Some Puzzling Foundational Issues: The Reading Program*

Noam Chomsky
University of Arizona & M.I.T.
chomsky@mit.edu

Received: May 23, 2019
Accepted: September 23, 2019

Abstract

This is an annotated transcription of Noam Chomsky's keynote presentation at the University of Reading, in May 2017. Here, Chomsky reviews some foundational aspects of the theory of structure building: essentially, Merge and Label. The aim is to eliminate what he refers to as *extensions* of Merge which are seemingly incompatible with the Strong Minimalist Thesis while still accounting for recursive structure, displacement, and reconstruction (as the main empirical goals of the Minimalist Program). These include sidewards movement, multi-dominance, and late-Merge; all of which have been developed throughout the life cycle of transformational generative grammar. Furthermore, Chomsky formulates a series of conditions that an adequate formulation of Merge must meet, and sketches how the aforementioned *extensions* may violate these conditions. Chomsky arrives at a formulation of an operation MERGE, which maintains the core properties of Merge but is further restricted by limitations over what MERGE can do to the *workspaces* where syntactic operations apply.

Keywords: Strong Minimalist Thesis; workspaces; MERGE; recursion

Resum. *Alguns problemes fonamentals desconcertants: el programa de Reading*

Aquesta és una transcripció anotada de la presentació principal de Noam Chomsky a la Universitat de Reading, el maig de 2017. Aquí Chomsky revisa alguns aspectes fundacionals de la teoria de la construcció d'estructures: fonamentalment, fusió i etiquetatge. L'objectiu és eliminar allò que aquest autor anomena extensions de fusió, que aparentment són incompatibles amb la tesi minimalista forta, i continuar donant compte de l'estructura recursiva, el desplaçament i la reconstrucció (com a principals objectius empírics del programa minimalista). Aquests inclouen el moviment lateral, la multidominància i la fusió tardana, que s'han desenvolupat al llarg del cicle de vida de la gramàtica generativa transformacional. A més a més, Chomsky formula una sèrie de condicions que ha de complir una formulació adequada de fusió i indica que les extensions esmentades poden violar aquestes condicions. Chomsky arriba a la formulació d'una operació fusió que manté les

* This talk took place at the University of Reading, in May 11th, 2017 as part of an international Colloquium: *Generative Grammar in the 21st century: the evidence and the rhetoric*. The video recording of the talk is available at <https://www.facebook.com/theuniversityofreading/videos/1115026661937460/>. The talk was transcribed and annotated by Diego Gabriel Krivochen and Douglas Saddy, who take responsibility for any mistakes.

ISSN 1695-6885 (in press); 2014-9718 (online) https://doi.org/10.5565/rev/catjl.287

propietats bàsiques de fusió però que queda restringida per limitacions sobre el que la fusió pot fer en els espais de treball on s'apliquen les operacions sintàctiques.

Paraules clau: tesi minimalista forta; espais de treball; fusió; recursió

When I hear about the work of the 1950s I always try to put myself in the position of a 17 year old kid in my first linguistics class in the 1940s. I imagine how it would have felt if someone got up and started talking about a book he wrote in 1885. Why bother? But that's what this [talk] is about.

I'd like to discuss some foundational issues which are unsettled and I think are rather troublesome, and that bear directly on a number of very important issues in current work and I think raise questions about the legitimacy of problems and challenges we have faced. So just a brief comment on background assumptions. The basic question we face is 'what is the Language Faculty?', 'what is Universal Grammar (UG)?'; there is good reason to seek the simplest possible answers to this.

One reason is just general methodology: simplicity is approximately the same as explanatory depth, we're looking for explanation – topics that were explored extensively in the 1950s by Nelson Goodman.[1] Another interesting comment that bears on our own history, I think, is by Richard Feynman, the Nobel laureate. When he received the Nobel Prize, he reviewed a number of developments in Physics and pointed out that, in every case the interesting results were reached from several different points of view and suggested that the characterization of explanatory depth – simplicity – is reaching the same conclusion by independent paths.[2] He pointed out that although these approaches turned out to be physically equivalent, they were psychologically different in that the different approaches suggested different ways to approach the innumerable unanswered questions that arise whenever a result is reached, and it opens up new puzzles. I think we find things like that too [in Linguistics]. So I think that's one reason for seeking the simplest solution.

The second reason is a dictum of Galileo's, which has served sciences pretty well for 500 years – therefore, worth taking seriously –, namely, that Nature is

1. See, for example, Goodman (1968: Chapter 4), which summarizes much of his previous work on the syntax of symbolic systems.
2. The relevant fragment is the following (Feynman, Nobel lecture, Dec. 11th, 1965)

   It always seems odd to me that the fundamental laws of physics, when discovered, can appear in so many different forms that are not apparently identical at first, but, with a little mathematical fiddling you can show the relationship. An example of that is the Schrödinger equation and the Heisenberg formulation of quantum mechanics. I don't know why this is - it remains a mystery, but it was something I learned from experience. There is always another way to say the same thing that doesn't look at all like the way you said it before. I don't know what the reason for this is. I think it is somehow a representation of the simplicity of nature. A thing like the inverse square law is just right to be represented by the solution of Poisson's equation, which, therefore, is a very different way to say the same thing that doesn't look at all like the way you said it before. I don't know what it means, that nature chooses these curious forms, but maybe that is a way of defining simplicity. Perhaps a thing is simple if you can describe it fully in several different ways without immediately knowing that you are describing the same thing.

in fact simple and it's the task of the scientists to show how that's the case,[3] be it the motion of planets, the tides, the flight of birds or, in our case, the nature of Language.

And there's a third reason, which is specific to the study of Language and fairly recent in our understanding of its import, and that has to do with human evolution and the evolution of language (by which I mean the evolution of the Language Faculty). We don't know a lot about it, but we know something, and there are recent discoveries that strongly suggest that language emerged along with anatomically modern humans or maybe slightly later and has remained stable ever since.[4] Certainly, it arose before the separation of humans, which by a number of genome analyses has been shown to be not long after humans emerged, roughly 200.000 years ago. There's a very interesting paper on this by Riny Huijbregts.[5] Well, if that's true, and it seems to be, then whatever emerged just has to be very simple: there were no selectional factors involved, there was just resort to some natural principle. Since it [Language] emerged suddenly and never changed, then it has to be a simple object, and that's what we should be looking for.

As a side comment, I think there are by now some reasons (I am not going to review them) to suppose that the core I-language (internal language) generates solely representations on one interface: C-I (Conceptual Intentional interface), essentially a kind of language of thought.[6] And that's probably close to, or probably we will discover totally invariant among human beings. It seems that the complexity, the variety of language arise overwhelmingly if not completely from the ancillary operations which lead to externalization which we know draws upon our sensory motor system. And it's pretty natural that that should be complex and vary because you have to match two systems that essentially have nothing to do with one another. The internal system seems to have arisen pretty suddenly along with modern humans and the SM (Sensory-Motor) system have been around for hundreds of thousands, in some cases millions of years, and have absolutely nothing to do with language. So when we try to connect these two things, it's necessarily going to be a complex operation, and in fact the external operations, although they certainly follow principles and rules of a restricted variety they nevertheless violate just about any principle of computational complexity one can imagine, and they do vary a lot, change a lot, generation to generation and so on. So I'll just assume that, admittedly without

3. This idea can be found, for instance, in *Dialogo sopra i due massimi sistemi del mondo* (1632).
4. See Berwick & Chomsky (2016) for discussion.
5. See, for instance, Huijbregts (2017).
6. See Fodor (1975, 2008). It may be worth pointing out that:

> *the language-of-thought hypothesis endorsed in LOT 1 wasn't just any old hyper-realism about the mental; it was, in particular, a species of RTM (that is, of the representational theory of mind). Roughly, but near enough for present purposes, RTM is a claim about the metaphysics of cognitive mental states and processes: Tokens of cognitive mental states are tokens of relations between creatures and their mental representations. Tokens of mental processes are 'computations'; that is, causal chains of (typically inferential) operations on mental representations. There is no tokening of a (cognitive) mental state or process (by a creature, at a time) unless there is a corresponding tokening of a mental representation (by that creature, at that time)*. (Fodor 2008: 5-6)

any arguments – it's been discussed elsewhere – and take a look at the generative mechanisms for the core I-Language mapping to C-I.

There are a number of questions, and problems, that arise; I will give a couple of examples of problems that I think are troublesome, and then go back and talk about how we can address them.

So to take one, consider the rule *raising to subject* and then the same rule of *raising to object*.[7] *Raising to subject* is the consequence of accepting the predicate-internal subject hypothesis. So there's this two rules and how they work. I had a proposal in my paper Problems of Projection, which was counter-cyclic[8] but it worked at the time. There was a good critique of it by Sam Epstein, Kitahara and Seely. They pointed out that first of all it violated the Extension Condition,[9] but much more seriously it involves an extra operation, a complex operation of substitution[10] of the newly Merged element in exactly in the place where it originally

7. In Rosenbaum (1965: 12), *raising to subject* was known as *Pronoun Substitution. Raising to object* is a case of NP complementation with *for-to* COMP deletion. Perlmutter (1968: 36) proposes a general raising rule which '*takes an NP out of the embedded sentence and moves it up into the higher sentence*'. Postal (1974) describes both processes, *raising to object* and *raising to subject* as a single rule of movement (1974: 267), and defines the rule *Raising*, crucially, as a *cyclic rule*.
8. It is useful in this context to first review the so-called 'Cyclic principle': essentially, when one domain to which transformations can apply is contained in another, relevant transformations apply to the smaller domain first, then proceeding to the wider domain. To quote Halle & Chomsky (1960: 275):

    The modifications [i.e., transformations] *are introduced in a stepwise fashion, successive steps reflecting the influence of successively higher constituents. Note also that the same modifications apply to all constituents regardless of their place in the constituent hierarchy; the same rules are reapplied to each constituent in a repeating cycle until the highest constituent is reached. The final result of such a cyclical reapplication of the same rules reflects to a certain extent the stress distribution of the morphemes as parts of lower constituents*.

    Chomsky (1973) formulates the Strict Cyclic Condition (SSC) as follows:

    No rule can apply to a domain dominated by a cyclic node [at the time, NP, S'] *in such a way to affect solely a proper subdomain of A dominated by a node B which is also a cyclic node*.

    Thus, a rule that targets a proper subdomain of a cyclic node is referred to as *counter-cyclic*. For example, in Chomsky (1995: 190), the ungrammaticality of **How did John wonder what Mary fixed* t_{how} is blamed on the counter-cyclicity of the operation that raises *how* to the matrix Spec-CP, later raising *what* to the embedded Spec-CP.

    The 'counter-cyclicity' that Chomsky mentions here can be exemplified as follows: if operations at T are triggered by C, then [C [T seem [$_\alpha$ Bill to have left]]] must be derivationally prior to the structure [C [T Bill seems [$_\alpha$ ~~Bill~~ to have left]]]. But because raising *Bill* affects, precisely, a proper subdomain of the cycle CP, this application of *raising-to-subject* is counter-cyclic. See Epstein, Kitahara & Seely (2014).
9. Formulated as follows:

    Suppose we restrict substitution operations still further, requiring that Ø be ***external*** *to the targeted phrase marker K. Thus, GT and Move-α extend K to K*, which includes K as a proper part.* (Chomsky 1995: 190)

    An even stricter version of the SSC, as Chomsky (1995: 190) observes, includes the effects of the EC.
10. Chomsky (1995: 248) defines that

    substitution forms L = {H(K), {α, K}} where H(K) is the head (= the label) of the projected element K, where K is a term of K. If L is a term of K, then the members of the members of L are terms of K (1995: 247), and adds that appeared, which is quite a complex operation. Counter-cyclicity is about the same as Late Merge,[11] so this critique holds for everything that is done with what's called Late Merge: it's completely unacceptable, because it involves operations that are complex, unmotivated, they have nothing to do with the goal we think we ought to obtain, something like the Strong Minimalist Thesis (SMT).[12] These considerations become far more significant in the case of what are sometimes called *exotic* constructions, those which have virtually no evidence, maybe none, available for the child; things like Antecedent-Contained deletion[13] or Across-the-Board movement, Parasitic gaps… It's simply impossible to propose a new principle for those, it can't be. The child has no evidence for them if he has to understand them. It must be the application of principles that are available for simple, easy, normal cases. So, in fact every kind of construction is in fact pretty exotic of the kind that Charles (Yang) was talking about but some are extremely so thus leading to the invocation of operations like counter-cyclicity, Late-Merge… completely unacceptable. We will rule this as part of the problems that we have to deal with. A lot of these proposals about Late Merge give very interesting descriptive results, bring all sorts of interesting ideas but without any basis, so what are apparently solutions are in fact problems, problems that now have to be addressed. And there's quite a lot of work like that, counter-cyclicity, Late Merge… One example is called Parallel Merge,[14]

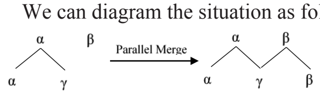
> *For the case of substitution, terms correspond to nodes of the informal representations, where each node is understood to stand for the subtree of which it is the root* (1995: 247)

thus respecting the Extension Condition (and therefore, the SSC). See fn. 31 below for the original (1995) formulation of Merge in terms of substitution as a Generalized Transformation.

11. See Stepanov (2001), in turn based on Lebeaux (1988): their proposal concerns the post-cyclic adjunction of XPs given (i) certain anti-reconstruction effects with respect to Condition C of Binding Theory, (ii) asymmetries on opacity in *wh*-extraction between arguments and adjuncts, (iii) Nissenbaum's (1998) analysis of parasitic gaps, (iv) the timing of *Affix Hopping* and adjunct intervenience effects, (v) multiple *wh*-fronting in Slavic and its interaction with Superiority (and, incidentally, the SSC), (vi) Principle A effects in relative clauses (which are taken to be adjoined structures). It is essential to bear in mind that Chomsky-adjunction must not be confused with Joshi-style adjunction in the framework of Tree Adjoining Grammars (Joshi 1985 and much related work).
12. In the words of Chomsky (2000: 96): *Language is an optimal solution to legibility conditions*. These legibility conditions are imposed by the external systems, Conceptual-Intentional and Sensory-Motor over the outputs of the *Narrow Syntax* (Merge + Agree + Move). See the discussion in Chomsky (2000: 112, ff.) for more details.
13. Cases like *John likes every boy Mary does* [~~like *t*~~]. See Fox (2002) for an analysis along the lines Chomsky is referring to here.
14. Citko (2005: 476)

> *The existence of External Merge and Internal Merge predicts the existence of a third type, combining the properties of both. This third type, which I will refer to as Parallel Merge, is like External Merge in that it involves two distinct rooted objects (α and β), but it is like Internal Merge in that it combines the two by taking a subpart of one of them.*

We can diagram the situation as follows, where α is a complex object {α, {α, γ}}:



which relies on work on multidimensionality, Sidewards Merge,[15] a lot of things in the literature. I'll come back to. But all of them are equally problematic for these reasons, particularly when they are used for more or less exotic constructions. In this particular case, *raising to subject* and *object* there seems to be an easy answer. The easy answer, which is in my recent papers, is simply to drop the condition that Internal Merge (Movement) has to be triggered, so it's free, like External Merge.[16] In fact, that's an improvement, we should never have had that condition. So dropping that condition, there's quite a straightforward cyclic analysis so we're in good shape in this particular case. But that's not true generally: if you look at the uses of Late Merge in the literature, a lot of them have interesting descriptive consequences, but don't have easy answers. There might be *some* answers, but they have to be worked on, these are challenges.

Well, that's the first kind of problem, but I am concerned with an extension of it, and there are many extensions of what is called Merge, but are not really Merge, in the literature, which raise very serious questions of legitimacy. Some of them in fact yield direct violations of quite sound principles. So they raise several questions.

15. As per Nunes (2004: 93, ff.) we can exemplify and illustrate Sidewards Movement (which assumes the Copy + Merge theory of movement) as follows:
    i) a given constituent α is copied from K and merges with the independent syntactic object L. $[_K \ldots \alpha_i \ldots] \rightarrow$ Copy $\alpha_i \rightarrow$ Merge(α_i, L), where $[_L \ldots]$, yielding $[_K \ldots \alpha_i \ldots]$, $[_M\ \alpha_i\ [_L \ldots]]$
    ii) at this point, the two copies of α cannot form a chain because they are not in a c-command relation. Thus, K and M are put together forming a new syntactic object XP, in which the copies of α can be in a c-command relation.
    $[_{XP}\ [_K \ldots \alpha_i \ldots]\ [_{X'}\ [X]\ [_M\ \alpha_i\ [_L \ldots]]]]$
    Note that K c-commands M, but neither copy c-commands the other (being embedded within K and M respectively).
    iii) If there is a higher head Y which selects a copy of α, such that YP transitively dominates XP and there is a copy of α in Spec-Y, this higher copy can form two distinct chains with the instances of α in *both* K and M.
16. An idea that goes all the way back to Lasnik & Saito's (1984) *Affect-α*. For triggered Internal Merge, see Chomsky (1995) and the argument about displacement, uninterpretable features, and language perfection in Chomsky (2000). Chomsky (2004: 110) defends Free Merge in the following terms:

    NS [Narrow Syntax] *is based on the free operation Merge. SMT entails that Merge of α, β is unconstrained, therefore either external or internal*.

    And, in footnote 29 (2004: 125):

    For over forty years, there have been efforts to motivate displacement. That seems to have been a mistake. Recourse to any device to account for the displacement phenomena also is mistaken unless it is independently motivated (as is Internal Merge). If this is correct, then the radically simplified form of transformational grammar ('Move-α' and its variants) is a kind of conceptual necessity, given the undeniable existence of displacement phenomena.

    Chomsky (2008: 139) adopts a different position, though, imposing a feature requirement on External Merge:

    For an LI [Lexical Item] *to be able to enter into a computation, merging with some SO* [Syntactic Object], *it must have some property permitting this operation. A property of an LI is called a feature, so an LI has a feature that permits it to be merged. Call this the edge feature (EF) of the LI. If an LI lacks EF, it can only be a full expression in itself; an interjection.*

The first question is 'how can they be ruled out?', 'what's a proper definition of Merge that rules them out?', and the second question is 'how do we deal with the descriptive results that are presented and described as solutions but we should think of as problems of the analyses?' A third kind of problem is raised by a very important part of what Charles Yang just talked about, his work on language acquisition,[17] which has some pretty remarkable results on the basis of very simple assumptions, basically the Elsewhere condition[18] and the assumption that listing something carries a cost for acquisition (which is almost tautological). On the basis of just those two assumptions he, along with Sam Gutmann, have managed to reach a very specific result, a definite tipping point that tells you –as much as possible– exactly where you should set a productive rule instead of listing.[19] And there are some pretty remarkable empirical results also for the first time on principles determining the core versus periphery distinction, quite valuable.

Now, the relevance for my concerns here however is different. One is that the work highlights the fact that we should keep in mind that an important condition on language design is that the languages should be learnable, it should be possible to acquire them. In fact, if language couldn't be acquired, it wouldn't survive. And that highlights the fact that we are not doing general recursion, we are not studying proof theory or metamathematics, we are studying a particular organic system which has its own natural conditions that it must meet, and that turns out to be important.[20]

17. Specifically, Yang (2016).
18. First formulated with this name in Kiparsky (1973: 94), see also Anderson (1969: 139-144), who calls it '*principle of disjunctive ordering*' and points towards antecedents in Panini. The general formulation in Kiparsky (1982: 8) suffices for present purposes:

    Rules A and B in the same component apply disjunctively to a form ϕ if and only if:
    (i) the structural description of A (the special rule) properly includes the structural description of B (the general rule)
    (ii) the result of applying A to ϕ is distinct from the result of applying B to ϕ.
    In that case, A is applied first, and if it takes effect, then B is not applied.
19. Chomsky is referring here to the Tolerance Principle, proposed in Yang (2016: 64):
    Let R be a rule applicable to N items, of which e are exceptions. R is productive if and only iff
20. See Turing (1952). The combination of purely formal and biological considerations is evident in the following passage:

    The model [a mathematical model of the growing embryo] *takes two slightly different forms. In one of them the cell theory is recognized but the cells are idealized into geometrical points. In the other the matter of the organism is imagined as continuously distributed. The cells are not, however, completely ignored, for various physical and physico-chemical characteristics of the matter as a whole are assumed to have values appropriate to the cellular matter*. (Turing 1952: 37)

    Given the enormous complexity of the matter, Turing makes the following methodological choice, which echoes much biolinguistic work

    The interdependence of the chemical and mechanical data adds enormously to the difficulty, and attention will therefore be confined, so far as is possible, to cases where these can be separated. The mathematics of elastic solids is a well-developed subject, and has often been applied to biological systems. In this paper it is proposed to give attention rather to cases where the mechanical aspect can be ignored and the chemical aspect is the most significant. (Turing 1952: 38)

Beyond that, there's a specific consequence of Charles' work, namely, if you take a look at it, his results depend on the assumption that rules are determinate [i.e., *deterministic*]: that means, if the structural conditions for a rule are met, the structural change has to take place in a fixed and determinate manner, and if you don't have that property, his results don't follow. And it's a pretty natural property, except it is violated all over the place. So take Phrase Structure Grammar (PSG):[21] if in a PSG you generate the structure, say:

1) NP, V, NP

And you have the rule

2) NP → Det, N

Then it's not deterministic. It could be either of the two NPs: the structural description is met, but you don't know what the result is. So that kind of rule really ought to be ruled out by the principle of determinacy. And it turns out that for these extensions of Merge that I've mentioned it's almost always violated. And it turns out one of the many considerations as to why they are unacceptable. Again, all these problems are particularly acute with the exotic constructions, the methodological principle is violated all over the place in descriptive practice, and it should be kept in mind.

Over the years, whenever some descriptive device has been introduced, and whatever it is (PSG, transformations, X-bar theory, parameters, phases, whatever it might be), almost always it tends to be used pretty extravagantly, well beyond the basis, of any solid foundation for the rule. And that's partly because it's not characterized explicitly enough, so there's a lot of vagueness in the periphery which is exploited for the descriptive purposes. And that's not necessarily a criticism. A good example, in fact, is Generative Semantics:[22]

21. See, e.g., Post (1943: 203, ff.), who presents what are known as 'Post canonical systems'. Also Chomsky (1959); Postal (1964); Kuroda (1964) and, for a perspective closer to automata theory, Hopcroft and Ullman (1969). Greibach (1965: 43, Definition 1.1) defines PSGs as follows:

    By a psg (I, T, X, ℘) we mean a context-free phrase structure grammar where
    (1) I is a finite vocabulary of intermediate symbols,
    (2) T is a finite vocabulary of terminal symbols and $I \cap T = \phi$,
    (3) X is the designated initial symbol [i.e., the *root* in terms of trees, the *axiom* in proof theoretic terms] *and* $X \in I$,
    (4) The rules of ℘ are of the forms, $Z \rightarrow AY_1 \ldots Y_n$, $n \geq 1$, $Z \in I$, $A, Y_i \in I \cup T$, *and* $Z \rightarrow \alpha$, $Z \in I$, $\alpha \in T$.

    Linguistically, they neatly capture Immediate Constituent analyses, and are opposed to Dependency Grammars.

22. See McCawley (1968 [1973]); Lakoff (1971); Lakoff & Ross (1976) (and also Dowty 1979). The basic tenets of a 'vanilla' GS (the common aspects to all variants) are presented clearly in McCawley (1968 [1973]: 155-156):

    (1) Semantic structures are claimed to be of the same formal nature as syntactic structures, namely labeled trees whose non-terminal node-labels are the same set of labels that appear in surface structure.

transformations were around, the Katz-Postal hypothesis[23] was around, Deep Structure interpretation was around, and that led to really quite extravagant use of these devices, which had both a positive and a negative aspect, finally collapsed from its own weight because it was so extensive that it didn't make sense. But the advantages were that it led to a lot of discoveries, there were lots of insights about language that came out of it. They're not solutions, they're problems, and it's good to have problems, and led to explorations of new domains that hadn't been looked at. All of that's positive, and that's commonly true for the promiscuous use of devices that are invented. The negative aspect is that it doesn't lead us to the goal of trying to understand UG and the language faculty, and it's also misleading in that it tends to present problems, which are interesting problems,

---

(2) The notions of a 'deep structure' which separates syntax from semantics and a distinction between 'transformations' and 'semantic interpretation rules' are given up in favor of a single system of rules which relates semantic structure and surface structure via intermediate stages which deserve the name 'syntactic' no more and no less than 'semantic'.

(3) It is held that the rules needed to determine what a grammatical sentence may mean are also needed to determine what is grammatical at all.

(4) A grammar is taken not to generate a set of surface structures but a set of derivations, and to consist of a set of derivational constraints: constraints on what combinations of elements may occur in semantic structure, on what combinations of elements may occur in surface structure, and on how different stages of a derivation may differ from each other.

23. Katz & Postal (1964). Two aspects of that proposal are relevant here:

The semantic interpretation of the sentence
The man hit the ball
Must represent the meanings of the constituents of this sentence, i.e., <u>the</u>, <u>man</u>, <u>hit</u>, <u>the</u>, <u>ball</u>, <u>the man</u>, <u>hit the ball</u>, and <u>the man hit the ball</u>. But it must not provide any meaning for such substrings […] as <u>the man hit</u> or <u>hit the</u>. Obviously, this condition of adequacy can be fulfilled only if the syntactic component provides, for each sentence it generates, an enumeration of all, and only, its constituents (Katz & Postal 1964: 20).

The fragment above explicitly adopts the definition of 'generative' used in Post (1943) and requires determinacy to hold. Also, and perhaps more importantly for Chomsky's point,

there are also many cases in the literature of syntactic facts characterized by optional singulary transformations where the output P- marker must have a semantic interpretation quite different from that of the input P- marker. Among these are the question transformation, the imperative transformation, the wh-attachment transformation, etc. Thus there are three possibilities: first, that no <u>correctly formulated</u> singulary transformation has an output with a semantic interpretation distinct from its input and that those transformations in the literature which violate this claim are incorrect; second, that all singulary transformations affect meaning and those in the literature which do not are incorrect; third, that some do and some do not affect semantic interpretation and it is some specific feature of the particular transformations that determines which do and which do not.
The first two alternatives are clearly preferable, even though what at present appear to be the facts throw more doubt upon them than upon the third, because they make no reference to specific features of a class of transformations. […] *on a priori methodological grounds, the first of the three alternatives is the one which deserves to be provisionally accepted. This alternative claims that P2* [Type 2 Projection rules, a set of rules that apply to strings that have been applied transformations] *play no role in the semantic interpretation of any sentoid* [a sequence of grammatical formants with an associated semantic interpretation] *without a generalized transformation in its T- marker*. (Katz & Postal 1964: 32. Underline in the original)

as if they were solutions, and they are not solutions: they are ways of stating a problem that we have to look at.

Well, I think that a lot of this is now happening with Merge-style systems. These systems have had quite a lot of real and, I think, significant contributions. So, there's accounts for the existence and ubiquity of displacement and reconstruction, which has always been regarded as an oddity of language, but it turns out to follow from the null hypothesis: if you pick the simplest combinatorial operation you get displacement and reconstruction, so that's a pretty significant result, I think. It accounts for the deep and quite puzzling property of structural dependence, which has been worried about for 50 years: why do languages have this strange property, which increases computational complexity of use of language, since dealing with linear order is much more trivial computationally? But it's nevertheless ubiquitous. Why is that the case? Well, it turns out again that it follows from the null hypothesis, if you pick the simplest combinatorial operation, that's what you get. Incidentally, this alone has many consequences for the future, raises difficult problems: it follows from this that anything involving linear order or any other arrangement cannot feed C-I. But there's overwhelming evidence to the contrary: in fact, the whole history of linguistics assumes the opposite, right up to the present. It assumes that things like linear order and arrangement are what yield semantic interpretation and interact all over the place with syntactic operations. And there's plenty of interesting contemporary work that seems to suggest the same thing. But if this is correct, and there's good reason to think it is, those indeed give the *only* explanation for structural dependence, in fact the best possible explanation. We have a real problem, like the problem of the extravagant use of devices: major areas of descriptive wealth have to be completely rethought. We have to ask how we can show the descriptive consequences of using linear order or arrangement have to be settled in some other way, they have to be assigned to the externalization system. There's also interesting neurolinguistic and psycholinguistic experiments that suggest exactly the same thing, including work that Ianthi [Tsimpli] talked about before.

These results that I mention, if you take a look at them, were all achieved within a *narrow* version of Merge, not using the eccentric versions I've been raising questions about. The narrow version is pretty well defined in itself: it involves simple combinatorial operations and relies on the observation that there are two logical possibilities, External and Internal Merge. But the *narrow* version has been used within a framework that has been left kind of vague and unspecific. And that's a problem. It's the vagueness that has been exploited for the extensions of Merge like counter-cyclic Late Merge, Parallel Merge to yield Multidominance,[24] and so on. Well, there are the usual advantages and problems that I've mentioned, but I

24. The various theories grouped under the term 'Multidominance' can be said to have in common the rejection of the so-called *Single Mother Condition* (Sampson, 1975), such that a node in a structural description can be dominated by more than a single node (i.e., can have more than one 'mother'). See Citko (2005); Peters and Richie (1981); McCawley (1982); Levine (1985). The locally multi-rooted graph proposal in Morin & O'Milley (1969) (who refer to these structures as *vines*) is also relevant.

think that these extensions are illegitimate and we have to somehow show that and show that the *narrow* version, which yields the interesting results, is somehow the only legitimate one. That's the question I will look at.

Well, there are two ways to proceed. One is the boring way, so I'll ignore it: that's to stipulate the narrow version explicitly: say 'this is what it is, the others don't work'. We don't want to do that. The interesting way is to take a look at the computational operations of language from a completely different point of view, to start by asking 'what are the general desiderata that any computational operations for language should meet?' Notice there are two considerations here: first, what should *any* computational operation be like, simply on grounds of computational complexity and third factor conditions?[25] And second, it's got to be *specific to language*, an organic system which has its own properties. So there are those two conditions or desiderata. And the idea will be that the program will be constructed in a general framework that accommodates a wide range of alternatives including all the extensions, including other things that we can think of that haven't been used yet. And then ask: what survives careful analysis in terms of these conditions? In the hope of showing that only the *narrow* version passes muster under these conditions, and which then leaves us with the challenge of facing the wealth of descriptive results that have been reached by what I hope to show are illegitimate means and methods. The better approach is the one that provides insight into the nature of language and explanatory adequacy, so let's proceed with it first formulating some general principles that *any* operation for language ought to meet.

The first and most obvious one is simply *descriptive adequacy*. It ought to get the facts right. And of course we all know that's not innocuous, as we don't know what the facts are *a priori*. Maybe something we think is a fact about language turns out to be a *performance fact*. What counts as a fact depends on theoretical understanding and empirical discoveries. Nevertheless, it's a pretty good guideline to proceed and I should say that, now that we have –thanks to Charles [Yang]– a sharp *core-periphery* distinction, that helps independent study on the fact of what is a performance property, that also helps. And we have reasonably good guidelines, I think it's safe to start from that. But it poses a problem. One problem is that Merge violates it. Merge does *not* satisfy descriptive adequacy. This is a fact that's pretty crucial. So if you take a look at the *simplest* case of Merge, that is Internal Merge, that's the one that involves least search, External Merge involves huge search, it's a complicated operation.[26] Internal Merge only involves search within the syntactic

25. In the words of Chomsky (2005: 6), the *third factor* in language design includes:

iii) Principles not specific to the faculty of language.

The third factor falls into several subtypes: (a) principles of data analysis that might be used in language acquisition and other domains; (b) principles of structural architecture and developmental constraints that enter into canalization, organic form, and action over a wide range, including principles of efficient computation, which would be expected to be of particular significance for computational systems such as language. It is the second of these subcategories that should be of particular significance in determining the nature of attainable languages.

26. The 'search space' in IM is limited to the local phrase marker under consideration, in EM the search space is potentially the entire Lexicon.

object that we are looking at. So suppose we apply just Internal Merge, then what do we get? Well, basically we get the successor function. So suppose you have a single element lexicon and we apply Merge to it; we have a single element, call it 0, and we apply Merge to it and you get the set containing 0, call that 1. Apply it again, you get the set containing 0 and 1, it's 2. And basically you get the successor function.[27] But language isn't the successor function, maybe arithmetic is. And in fact that's part of the argument that maybe that's why humans know arithmetic, because it's the simplest case of language. But that's not language. Well, suppose you apply Internal Merge. Then, if you think about it, what you get is (if you look at it in terms of trees) a tree in which the leaves are lexical items, whatever lexical items are, you get a tree with lexical items coming off. Well, that's not language. So, these two operations, incidentally, are appropriate for standard *formal* systems.[28] So if you are making standard formal systems you can use these methods, but not language, because language has a different property: language has exocentric constructions, it has things of the form {XP, YP} like, say, {NP, VP}, and you can't get them by just internal Merge, so there's a problem. The operation Merge violates the simplest condition: descriptive adequacy. Well, that has (kind of) been overcome by a tacit assumption, and this tacit assumption has to be made precise, as it has consequences. And the tacit assumption is that you can construct syntactic objects in parallel and then bring them together somewhere. Now, that presupposes that you have a workspace in which operations are being carried out. And, what's the workspace? Well, that hasn't been properly answered. And fixing it has consequences. The one immediate consequence is that operations, say, the right version of Merge should be operations on the workspace, not on a particular syntactic object, because they can change the workspace. And in fact, if you look at these –what I think are- illegitimate operations like, say, Parallel Merge, they in fact involve *separate* elements of the workspace, so they are modifying the entire workspace. And since the program I am suggesting is to present desiderata that include everything, these fall within it, just like much else.

27. The recursive set-theoretic construction of the naturals is due to Von Neumann, e.g., (1923) (within the more general framework of Zermelo-Fraenkel set theory), and it is equivalent to Peano's (which was however *not* based on set theory). First, let

$s(\mathrm{a}) = \mathrm{a} \cup \{\mathrm{a}\}$, call s the *successor function*

And

$0 = \emptyset$

Then, define

$1 = s(0) = s(\emptyset) = \emptyset \cup \{\emptyset\} = \{0\}$
$2 = s(1) = s(\{0\}) = \{0\} \cup \{\{0\}\}$

But we know that $\{0\} = 1$, so

$2 = \{0\} \cup \{1\} = \{0, 1\}$

Which means that 2 is defined as the ordered set containing 0 and 1. In the same way, 3 is defined recursively as

$s(\{0, 1\}) = \{0, 1\} \cup \{2\} = \{0, 1, 2\}$ (or, equivalently, $\{0, 1, \{0, 1\}\}$)

28. Standard here just means the kind usually discussed in general (including linguistic) contexts: propositional calculus, quantification theory, arithmetic…. Not say category theory. (Chomsky PC)

So the operations, including the right version of Merge, have to be operations on the workspace. That raises all kinds of questions, to which I'll turn directly.

The first condition is *descriptive adequacy*.

The second condition is some version of the Strong Minimalist Thesis, I mentioned reasons why we ought to be able to approach that. There are specific consequences like, for example, Inclusiveness:[29] the operations should not *add* anything: they should not add order, they should not add new features, or anything like that. Of course, externalization violates all these conditions, it violates just about everything, so that's not surprising.

A third condition, which is specific to language, is not true of, say, formal proof theory or something like that, is that we should be restricting the computational resources. We are dealing with an organic system with limited computational resources. In fact, quite limited, if you think about the speed of neural transmission and so on and so forth. So a third principle is *restrict computational resources*, and in the best case (the case that we ought to try to achieve, if possible), the operations should never extend the workspace; they should maybe contract it, but not expand it.

The fourth principle, I already mentioned, is *determinism*, the principle that was required for Charles' results. So, if the structural conditions for a rule holds for some workspace, then the structural change must be unique, it must be determinate, unlike, say, phrase structure grammar.

The fifth condition is, centrally, a condition of *coherence* or *stability* that says that the properties of a syntactic object can't change in the course of the derivation,[30] so something that refers to *Mary* on line 1 cannot refer to *John* on line 3. There's an interesting history about this, that there's no time to go into, but in the history of science it turns out that in classical physics and mathematics this condi-

29. Chomsky (1995) defines the Inclusiveness Condition as follows:

> *A "perfect language" should meet the condition of inclusiveness: any structure formed by the computation (in particular, π and λ) is constituted of elements already present in the lexical items selected for N; no new objects are added in the course of computation apart from rearrangements of lexical properties (in particular, no indices, bar levels in the sense of X-bar theory, etc.* (Chomsky 1995: 228)

Also, footnote 7 of Chapter 4, p. 381:

> *Note that considerations of this nature can be invoked only within a fairly disciplined minimalist approach. Thus, with sufficiently rich formal devices (say, set theory), counterparts to any object (nodes, bars, indices, etc.) can readily be constructed from features. There is no essential difference, then, between admitting new kinds of objects and allowing richer use of formal devices; we assume that these (basically equivalent) options are permitted only when forced by empirical properties of language.*

30. See, e.g., Lasnik & Uriagereka's (2005: 53, 112) Conservation Laws:

> *1st Conservation Law: Conservation of Lexical Information*
> *All information in a syntactic derivation comes from the lexicon and interpretable lexical information cannot be destroyed.*
> *2nd Conservation Law: Conservation of Structural Information*
> *Interpretable structural units created in a syntactic derivation cannot be altered*

The notion of *faithfulness constraint* in Optimality Theory is also consistent with this dictum.

tion was crucially violated, as in Newton's work, with quite major consequences. But anyway, this condition *has* to be satisfied.

And the sixth condition is the fact that language crucially involves *recursion*. That's a universal human property of the human faculty of language (there's a lot of confusion about this in the popular literature, which I will not go into). It's an invariant property of humans that the language faculty involves recursion. Well, what's the basic idea of recursion? It's that every object that's generated must be available for later computations. So, for example, if you're doing formal *proof theory*, and you prove a theorem, you have to be able anywhere later in the proof to go back to that theorem and follow its consequences. That's formal proof theory, but remember, we are not doing formal proof theory, we have an organic object which has to meet other conditions. But in our case we want to try this sixth condition given condition 2, the SMT, we want to formulate *recursion* in a way that stipulates no specific properties, so we don't put any extra conditions on it: recursion ought to be *free*. So what we'll say is that a syntactic object is *accessible* (that's a technical term) to further operations if it's been generated, period. So that's general recursion without further stipulations.

Well, I'll stop with these conditions for time reasons. The basic ones are the first and second (descriptive adequacy and SMT), if we think it through, the others pretty much follow.

There are consequences right away from the definition of *recursion*. One consequence shows that what's been called the Extension Condition is a mistake, because the Extension Condition simply stipulated that the only accessible syntactic objects are the whole syntactic objects, that's the Extension Condition. But the general definition of recursion tells us that *anything* inside should be accessible. And that means that one of the arguments against Late, counter-cyclic Merge doesn't work, the one that said 'it violates the Extension Condition'. It doesn't matter in this case, because the *major* argument, the EKS [Epstein, Kitahara, Seely] argument against Late, counter-cyclic Merge is the *substitution* operation, which is unacceptable. So the conclusion holds, but not the entire argument, the Extension Condition [one] has to be withdrawn.

Much more interesting is what happens if we take the simplest cases. Suppose the workspace consists of two elements, the set $\{a, b\}$ (whatever a and b are). So we have the result of Merging a and b. That's one of the things in the workspace. The second thing in the workspace, call it c, a syntactic object. So that's the simplest case, and it's worth a very close look because it turns out that the problems that undermine, I think, *all* of the extensions of Merge, already show up with this simple case, so it's worth looking at this simple case carefully:

3) $\mathrm{W} = \{\{a, b\}, c\}$

So let's take a look at that and the problems that it raises.

Notice first that a and b have to be accessible to further operations by the definition of *recursion*. And in fact they are, because they are part of the first element, the set $\{a, b\}$. It has already been generated, so it's accessible. However, if you look further, the condition of *recursion* does suggest that the workspace that results from the operation that yields (3) should be as follows. If we Merge:

4) $\mathrm{W} = \{\{a, b\}, c\}$

Merge

The definition of recursion suggests that the workspace should be:

5) $\{\{\{a, b\}, c\}, a, b\}$

Because *a* and *b* are both accessible. And, as I mentioned, that's not a strong argument. But the general notion of recursion works like this. So if you're doing, say, *proof theory*, the axioms and every line you've generated are still there, they are part of the workspace (we don't need a workspace for proof theory, but the set created by the operation would be (5)). So that raises a question: is that the right answer? Should the result of Merge, or the improvement of Merge, be an operation on the workspace which yields (5)? Well, suppose we are to do this.

Notice that first of all we violate the condition of *descriptive adequacy* [condition 1], and the reason is quite simple: we have (4) in the workspace and it has to be available for further operations, which can yield (6)

6) $\{\{\{a, b\}, c\}, a, b\}$

X

X is a structure of arbitrary complexity. Since *a* and *b* are in the workspace we could merge *a* to X

7) $\{\{\{a, b\}, c\}, a, b\}$

X

And it means that we have what amounts to a movement operation which violates every possible rule on movement, so we can't accept that. Furthermore, it violates condition 3, the condition on *not expanding the workspace* [compare the size of the workspace in (4) and (5)].

Another reason is that it violates *determinism* [condition 4]: suppose that there's some other operation going on that targets, say, *a*, we don't know which *a* it applies to. So already there are three conditions that are violated.

And this goes back to Feynman's argument: we are getting somewhere because we have several independent lines of argument (each reasonable in itself) which yield the same result.

Well, there is a proposal, very widely used in the literature that we are talking about, which says we can overcome this by developing a new theory of movement, one which doesn't involve Internal Merge. So suppose we bar Internal Merge and develop a *representational* theory of movement, which just involves External Merge. That's very widely used, in multidimensionality. That's a terrible idea: it violates every condition you can think of, and it has its own problems.

So take for example Topicalization or Left Dislocation:

8) [Mary's book]$_1$, John read [Mary's book]$_2$

This is the underlying structure. If you interpret it as Topicalization, then [Mary's book]$_1$ has to be identical in all respects to [Mary's book]$_2$. So, if Mary owns the book, that's the topic, the book that Mary owns. But then what John read, it has to be the book that she owns, not the one that she wrote, and it's the same Mary. On the other hand, in the Left Dislocation case this isn't true, they are totally unrelated. So we can say 'as for Mary's book', that's what we're talking about, John read Mary's book, maybe it's a different Mary, maybe it's the book she owns…

In the Topicalization case, [Mary's book]$_1$ and [Mary's book]$_2$ need to be *copies*. In the Left Dislocation case, they are repetitions. Well, that's a critical distinction. And in this new representational theory, there's no way of describing it. And in the old theory, the *narrow* theory, there are *trivial* ways of describing it: simply define the concept of *copy* as something formed by Internal Merge, everything else is a repetition. And that captures exactly the intuition behind this, namely, that if there's something *new* that comes into the derivation from the outside it's a repetition, that has nothing to do with what's inside. So if I say

9) John saw John

They are different guys. If it's a *copy*, it's something *inside* the derivation. It's not adding anything new from the outside. And that's trivially computed at the phase level. So within the *narrow* version a simple answer can be made to work, that's all. But in the new theory there's no answer at all, other than abandoning the hope of distinguishing copies from repetitions. And that's incidentally only the beginning of a lot of other difficulties.

Another difficulty with this approach is that you are barring Internal Merge. But that's the simplest possible operation. So in order to follow this approach you are saying 'the simplest operation is ruled out' for no reason. And incidentally, barring Internal Merge means losing the explanation for the ubiquity of displacement and reconstruction, which is a pretty big result.

Another consequence is that the elements manipulated by External Merge can be of arbitrary complexity, it could be anything at all, which means a huge amount of extra computation. And to amplify that, this new object, as we know, has to be inserted at every point of the successive cyclic operation, because as you know there are consequences at the various points of insertion in movement, at the *v*P phase and the CP phase, there are both semantic and phonological output conditions (as Doug [Saddy] showed many years ago in Indonesian).[31] So you have massive new computation which has to be introduced at every point of the successive cyclic operation, and it goes on like that.

The result is, there's lots of loss and no gain whatsoever, because these constructions are still ruled out by other conditions. So I think that's a non-starter. But how do we approach the matter? We have to redefine Merge as an operation of *replacement*, so that Merge would say:[32]

10) Replace(a, b) by $\{a, b\}$

That's the way Merge was defined, back in 1995,[33] in the initial publications in the Minimalist Program. Incidentally, it was defined as follows

10') a. Replace(a, b) by $\{a, b\}$

b. Eliminate(a, b)

Some of the recent formalizations call it Remove. But, where do you eliminate it *from*? Well, that was not answered. But now we have an answer, we know where it's eliminated from, it's the *workspace*. So we have an operation which replaces a and b by the set $\{a, b\}$ and eliminates a and b from the workspace; that will overcome the problems that we discussed with the earlier example (5). But we want to do this without a new rule *Eliminate*. We're trying to keep to the SMT, we don't want new rules or anything like that.

The simplest way to do this is the following:

Suppose we have a workspace with a set of objects. From that workspace we can determine a sequence Σ:

31. Saddy (1990, 1991).
32. See also Collins (2017); Epstein, Kitahara & Seely (2015).
33. The relevant definitions are the following:

> *We now adopt (more or less) the assumptions of LSLT, with a single generalized transformation GT that takes a phrase marker K^1 and inserts it in a designated empty position Ø in a phrase marker K, forming the new phrase marker K*, which satisfies X-bar theory. Computation proceeds in parallel, selecting from the lexicon freely at any point. At each point in the derivation, then, we have a structure Σ, which we may think of as a set of phrase markers. […] GT is a substitution operation. It targets K and substitutes K^1 for Ø 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. Note that this is a description of the inner workings of a single operation, GT.* (Chomsky, 1995: 189)

In this context, GT (Merge) is defined as a *binary substitution operation*, and Move-α is its *singulary* version. In 'Categories and Transformations' Chomsky defines GT as follows:

> $C_{(HL)}$ *must include a second procedure* [other than *Select*] *which combines syntactic objects already formed. A derivation converges only if this operation has applies often enough to leave us with just a single object, also exhausting the initial Numeration. The simplest such operation takes a pair of syntactic objects (SO_i, SO_j) and replaces them by a new combined syntactic object SO_{ij}. Call this operation Merge.* (Chomsky 1995: 226)

11) $\Sigma = (X_1, X_2, \dots X_n)$

That has the following properties: it's the shortest sequence such that

(i) Each X_i is *accessible* (that's the definition of *recursion*)
(ii) Σ exhausts the workspace

And we can define a new operation called MERGE:

12) $\text{MERGE}(\Sigma) = \{\{X_1, X_2\}, X_3, \dots X_n\}$

That's a *replace* operation. It replaces X_1 and X_2, the first members of the sequence, by a *set*, and it doesn't have any *Remove* operation.

If you look at this, a couple of things follow.

First of all, this says that you can take *any* two accessible elements in the workspace, any at all, and you can MERGE them in the 'capitals' sense [i.e., in the sense of (12)], and you map the workspace into a new workspace. This happens to accommodate all of the extensions that are around, plus more than you can think of. And the next step you have to take is to make sure that the *legitimate* operations External Merge and Internal Merge are included. And they are, EM, in fact, yields exactly the same results if you think it through you'll see that EM in the old definition (now extended to a definition of the workspace) yields (12) and IM yields exactly the same thing, which makes sense, because they are the same operation; they are just two possible cases of the same operation. So, naturally, they yield the same results: Internal Merge is, again, the simplest operation, with the least search.

MERGE, in this new definition, satisfies condition 2, SMT: it's the simplest computational operation you can think of on the workspace, and it excludes the bad cases. It keeps the workspace from expanding: notice that the new result is not larger than the original one. And in fact External Merge always *reduces* the workspace by one; Internal Merge keeps it the same. So we are *not* violating condition 3, as the workspace is not expanding indefinitely.

Now, there's an obvious qualification here: sentences can be longer and longer, so there's got to be some way of building up the workspace. The minimal way to extend the workspace, therefore the one we want to keep to, is to take two lexical items *a*, *b* out of the Lexicon (it doesn't matter what the Lexicon is; certainly not words, but that's been recognized since LSLT) and form from them the set {*a*, *b*}. If the Borer-Marantz idea about roots and categorization of roots is correct[34], then

34. Marantz (1997: 215) defines this position as follows:

Roots like √DESTROY and √GROW [...] *are category neutral, neutral between N and V. When the roots are placed in a nominal environment, the result is a "nominalization"; when the roots are placed in a verbal environment, they become verbs.*

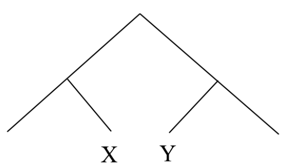
These 'environments' are defined shortly after:

Among the functional heads in whose environments roots become verbs (these may be "aspectual" in some sense), one, call it "v-1," projects an agent while another, call it "v-2," does not. These little "v's" could be different flavors of a single head, or perhaps there is some unified these will always be composed of a root and a categorizer, for example *n*, *v*, that says what category it is. In any event, two things: notice that it's not enough to take *one* of them out of the workspace because that will just External Merge whatever we have and it won't add anything. But if you pick these two things out of the workspace then you can go on to build a new syntactic object. So that's the minimal way to allow the workspace to grow, if you can get away with it, of course, and still keep to the SMT. Other than that, the operations themselves never expand the workspace, so that's satisfying the third condition.

I'll mention some things than have to be done. You have to show that the two legitimate operations, EM and IM (which are the same operation) satisfy *determinacy* [condition 4]. That's not trivial. But if we look at it, there's an answer, as to how they satisfy determinacy.

Then, you have to look at other cases. I'll just mention a couple I won't go through. So, for example, one case that follows from the general framework looks like this:

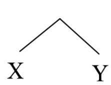
13)



And we Merge {X, Y}. You take *two* things inside the syntactic object, and you Merge them. I don't think anybody suggested that, but it falls within the general framework. And if you think it through, you'll see that it violates all the conditions that the simplest case violates.

So now let's take one that is in the literature for which there has been quite a lot of descriptive work and that is Parallel Merge. So we have an object

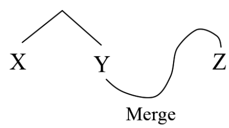
14)



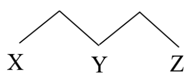
That's one. And we have some other one, call it Z, and that's the whole workspace. And then we Merge:

account that could have a single head optionally project an agent and thus cover both v-1 and v-2. (Marantz 1997: 217)

Conversely, the 'nominalizing environment' is taken in Marantz (1997: 218) to be the domain of D. More recent proposals have proposed a 'little *x*' for each environment, such that *v* is a verbalizer, *n* is a nominalizer, etc. (see Borer 2005a, b). The root (pun intended) of the transformational approach to nominalizations, which is essential to understand these proposals, is to be found in Lees (1960).

15) 

Forming {Y, Z}. That's what's called Parallel Merge. And it's usually described as Internal Merge, so that Y is internally merging to Z. But that's *not* Internal Merge. It's some new kind of Merge which doesn't fit the *narrow* version. It's usually written like this:

16) 

That's the notation used in the Multidominance literature. But that notation doesn't *mean* anything, there's no such object. In fact, in general, trees are an extremely misleading notation, the root may have nothing there, there may not be a root. And this temptation to draw all sorts of lines, that's all over the place, and much of it can't be reduced to Merge. I won't go through this today, I'll leave it as an exercise. But (16) has exactly the problems of the trivial case, it violates all the same conditions.

That, incidentally, eliminates all the work on merge-based Multidominance, which has led to very interesting results, ATB, all these things. But they are left as problems. And again, if you think about Richard Feynman's observation, that's what we are finding, case by case.

Well, pretty much the same is true of other conditions; Sidewards Merge is even more complicated, there's lots of other problems. But as far as I know, every case that's in the literature aside from the narrow version and other cases that you can dream of that aren't in the literature violate all the conditions that I've mentioned. Again, I'll leave it as an exercise. But what was intuitively the old version of EM and IM (*intuitively*, because it was never stated precisely), with the notions *Replace* and *Eliminate*, the question '*Eliminate* from what?' was never answered. And then when you answer it, you conclude that the operation Merge has to be redefined as an operation on the *workspace*, one which doesn't have this word *eliminate* and meets the conditions we listed.

And I think that converges down to what pretty much was assumed intuitively. Which means, in fact, that for practical purposes (not being precise) you *can* use the old definition, which you know is wrong, because it's the only definition, I think, which satisfies a range of legitimate general desiderata on what every operation ought to be. So all of those background conditions are a framework in which any computational operation for language must be selected. And it happens, I think, to work out [that] that was what we were using intuitively in the narrow version.

Well, that leaves a couple of big problems, first of all showing that what I just said is correct. Which would mean solving a lot of problems which I am leaving for you to solve.

But what doesn't work and is easily a big problem is taking everything that's been described in the extensions of Merge, like everything in the Multidominance theory and all the rest, and showing that there is a legitimate way to pay off those promissory notes – I think if you look at that there may be some interesting ways. So there's lots of work to do. I'll stop there.

References

Anderson, Stephen R. 1969. *West Scandinavian Vowel Systems and the Ordering of Phonological Rules*. PhD dissertation, MIT.

Berwick, Robert. 1985. *The acquisition of syntactic knowledge*. Cambridge, MA: MIT Press.

Borer, Hagit. 2005a. *In Name Only. Structuring Sense, Volume I*. Oxford: Oxford University Press.

Borer, Hagit. 2005b. *The Normal Course of Events. Structuring Sense, Volume II*. Oxford: Oxford University Press.

Borer, Hagit. 2017. The Generative Word. In James McGilvray (ed.). *The Cambridge Companion to Chomsky*. 2nd Ed. Cambridge: CUP, 110-133.

Chomsky, Noam. 1955. *The Logical Structure of Linguistic Theory*. Mimeographed, MIT. Available at <http://alpha-leonis.lids.mit.edu/wordpress/wp-content/uploads/2014/07/chomsky_LSLT55.pdf>.

Chomsky, Noam. 1959. On Certain Formal Properties of Grammars. *Information and Control* 2: 137-167.

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

Chomsky, Noam. 2000. Minimalist Inquiries: The Framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.). *Step by Step – Essays in Minimalist Syntax in Honor of Howard Lasnik*. Cambridge, Mass.: MIT Press, 89-155.

Chomsky, Noam. 2004. Three Factors in Language Design. *Linguistic Inquiry* 36(1): 1-22.

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

Chomsky, Noam. 2015a. Problems of Projection: Extensions. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.). *Structures, Strategies and Beyond: Studies in honour of Adriana Belletti*. Amsterdam: John Benjamins, 1-16.

Chomsky, Noam. 2015b. *The Sophia Lectures*. Sophia Linguistica 64. Sophia University, Tokyo.

Citko, Barbara. 2005. On the Nature of Merge: External Merge, Internal Merge, and Parallel Merge. *Linguistic Inquiry* 36(4): 475-496.

Collins, Chris. 2017. Merge(X, Y) = {X, Y}. In Leah Bauke & Andreas Blühmel (eds.). *Labels and Roots*. Berlin: Walter de Gruyter, 47-68.

Dowty, David. 1979. *Word Meaning and Montague Grammar*. Springer Verlag.

Epstein, Samuel D., Hisatsugu Kitahara & T. Daniel Seely. 2014. Labeling by Minimal Search: Implications for Successive-Cyclic, A-Movement and the Conception of the Postulate "Phase". *Linguistic Inquiry* 45(3): 463-481.

Epstein, Samuel D., Hisatsugu Kitahara & T. Daniel Seely. 2015. Simplest Merge Generates Set Intersection: Implications for Complementizer-Trace Explanation. In Samuel Epstein, Hisatsugu Kitahara & T. Daniel Seely. *Explorations in Maximizing Syntactic Minimization*. New York: Routledge, 175-194.

Fodor, Jerry A. 1975. *The Language of Thought*. Cambridge, Mass.: Harvard University Press.

Fodor, Jerry A. 2008. *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.

Fox, Danny. 2002. Antecedent-Contained Deletion and the Copy Theory of Movement. *Linguistic Inquiry* 33(1): 63-96.

Goodman, Nelson. 1968. *Languages of Art: An Approach to a Theory of Symbols*. New York: The Bobbs-Merrill Company.

Greibach, Sheila. 1965. A New Normal-Form Theorem for Context-Free Phrase Structure Grammars. *Journal of the ACM* 12(1): 42-52.

Halle, Morris & Noam Chomsky. 1960. The morphophonemics of English. *MIT RLE Quarterly Progress Report* 58: 275-281.

Huybregts, M. A. C. (Riny). 2017. Phonemic clicks and the mapping asymmetry: How language emerged and speech developed. *Neuroscience & Biobehavioral Reviews* 81: 279-294.
<https://doi.org/10.1016/j.neubiorev.2017.01.041>

Katz, Jerry & Paul Postal. 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: MIT Press.

Kiparsky, Paul. 1973. 'Elsewhere' in Phonology. In Stephen R. Anderson & Paul Kiparsky (eds.). *Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston, 93-106.

Kuroda, Sige-Yuki. 1964. Classes of languages and Linear-Bounded Automata. *Information and Control* 7: 207-223.

Lakoff, George. 1965. *On the Nature of Syntactic Irregularity*. PhD dissertation, Indiana University.

Lakoff, George. 1968. Instrumental Adverbs and the Concept of Deep Structure. *Foundations of Language* 4(1): 4-29.

Lakoff, George. 1971. On Generative Semantics. In Danny D. Steinberg & Leon A. Jakobovits (eds.). *Semantics - An Interdisciplinary Reader in Philosophy, Linguistics, Anthropology and Psychology*. Cambridge: Cambridge University Press.

Lakoff, George & John Robert Ross. 1967 [1976]. Is Deep Structure Necessary? Personal letter to Arnold Zwicky, printed in James D. McCawley (ed.). *Syntax and Semantics 7: Notes from the Linguistic Underground*. New York: Academic Press, 159-164.

Lasnik, Howard & Mamoru Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15: 235-289.

Lasnik, Howard & Juan Uriagereka. 2005. *A course in Minimalist Syntax*. London: Wiley-Blackwell.

Lebeaux, David. 1988. *Language acquisition and the form of the grammar*. Ph.D. dissertation, University of Massachusetts, Amherst.

Lees, Robert. 1960. *The grammar of English nominalizations*. Bloomington: Indiana University Press.

Levine, Robert D. 1985. Right node (non-)raising. *Linguistic Inquiry* 16(3): 492-497.

Marantz, Alec. 1997. No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark & A. Williams (eds.). *Proceedings of the 21st Penn Linguistics Colloquium*, UPenn Working Papers in Linguistics, 201-225.

McCawley, James D. 1968. Lexical insertion in a transformational grammar without deep structure. *CLS 4*: 71-80. Cited by the reprint in McCawley. 1973. *Grammar and Meaning*. Tokyo: Taishukan, 155-166.

McCawley, James D. 1982. Parentheticals and Discontinuous Constituent Structure. *Linguistic Inquiry* 13(1): 91-106.

Morin, Yves-Charles & Michael O'Malley. 1969. Multi-rooted vines in semantic representation. In Robert Binnick et al. (eds.). *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*. Chicago: University of Chicago, 178-185.

Nunes, Jairo. 2004. *Linearization of Chains and Sidewards Movement*. Cambridge, Mass.: MIT Press.

Peters, Stanley & R. W. Richie. 1981. Phrase Linking Grammars. Ms. Stanford University and Palo Alto, California.

Post, Emil. 1943. Formal Reductions of the General Combinatorial Decision Problem. *American Journal of Mathematics* 65(2): 197-215.

Post, Emil. 1944. Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society* 50: 284-316.

Postal, Paul M. 1964. *Constituent Structure*. Bloomington, Indiana: University of Bloomington.

Postal, Paul M. 1974. *On Raising*. Cambridge, Mass.: MIT Press.

Rosenbaum, Peter. 1965. *The Grammar of English Predicate Complement Constructions*. PhD dissertation, MIT.

Saddy, Douglas. 1990. *Investigations into Grammatical Knowledge*. PhD dissertation, MIT.

Saddy, Douglas. 1991. Wh-scope mechanisms in Bahasa Indonesia. MITWPL #15 ed. Lisa Lai-Shen Cheng and Hamida Demirdache, 183-218.

Sampson, Geoffrey. 1975. The Single Mother Condition. *Journal of Linguistics* 11(1): 1-11.

Stepanov, Arthur. 2001. Late adjunction and Minimalist Phrase structure. *Syntax* 4(2): 94-125.

Turing, Alan M. 1952. The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society of London* B 237(641): 37-72.

von Neumann, Johann. 1923. Zur Einführung der transfiniten Zahlen. *Acta litterarum ac scientiarum Ragiae Universitatis Hungaricae Francisco-Josephinae, Sectio scientiarum mathematicarum* 1: 199-208.

Yang, Charles. 2016. *The Price of Linguistic Productivity: How children learn and break rules of language*. Cambridge, Mass.: MIT Press.