP has many consequences (beyond those traced in POP itself), some of which are just beginning to be explored. Epstein et al. (2014; henceforth EKS), for example, argue that POP's account of obligatory exit from intermediate positions in A'-movement, as in

(7) *you think [$_{CP}$ who [Jon likes] (*cf.* who do you think Jon likes)

carries over to A-movement as well. Thus, (7) is problematic at CI since the embedded CP will not be labeled. Similarly, EKS argue, for A-movement:

(8) *seems [$_{TP}$ a man [to like Jon]] (*cf.* a man seems to like John)

where the embedded TP fails to be labeled by CI. A host of technical mechanisms that have been proposed over the years to account for cases such as (8) are eliminated in favor of the single, simple labeling analysis. Chomsky (2014) extends his labeling-as-minimal-search analysis to the unification of the EPP and the ECP, reducing both to independently motivated properties of (3rd factor) labeling. Chomsky (2014) also suggests an analysis of 'obligatory halting' in a Criterial Freezing position (see also Epstein, Kitahara and Seely. to appear), reducing it too to labeling.

The 50 years since *Aspects* have brought many technical changes, but also ever deeper insight into the human language faculty. What in *Aspects* was two separate subsystems: (i) construction specific, category specific, language(s) specific and UG specific PS rules, accounting for word order and structure, and (ii) Transformations, accounting for 'displacement,' has now been reduced to a *single*, (maximally) *simple*, and necessary operation, namely, Merge (X, Y) = {X, Y}. This represents extraordinary progress in that the entire PS and Transformational components have been reduced to a single simple, linguistically unconstrained, operation, rendering explanation of the radically simplified theory of UG, conceivable and perhaps tractable so that the evolutionary 'great leap forward' of humans roughly 75 thousand

[28] Note that in (B), the, or a, featural intersection of X and Y is taken as the label of {XP, YP}. Interestingly, as pointed out to us by an anonymous reviewer of EKS (to appear), Chomsky (1995b) proposed this very idea, but at the time rejected the existence of projection of an intersect of features borne by X and Y, since X, Y might not have intersecting features. However, EKS (to appear) argue that the central phase-motivating (i)

(i) *there is likely [[$_{XP}$ a man] [$_{YP}$ to arrive]]

is now deducibly disallowed precisely because there is no intersection of features between XP, YP and hence no label for (what would be) the embedded TP (with gibberish resulting at CI).

years ago, might one day be explained. And what were syntax-internal (and ultimately stipulative) constraints on rules or (non-explanatory) level ordered filters, are now by hypothesis being reduced to 3rd factor laws of nature. There have been many important discoveries since *Aspects*, and the current research program seems more intriguing and exciting than ever, particularly for those who are interested in scientific (theory-based) explanation and are willing to be surprised by that which seems so obvious, but demands (demanding) explanation.

References

Boecks, Cedric. 2006. 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, 80-104. Amsterdam: John Benjamins.

Brody, Michael. 1995. *Lexico-logical Form: A Radically Minimalist Theory*. Cambridge, MA: MIT Press.

Chametzky, Robert. 2000. *Phrase Structure: From GB to Minimalism*. Oxford: Blackwell.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick Jacobs and Peter Rosenbaum (eds.), *Readings in English Transformational Grammar*. Waltham, MA: Ginn & Co., 184–221.

Chomsky, Noam. 1993. A Minimalist program for linguistic theory. In Ken Hale and Samuel Jay Keyser (eds.), *The View from Building 20.* Cambridge MA: MIT Press, 1-52.

Chomsky, Noam. 1995a. Bare phrase structure. In G. Webelhuth (ed.), *Government and Binding Theory and the Minimalist Program*. Oxford: Blackwell, 285-439.

Chomsky, Noam. 1995b. *The Minimalist Program.* Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels and Juan Uriagereka (eds.), *Step by Step.* Cambridge MA: MIT Press, 89-155.

Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland and Hans Gärtner (eds.), *Interfaces + Recursion = Language?* New York: Mouton de Gruyter, 1-29.

Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos Otero and Maria-Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory.* Cambridge MA: MIT Press, 136-166.

Chomsky, Noam. 2013. Problems of projection. *Lingua* 130, 33-49.

Chomsky, Noam. 2014. Problems of projection: Extensions. Manuscript, MIT.

Collins, Chris. 2002. Eliminating labels. In Epstein and Seely (2002).

Donati, Caterina. 2006. On *wh*-head movement. In Lisa Lai-Shen Cheng and Norbert Corver (eds.), *WH-Movement: Moving On*. Cambridge, MA: MIT Press, 21-46.

Epstein, Samuel D. 1998. Overt scope marking and covert verb-second. *Linguistic Inquiry* 29, 181-228.

Epstein, Samuel D. 2007. On I(nternalist)-functional explanation in Minimalism. *Linguistic Analysis* 33, 20-53.

Epstein, Samuel. D., Erich Groat, Ruriko Kawashima, and Hisatsugu Kitahara. 1998. *A Derivational Approach to Syntactic Relations.* Oxford: Oxford University Press.

Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely. 2012. Structure building that can't be! In Myriam Uribe-Etxebarria and Vidal Valmala (eds.), *Ways of Structure Building*. Oxford: Oxford University Press, 253-270.

Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely. 2013. "Simplest Merge generates set intersection: Implications for complementizer 'trace' explanation." In The Proceedings of *GLOW in Asia IX*, 77-92. Mie University, Japan.

Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely 2014 "Labeling by minimal search – implications for successive cyclic A-movement," *Linguistic Inquiry* 45.3, 463-481.

Epstein, Samuel D., Hisatsugu Kitahara, and T. Daniel Seely. to appear. "*What do we wonder is not syntactic?" In Samuel D. Epstein, Hisatsugu Kitahara, and T. Daniel Seely, *Explorations in Maximizing Syntactic Minimization.* London: Routledge.

Epstein, Samuel D. and T. Daniel Seely (eds). 2002. *Derivation and Explanation in the Minimalist Program*. Oxford: Blackwell.

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

Harris, Randy Allen. 1993. *The Linguistics Wars*. Oxford: Oxford University Press.

Hauser, Marc, Charles Yang, Robert Berwick, Ian Tattersall, Michael J. Ryan, Jeffrey Watumull, Noam Chomsky and Richard C. Lewontin (2014) The mystery of language evolution. *Frontiers in Psychology*. May 07. doi: 10.3389/fpsyg.2014.00401

Hornstein and Uriagereka. 2002. Reprojection. In Epstein and Seely (2002).

Jackendoff, Ray. 1977. *X-bar Syntax: A Study of Phrase Structure*. Cambridge, MA: MIT Press.

Katz, Jerrold and Paul Postal. 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, MA: MIT Press.

Lasnik, Howard and Terje Lohndal. 2013. Brief overview of the history of generative syntax. In Marcel den Dikken (ed.), *The Cambridge Handbook of Generative Syntax*. Cambridge: Cambridge University Press. 26-60.

Muysken, Pieter. 1982. Parametrizing the notion 'head'. *Journal of Linguistic Research* 2, 57-75.

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

Stowell, Tim. 1981. Origins of Phrase Structure. Ph.D. dissertation, MIT.

Travis, Lisa. 1984. Parameters and Effects of Word-order Variation. Ph.D. dissertation, MIT.

Uriagereka, Juan. 1999. Multiple Spell-out, in S. D. Epstein and N. Horstein (eds), *Working Minimalism*, MIT Press. 251-282.