Due to Martha by 11/6/17
0 Figures
0 Tables

ANNUAL REVIEWS

Annual Review of Linguistics

The Minimalist Program After 25 Years

Norbert Hornstein

Department of Linguistics, University of Maryland, College Park, Maryland 20742; email: nhornste@umd.edu

Annu. Rev. Linguist. 2018. 4:3.1–3.17

The *Annual Review of Linguistics* is online at linguist.annualreviews.org

https://doi.org/10.1146/annurev-linguistics-011817-045452

Keywords

minimalism, Merge, generative grammar

Abstract

The Minimalist Program (MP) has been around for about 25 years, and anecdotal evidence suggests that conventional wisdom thinks it a failure. This review argues that MP has been a tremendous success and has more than met the very high goals it had set for itself. This does not imply that there is not more to be done. There is, a lot more. But the problems are those characteristic of successful and ongoing research programs. Why the perception of failure? It arises from a misunderstanding concerning the aims of the minimalist project and what, given these aims, it is reasonable to expect. Once we clear up the nature of MP's goals, we will be better placed to judge (and appreciate) how far it has come.

1. INTRODUCTION

The Minimalist Program (MP) has been around for about 25 years, and anecdotal evidence suggests that conventional wisdom thinks it a failure. I could not disagree more, and in this article I explain why. Before I start, here is the bottom line: MP has been a tremendous success and has more than met the very high goals it had set for itself. This, of course, does not imply that there is not more to be done. There is, a lot more. But the problems are those characteristic of successful and ongoing research programs. Why the perception of failure? It arises from a misunderstanding concerning the aims of the minimalist project and what, given these aims, it is reasonable to expect. Once we clear up the nature of MP's goals, we will be better placed to judge (and appreciate) how far it has come.

This review is organized as follows. First, I locate MP historically in the generative grammar (GG) tradition and thereby identify the central questions on its research agenda. Second, I outline one MP theory that builds on results over the last 25 years that, were it true, would go a long way toward redeeming MP's ambitions. Third, I argue that there is decent (although not dispositive) evidence that the story outlined has nonnegligible verisimilitude. Does this mean that it is the one right, true, and correct, theory? No. Rather, I argue that it is the right kind of theory and that there is a rational basis for thinking that it could be right. In other words, I argue that 25 years of research has shown that the MP project is viable and timely (which was not certain or even likely 25 years ago) and that it has produced novel insights that have a good chance of being correct about the nature of the faculty of language (FL). As this is really all anybody should ever expect of any research program, I conclude that MP has been wildly successful (justified applause ensuing).

2. THE GENERATIVE GRAMMAR ROAD TO THE MINIMALIST PROGRAM

Before MP, GG focused on two salient and obvious facts and their implications: (*a*) the fact of linguistic creativity (LC) [e.g., the fact that a native speaker's linguistic facility extends over an unbounded domain of sound–meaning (<s, m>) pairs] and (*b*) the fact of linguistic promiscuity (LP) [e.g., the fact that any child (absent pathology) acquires any language in (roughly) the same way as any other child does]. LC refers to the fact that a native speaker's linguistic facility is effectively unbounded. Chomsky (1964, p. 7) described it thus:

> . . . a mature native speaker can produce a new sentence of his language on the appropriate occasion, and other speakers can understand it immediately, though it is equally new to them. Most of our linguistic experience, both as speakers and hearers, is with new sentences; once we have mastered a language, the class of sentences with which we can operate fluently is so vast that for all practical purposes (and, obviously, for all theoretical purposes), we may regard it as infinite.

GG has inferred that this unbounded capacity rests (in part) on the internalization of a grammar (G):[1] a finite set of rules/procedures that recursively apply to generate the unbounded set of linguistic objects that native speakers easily use (i.e., produce and understand). In other words,

[1]I say "in part" because Chomsky describes actions that native speakers can perform (producing and understanding) and these actions rely both on a certain body of knowledge (competence) and on mechanisms for using this body of knowledge (performance). The important feature is that the knowledge is unbounded and undergirds the performance, and hence Gs are implicated.

part of the explanation of LC rests on the fact that native speakers of a given language L have as part of their mental makeup a grammar of L (i.e., G_L) that characterizes their knowledge of L.

LP is the second salient fact. It names the truism that any child C can learn any language L if placed in the appropriate linguistic environment (i.e., if exposed to samples of L). The samples of L that C [also known as the language acquisition device (LAD)] uses to acquire G_L are called the primary linguistic data (PLD_L). LP reflects the observation that, in a deep sense, a native speaker's actual language is an accidental and relatively superficial fact about him/her. Why? Because a speaker of a given language L could have spoken any other language L′ had the speaker (as child) had timely exposure to $PLD_{L'}$.

What follows from these two facts? LC combined with LP implies that humans come equipped with a meta-capacity to acquire any G if exposed to the appropriate PLD. It is a moral certainty that an endemic feature of being human is coming equipped with this meta-capacity (i.e., the FL).[2]

Two observations: First, to this point we have been dealing entirely with truisms. That native speakers have Gs is a trivial consequence of LC. That humans have FLs is a trivial consequence of LP. Again, these are truisms, and the facts that they rest on are evident and not rationally contested.

Second, we move beyond truisms when we specify the actual structure of a given G_L and the actual properties of FL that allow it to construct G_Ls on the basis of PLD_L. In doing so, we move beyond the obvious and commit hostages to facts and theories.

How does one go about finding these facts and elaborating these theories? The history of research in GG from the mid 1950s to the mid 1990s illustrates how this can be done. Early GG developed accounts of particular Gs and described the kinds of rules Gs contain. This research focused on different constructions (e.g., Relativization, Question Formation, Pronominalization, Clefting, Complementation) and described their intricate properties. It also showed how diverse rules could interact to allow for ever more elaborate linguistic structures.

Given a plausible inventory of rules prompts a further question: How does one describe the meta-capacities that allow children to build Gs in concert with PLD (i.e., a theory of FL)? These properties of FL are also known as Universal Grammar (UG).[3] Given that Gs have rules of the kind that they do, theories of UG aim to explain why Gs contain these kinds of rules and not others and why they have the properties they do and not others. Standard accounts rely on identifying locality conditions on rules (e.g., clause-mate, islands, phases), inventorying which operations must meet which conditions at which points in a grammatical derivation, and finding the properties that Gs are sensitive to (e.g., structure dependence, c-command, government). Thus, the results of research on the character of specific Gs become fodder for further research on the UG properties of FL that permit humans to develop Gs with such rules and not others.

Unsurprisingly, these two projects develop hand in hand; the description of the properties of particular Gs is particularly important in the first phase of research, and elaboration of the principles of FL rises to prominence later on. Why?

In exploring the UG properties of FL (i.e., the meta-capacity to acquire a G) it helps to have properties of actual Gs to focus on. To reiterate: That humans have Gs and the capacity to acquire them is trivial. But what the actual properties of these Gs and FL are is decidedly not trivial

[2]The standard view is that this information is part of the human genotype. But all that we can be sure of is that every human comes with this capacity and that every human can use it to acquire a G in the right PLD environment. It is consistent with this view that factors other than shared genotypes allow humans as such to share properties. All we can conclude is that having an FL is one of these shared properties.

[3]As a reviewer insisted I make clear: UG as used here denotes the principles of a system of inference that takes one from PLD to G. It is not a specific concrete G assumed to be universal. Nor, I should add, despite the widespread confusion in the literature has Chomsky ever used the term in the latter fashion. Greenberg did, but Chomsky never did. Nor do I.

and requires scientific inquiry of the standard variety. Thus, early research in GG focused on describing the kinds of rules and operations [i.e., generative procedures (GPs)] found in particular Gs and, just as importantly, not found in such Gs. The resulting plausible inventory of GPs feeds inquiry into the structure of FL, with the goal of finding those features of FL that would explain why the available GPs are in fact available and why the nonextant GPs are absent.

From the early 1980s to the mid 1990s, GG research crystalized in a specific concrete proposal, Government-Binding Theory (GB).[4] GB is a highly modular theory with different kinds of GPs defined over many different kinds of objects, many of which appear to be highly linguistically specific.

Specifically, GB identifies several different interacting modules, each with its own primitives and operations and domains. Here is a short outline of the relevant parts of a GB grammar:

1. Base rules
    a. X′ Theory
    b. Theta Theory
2. Movement rules (A and A′)
    a. Trace Theory
    b. Subjacency Theory
    c. Empty Category Principle (ECP)
3. Case rules
4. Binding rules
    a. Anaphors
    b. Pronouns
    c. R-expression
5. Control rules

There are some common relations (c-command and government being two that occur in many of the modules), but the rules and primitives for assigning case differ from those for assigning theta roles, and both differ from those that move expressions around or license anaphoric dependencies. Thus, each module comes with its own specific information concerning operations, domains of applications, and relevant relata. Importantly, this internal modularity is intrinsic to the GB model of FL, as is the linguistic specificity of the modules. In other words, the operations and primitives appear to be linguistically dedicated, reflecting a conception of FL that is strongly domain specific. Let me elaborate.

To fix ideas, let us adopt the following convention, which slightly repurposes earlier GG terminology. FL designates whatever combination of mental powers is required to build a natural language G. These powers can be domain general, computationally generic, or domain specific. Let UG designate those features of FL that are linguistically domain specific [i.e., are neither (cognitively) domain general nor computationally generic]. GB embodies the view that most of FL is UG-ish (i.e., domain specific). On the other end of the spectrum lie theories of FL in which FL (or most of FL) comprises domain-general operations and cognitively generic computational

[4]I concentrate on mainline GB research here as an embodiment of early GG investigations. This is not meant to cast aspersions on other "frameworks" like GPSG or RG or LFG. To my mind, these various approaches ended up converging on more or less the same kinds of Gs and UG principles, despite their apparent differences. In my opinion, these various frameworks turned out to embody the same basic insights in basically the same ways and thus converged on the same basic conceptions of Gs and FL.

operations and domains. An MP question is: How UG-ish is FL? MP takes this question as cynosure and proposes that although FL might be UG-ish, it is not very UG-ish. The reasoning involves two considerations concerning the evolvability of FL, and the fact of interest is that the human capacity for language appears to be quite recent and very stable. Let us call this the fact of linguistic recency (LR).

The evidence for LR is twofold; one fact is rather more evident than the other. First, human language appears to be a relatively recent biological innovation. More precisely, if one assumes that cultural artifacts are proxies for the emergence of language, then evidence of rich cultural artifacts is very recent, roughly 50–100 kya. Before about 50 kya, such language proxies are largely invisible. A common conclusion is that the efflorescence of such artifacts is tied to the emergence of language in humans.[5]

The second fact of interest is that FL arose in humans before the trek out of Africa and has remained unchanged since. In other words, whatever happened has not changed much. The evidence for this fact involves the fine detail of linguistic promiscuity (LP). From what we can tell, any human can learn any language and the course of acquisition for any human of any language is essentially the same. This suggests that human FLs are effectively the same across the species or, more pointedly, that FL has remained stable after humans went their separate ways. Unlike the first fact, which relies on linguistic proxies, this fact is quite robust. As far as we know, there is no robust genetic, ethnic, racial, or areal variation in the acquisition of any language.[6] We can all acquire the same Gs in essentially the same ways (i.e., all humans have the same FL).[7]

Both facts point to the same conclusion: that whatever prompted the emergence of FL is recent and very minor—a small, recent, simple addition to the prelinguistic cognitive apparatus. A simple or minor change could occur all at once. A simple or minor recent change would be similar across the species. In other words, if FL is composed largely of domain-general computational operations with a recent minor or simple domain-specific add-on, then the two facts above could be explained. To put this yet another way: The less UG-ish FL is, the easier it is to explain how it arose in the species so recently (fact 1) and why it has remained largely unchanged (fact 2). Thus, we can conclude that FL is only a tiny bit UG-ish; most of its operations and principles are domain general or computationally generic. The principal MP aim is to show that this line of thinking is plausible—that, properly understood, FL really can be factored into a smaller, language-specific UG part and a larger, domain-general, computationally generic FL part.

[5]This is a very weak argument. The proxies for language are cultural artifacts of increasing complexity. The assumption is that culture requires language and so the emergence of these artifacts should index language proficiency. However, we need to specify what feature of language culture supervenes on. It may be more than the capacity to entertain unboundedly complex thoughts. It is reasonable to think that the communicative aspect of language is crucial for the rapid advancement of culture, for it becomes a medium of cultural evolution (memes and so forth). But a standard MP assumption is that externalization is a late addition to a Merge-based thought system. If so, then it appears that the great leap forward is less an index for the emergence of Merge than it is for the externalization of language. If this is correct, then all we can be confident about is that Merge precedes the trek out of Africa, which follows from LP.

[6]A reviewer rightly observed that there is no evidence against the view that there is genetic, ethnic, racial, and/or areal variation. This issue has not been systematically studied. But I doubt that it is very robust, or we would have known about it, at least anecdotally. So, if I wanted to cover my posterior I could note this possibility and make noises about how it could be true and modulate my categorical assertions to appear more complaisant, but as it really doesn't affect much of what follows, and as to date there is no evidence that it is true or that the idealization to a uniform capacity has deleterious consequences, I resist this temptation, strong as it is, to be more mealy mouthed than usual.

[7]As far as we can tell, LP extends to all aspects of linguistic competence. So any child will learn the syntax and phonology (and phonetics) of any language in the same way. Therefore, humans not only share a common syntax, they also share common processes of externalization. This suggests that all aspects of FL, including the methods of externalization, were in place before the trek out of Africa and have not evolved significantly since.

Another line of reasoning points to the same conclusion. Methodologically, FLs with simpler UGs are preferable because, by assumption, they involve less cognitively idiosyncratic machinery. To the degree that cognitive architecture is shared across modules, to that degree such architecture has wider empirical support. All things being equal, therefore, such theories are methodologically preferable.

In summary, both methodological and evolutionary reasoning support a project wherein the UG features of FL are minimized, hence the MP program: Determine to what degree a more computationally generic domain-general view of FL is viable.

3. PURSUING THE MINIMALIST PROGRAM PROJECT IN TWO STEPS

Practically speaking, how does one pursue this MP agenda? The answer is that one should start with GB and simplify, simplify, simplify. In this section I describe a two-step process that does just that.

First, we eliminate the internal modularity of GB. In other words, the first step is to unify the various modules (Case, Binding, X′, Control, Movement, Bounding, ECP) and show that they are all instances of the same underlying operations and principles. This wing of the MP project has antecedents in earlier GB research; the prime example is the unification of Ross's island effects in terms of Subjacency Theory.[8] As discussed further below, this project is quite advanced conceptually, and in my opinion the empirical payoffs have been quite substantial.

Second, we show that the principles and operations of the unified theory (the results of step 1) have (very) few linguistically specific operations and principles. In other words, we show that the required operations and principles that the unification invokes are largely computationally/cognitively generic.

The minimalist project is, first of all, to demonstrate that this two-step process is conceptually possible. Are there ways of unifying the modules without paying too high an empirical price? Note that MP is here taking GB as the explanandum. Just as Subjacency Theory derived Ross's island effects, a successful MP theory will derive the various universals that that GB postulates (e.g., Binding, Subjacency Theory) to explain the effects it identifies (e.g., island effects, weak crossover effects, complex NP effects).

A caveat: Taking GB as the target of explanation does not imply that MP need believe that everything that GB claims to be true is true. Rather, MP requires that GB be a useful proxy for what is the case. It has the right kinds of properties, so showing how to minimalize GB provides a good proof of concept for MP and a good foundation for further MP inquiry. In other words, GB FL and the Gs it yields have the kinds of properties that the "true" FL has, so aiming to minimalize GB is a useful step forward in advancing the program.

There is, of course, a second step: moving beyond viability (i.e., proof of concept) to show that MP theories are empirically superior. I believe that there are a few scattered indications that this is so, although space limitations prevent me from going into detail. However, I believe it is fair to say that MP's greatest success has been at the conceptual level (again, something that I believe to be very nontrivial) and that the empirical payoffs, although interesting, have not yet been dispositive.

So, GB is the target of explanation because it identifies the kinds of properties that FL is likely to have. What are these? Here is a partial list:[9]

[8]The classic paper in this vein is that by Chomsky (1977).

[9]The list is intended to be representative, not exhaustive. For a longer list, see **https://facultyoflanguage.blogspot.com/2015/03/a-shortish-whig-history-of-gg-part-3.html**.

(6a) Hierarchical recursion
(6b) Displacement
(6c) Gs generate a natural format for semantic interpretation
(6d) Reconstruction effects
(6e) Movement targets c-commanding positions
(6f) No lowering rules
(6g) Strict cyclicity
(6h) Structure-dependent rules
(6i) Locality
(6j) Antecedents c-command their anaphors
(6k) Anaphors never c-command their antecedents
(6l) XPs move, X′s do not, X^0s might
(6m) Gs treat arguments and adjuncts differently, with the former less "constrained" than the latter
(6n) Control targets subjects of "defective" (i.e., tense or agreement deficiency) clauses
(6o) Control respects the Principle of Minimal Distance
(6p) Case and Agreement are X^0–YP dependencies
(6q) Reflexivization and Pronominalization are in complementary distribution
(6r) Selection/subcategorization relations are very local

A reasonable question is why FL has Gs with these properties. The most interesting feature of current minimalist research is that it offers theories that provide decent answers to this question. In the next section, I review some of these.

In sum, GB theory identifies plausible design features of FL, such as those listed above. MP aims to develop accounts of why FL has these design features (and not others).

Before going into detail, I note that asking such questions and trying to answer them is standard scientific practice. This is what science does: Identify some property P and ask why P, explain P in terms of Q and ask why Q, and so on. Thus, given the history of GG, MP is the next logical step for research into the structure of FL. It has been made possible by earlier GB-ish achievements that identified plausible UG principles of FL that could then serve as explananda for MP accounts. Absent plausible targets of explanation, attempts to explain them are nugatory. Given such targets, the aim of explaining them in more fundamental terms is, scientifically speaking, the obvious desired and required next step. In this sense, MP is ineluctable. Indeed, refusing to entertain such MP questions amounts to obscurantism.

4. SOME MINIMALIST THEORIES

4.1. Merge and the Gs It Delivers

Recall that the fact of LC implicates Gs able to generate unboundedly complex hierarchical structures. What is the minimum required to do so?

All GG theories have identified two essential components: (*a*) a repository of basic atoms (i.e., a lexicon) and (*b*) rules for combining these atoms into ever larger and more hierarchically complex structures (i.e., constituents). The simplest version (or at least a very simple version) of such a rule of combination would take linguistic objects and make them a unit/constituent and otherwise

leave the elements combined unchanged. MP calls this combination operation Merge and asks what kinds of objects such a very simple operation would generate.

What kind of unit does merging two elements a and b form? At the very least, Merge takes elements that are not a unit and turns them into one. The simplest rule would do nothing more than unify them. Thus, at a minimum, there must exist G rules that take linguistic expressions a and b and form them into something like sets (e.g., {a, b}), namely objects that specify nothing more than that the elements that constitute them are a unit. Thus, the only difference between a list of a and b and the set {a, b} is that the latter represents the fact that a and b are a unit.[10] The set specifies nothing else. For example, it imposes no order on the elements of the unit (i.e., {a, b} is the same object as {b, a}). Nor does it change a and b in any way beyond combining them. We have the same a after it combines with b (and ditto for b) as we had before, save for the fact that it now forms a unit with b (and b with a). Thus, sets are minimal specifications of constituency (i.e., anything that is a constituent says at least this much about its constituents).

More specifically, given the fact of LC, we know that every theory of G must contain an operation that takes syntactic objects (SOs) and combines them into larger syntactic units that are themselves combinable into larger units. And thus we know that every theory of G must contain an operation that combines smaller units into set-like objects because a set is the least encumbered specification of what being a unit is. Of course, saying this does not preclude the possibility that the required units have more structure than simple sets specify (e.g., a specification of the linear order or headedness of the constituents). But any theory will have at least this much. It is thus worth asking what, if anything, follows from this very minimal specification. Oddly—and this is an MP discovery—it turns out to be quite a lot, in the sense that postulating a simple recursive operation at the heart of FL that minimally combines elements into larger combinable set-like units explains some nontrivial properties characteristic of natural language Gs. Let us consider some details.

Merge can be recursively specified as follows:

(7a) If α is a lexical item, then α is an SO.[11]

(7b) If α is an SO and β is an SO, then Merge(α, β) is an SO.

(8) For α and β, SOs, Merge(α, β) $\rightarrow$ $\{\alpha, \beta\}$.

So specified, Merge indeed accounts for properties 6*a*–*b*, above. If Merge is an operation that forms sets as in statements 7 and 8, then we expect FL to allow for Gs that have properties 6*a*–*b*. Consider the details.

First, Merge generates an unbounded number of hierarchically structured sets (i.e., property 6*a*). Thus, given a lexicon with expressions α, β, γ, $\delta, \ldots$, we can form, using statements 7 and 8, objects 9*a*–*c* and then some (in fact, there is no limit to the "then some"):

(9a) $\{\alpha, \{\beta, \{\gamma, \delta\}\}\}$

(9b) $\{\{\alpha, \beta\}, \{\gamma, \delta\}\}$

(9c) $\{\{\{\alpha, \beta\}, \gamma\}, \delta\}$

[10]Or, to put this another way, whereas a list of a and b lists two objects (i.e., a and b), the set {a, b} comprises three objects (i.e., a, b, and the set {a, b}).

[11]The term lexical item denotes the atoms that are not themselves products of Merge. These roughly correspond to the notions of morpheme and word, although these notions are themselves terms of art and it is possible that the naïve notions only roughly correspond to the technical ones. Every theory of syntax postulates the existence of such atoms. Thus, what is debatable is not their existence but their features.

Here is a more user-friendly example. Derivation 11 specifies how to generate sentence 10 using Merge repeatedly on the lexical atoms specified. I leave it as an exercise to generate yet larger and deeper structures using the recursive Merge procedure in statements 7 and 8.

(10) The king of Georgia hugged the portrait of Lenin.

(11a) As *of* is an SO and *Lenin* is an SO, statements 7b and 8 license the construction of the SO {of, Lenin}.

(11b) The SO {of, Lenin} merges with the SO *portrait* to form the SO {portrait, {of, Lenin}}.

(11c) The SO *the* merges with the SO {portrait, {of, Lenin}} to yield the SO {the, {portrait, {of, Lenin}}}.

(11d) The SO *hugged* merges with the SO {the, {portrait, {of, Lenin}}} to yield the SO {hugged,{the, {portrait, {of, Lenin}}}}.

(11e) The SO *of* and the SO *Georgia* merge to form {of, Georgia}.

(11f) The SO *king* and the SO {of, Georgia} merge to form {king, {of, Georgia}}.

(11g) The SO *the* and the SO {king, {of, Georgia}} merge to form the SO {the, {king, {of, Georgia}}}.

(11h) The SO {the, {king, {of, Georgia}}} and the SO {hugged, {the, {portrait, {of, Lenin}}}} merge to form the SO {{the, {king, {of, Georgia}}}, {hugged, {the, {portrait, {of, Lenin}}}}}.

Second, Merge generates "movement" structures (i.e., phrase markers where the same expression occupies different positions in the structure; see property 6*b*, above). In other words, the same Merge operation that allows for the construction of arbitrarily complex hierarchical structures will also suffice to generate movement dependencies. And this is a very good thing. Why? Because 60 years of GG research has shown that human Gs are rife with movement dependencies. Typical examples are *wh*-movement and passivization in English.

Let us consider how movement structures might be formed using statements 7 and 8. Given structure 12*a*, consider how statements 7 and 8 yield structure 12*b*. Observe that in the latter, β occurs twice. This can be understood as coding a movement dependency, as β is both the sister of the SO α and the sister of the derived SO $\{\gamma, \{\lambda, \{\alpha, \beta\}\}\}$. Using statements 7 and 8 yields derivation 13:

(12a) $\{\gamma, \{\lambda, \{\alpha, \beta\}\}\}$

(12b) $\{\beta, \{\gamma, \{\lambda, \{\alpha, \beta\}\}\}\}$

(13) The SO $\{\gamma, \{\lambda, \{\alpha, \beta\}\}\}$ and the SO β (within $\{\gamma, \{\lambda, \{\alpha, \beta\}\}\}$) merge to form $\{\beta, \{\gamma, \{\lambda, \{\alpha, \beta\}\}\}\}$.

Note that this derivation assumes that once an SO, always an SO. Thus, merging α to form part of a complex SO does not change the fact α is an SO. This allows Merge to target a subpart of another SO for further merging. This is what enables Merge to generate structures with the properties of movement; it allows for the generation of structures in which a single expression can be member of two different "sets," which suffices to code movement dependencies.

Note further that the required assumption "once an SO, always an SO" follows from the assumption that Merge does nothing more than take SOs and form them into a unit. It otherwise leaves the combined objects alone. MP has dubbed this quality the No Tampering Condition (NTC). The idea is that Merge does not affect the properties of the combining elements in any

way. So, if α has some property before being combined with β (e.g., being an SO), it will have this property after it is combined with β.[12]

Unifying movement and phrase building is an MP innovation. Prior theories of grammar (and early minimalist theories) treated phrasal dependencies and movement dependencies as the products of entirely different kinds of rules (e.g., phrase structure rules versus transformations/Merge versus copy+Merge). Merge unifies these two kinds of dependencies and treats them as different outputs of a single operation. As such, the fact that FL yields Gs that contain both unbounded hierarchy and displacement operations is unsurprising.

Third, Merge generates representations sufficient for coding semantically relevant relations (property 6*c*). In particular, we can code both thematic and scope relations using Merge. To see this, let us distinguish two instances of the Merge operation. We refer to instances in which the mergees are distinct, neither contained in the other, as E-merges. Derivation 11 involves a series of E-merges, which serve to code theta role assignment. Thus, in example 11*d*, above, E-merging *hugged* and {*the*, {*portrait*, {*of*, *Lenin*}}} to yield {*hugged*, {*the*, {*portrait*, {*of*, *Lenin*}}}} enables *hug* to assign the internal theta role to {*the*, {*portrait*, {*of*, *Lenin*}}}. Thus, the thematic dependency between the predicate and its internal argument is syntactically established under E-merge.

The other instance of Merge, I-merge, involves merging SOs one of which is contained in the other. I-merge results in displacement structures that are able to support the scope properties of sentences. This is well known in the case of *wh*-movement in English, where I-merging a *wh*- with the clause containing it provides a logical syntax appropriate for a quantificational semantics of the relevant sort. For example, if sentence 14*a* has phrase marker 14*b* derived via an application of I-merge to the *what* and the containing clause *you eat*, then we end with a format easily mapped to interpretation 14*c*. Note that the two copies occupy the scope and variable positions involved in interpreting a quantifier:

(14a) What did you eat?

(14b) {what, {you, {eat, what}}}

(14c) For which thing x, you ate x.

In sum, the same simple Merge operation that can build unbounded hierarchical structures can support displacement, and each of the two instances of this Merge operation can serve to structurally code the two basic kinds of information relevant for semantic interpretation, namely argument structure (via E-merge) and scope (via I-merge).

Coding these basic dependencies via Merge suggests a very strong principle: All grammatical dependencies are established under Merge. So, for α and β to be grammatically linked, α and β must have merged. To the degree that this principle can be maintained, we can trace G dependencies to a single generative procedure and date the emergence of FL with the arrival of Merge. It also suggests a program: Show that all G dependencies (all those GB described in its various modules) are actually mediated by Merge. I return to this topic in the next section.

Fourth, the Merge-based conception of movement embodies a version of the copy theory of movement. Thus, in the above example, I-merging *what* results in a structure with an instance of *what* in two positions in phrase marker 14*b*. This contrasts with earlier theories in which the lower *what* GB had traces (i.e., categories without contents; [$_a$ e]) in place of copies. However, we know

[12]The NTC has two codicils: One forbids adding information to inputs, and the other pertains to losing information about the inputs. The Inclusiveness Condition (IC) says that expressions cannot be enriched in the course of the derivation. So, for example, indices cannot be "added" derivationally. The Extension Condition (EC) says that the structure of inputs must be preserved in the output. So, for example, operations cannot delete information concerning the inputs in the output. The NTC combines the IC and the EC.

that something like copies is what we want. How do we know this? Because natural languages are rife with reconstruction effects, cases in which displaced contents act as though they are in their base positions for purposes of interpretation.[13]

Consider example 15*a*, where the pronoun *she* must be disjoint in reference with *Sheila*. The Merge-generated phrase marker for example 15*a* is 15*b*:[14]

(15a) *Which book about Sheila$_1$ did she$_1$ enjoy?

(15b) {{which, {book, {about, Sheila}}}, {{she, {enjoy, {{which, {book, {about, Sheila}}}}}}}

Here, I-merging the *wh*-phrase to the matrix results in a copy of the *wh*-phrase in the base position. This fact can be (and has been) used to explain why example 15*a* is unacceptable with the indicated coindexing (the principle C effect). The point here is that given Merge, we obtain the copy theory, and given the copy theory, we expect reconstruction effects.[15] So, not only do I-merge-generated phrase markers support operator/variable readings but also the way they do so implicates reconstruction effects, so their crosslinguistic ubiquity is unsurprising.

We also get properties 6*e*–*g* as consequences. Recall that Merge embodies the NTC. Thus, Merge does no more than create linguistic units. In particular, it does not change the inputs in any other way. Whatever is a constituent in the input appears as a constituent with the same properties in the output. This implies that all instances of I-merge are to a c-commanding position and that lowering rules cannot exist.[16] The condition that movement always be upward and to a c-commanding position thus follows trivially from this simple specification of Merge (i.e., Merge with the NTC).

For example, the NTC prohibits the derivation structure 16*b* from structure 16*a*. Here, γ is merged with α. The output of this instance of Merge obliterates the fact that $\{\alpha, \beta\}$ had been a unit/constituent of structure 16*a*, the input to Merge. This is what the NTC prohibits. As NTC implies that Merge always target the "root," structure 16*b* is not a licit instance of I-merge (note that $\{\alpha, \beta\}$ is not a unit in the output); nor is structure 16*c* (note that $\{\{\alpha, \beta\}, \{\gamma, \delta\}\}$ is not a unit in the output); nor is a derivation that violates the strict cycle, as in structure 16*d*. Only structure 16*e* is grammatically acceptable because all the inputs to the derivation (i.e., γ and $\{\{\alpha, \beta\}, \{\gamma, \delta\}\}$) are also units in the output of the derivation. A new relation has been added, but no previous ones have been destroyed (again, once an SO, always an SO). In a derivation of cases 16*b* and 16*c*, one of the inputs (e.g., $\{\{\alpha, \beta\}, \{\gamma, \delta\}\}$) is no longer a unit in the output, so the NTC has been violated:

(16a) $\{\{\alpha, \beta\}, \{\gamma, \delta\}\}$

(16b) $\{\{\{\gamma, \alpha\}, \beta\}, \{\gamma, \delta\}\}$

(16c) $\{\{\gamma, \{\alpha, \beta\}\}, \{\gamma, \delta\}\}$

(16d) $\{\{\alpha, \beta\}, \{\delta, \{\gamma, \delta\}\}$

(16e) $\{\gamma, \{\{\alpha, \beta\}, \{\gamma, \delta\}\}\}$

[13] We also prefer copies to traces methodologically: All things being equal, we should resist theory internal constructs. Traces are theory internal objects (products of movement) that come with a slew of idiosyncratic licensing conditions (think ECP). Like all theoretical constructs, traces must earn their keep empirically. In other words, *ceteris paribus*, a theory of movement without traces is preferred to one with them.

[14] I abstract here from T to C movement.

[15] Note again that "copy" here means multiple occurrences of the same expression. Note, too, that there is no copy operation. "Copy" here describes the outputs of Merge. It is not a G operation (as in the earliest minimalist proposals).

[16] The first conjunct holds only if there is no interarboreal/sideward movement. For the nonce, I assume this to be correct.

Finally, this Merge-based conception also precludes non-structure-dependent operations (i.e., property 6*b*). How so? Recall that the products of Merge are sets and that sets impose a hierarchical organization only on the expressions merged. Importantly, they impose no linear requirements on the elements. If we understand a derivation to be a mapping of phrase markers into phrase markers and we understand phrase markers to effectively be sets (i.e., to only specify hierarchical relations), then it is no surprise that rules that leverage linear left–right properties of a string cannot be exploited. They do not exist, because phrase markers eschew this sort of information, so they cannot be defined. So, why are rules of G structure dependent? Because this is the only structural information that Merge-based Gs represent.

In summary, a Merge-based conception of grammatical operations has a long explanatory reach. On the basis of this very simple conception, it is possible to derive several (eight!) key properties that GG has argued characterize FL and the Gs it allows. This is a substantive result, and quite contrary to the notion that MP is a failure.

4.2. Extending the Logic to Unifying the Modules

GB is a strongly modular theory. In other words, there are many different kinds of rules involving different operations, locality conditions, and relata. If the aim of MP is to show how something like FL could have arisen quickly and remained stable, and we take GB to have provided a roughly accurate description of the kind of dependencies FL regulates, then MP's goal is to show why the modules have the properties they have, and to do so without multiplying syntactic operations. In fact, the ultimate aim would be to show how to incorporate the various modular operations using Merge alone. Why? Because, as discussed above, Merge has some very nice properties; in particular, it allows for a unification of phrase building and movement. Furthermore, given its simplicity, something like Merge must be part of any conceivable conception of grammar. Thus, to the degree that we can show that all GB dependencies are Merge-mediated dependencies, we can explain why FL has the properties it has.

This suggests an obvious project: Show how to unify construal rules with movement operations (i.e., reduce construal to I-merge). Before proceeding with this discussion, I note two things. First, what is presented below enjoys a sparser consensus than what I have reviewed above. So, although I believe that it builds on regulative minimalist ideals, the proposal does not enjoy universal approbation within MP. Second, due to space limitations, what I provide is merely a sketch and is at best suggestive. The aim is to illustrate that unifying all nonlocal dependencies with movement (and thereby with Merge) is feasible and has some empirical advantages. But details will unavoidably be sparse. Moreover, the discussion concentrates largely on those modules dependencies that can be unified with A-chains (e.g., Case, Obligatory Control, and A-anaphors). As the reader knows, there is more to movement than A-chains.

4.2.1. Case and movement. Case was the first module unified with Movement.[17] In GB, a special class of heads (i.e., finite T, V, and P) assign case to a special class of phrases (i.e., nominals). The configuration of case assignment is government. The earliest minimalist papers (see footnote 16) reanalyzed the GB theory and argued that Case is effectively an A-chain dependency between a case-assigning head and a case-needy nominal.

More specifically, early MP approaches took the configuration for nominative case, rather than accusative case, to be the basic case-licensing configuration. At the time, the conventional

[17] This was first proposed by Chomsky (1995, pp. 167–218).

wisdom was that nominative case is assigned in the specifier position (spec) of finite T. Given a standard theory of labeling, a nominal has its case licensed when it (I-)merges with (the label of) a case-assigning head (example 17, below). Given the predicate-internal subject hypothesis, Spec T is always a derived position for a nominal that occupies it. Importantly, in this approach, case and theta requirements are licensed in different configurations by different heads. This contrasts with GB accounts in which both the internal theta role and accusative case can be assigned by the same V in the same grammatical configuration. If one carries the technology of nominative case assignment over to accusative case, then V is not the case licenser of an accusative-marked nominal. Rather, there is some case head outside the thematic domain of V that serves to license accusative case when a nominal merges with a spec of that licensing head.[18]

The most interesting empirical consequence of this A-movement theory of case is that it implies that case, in particular accusative case, potentially affects scope. We know this to be true of the movement that subserves nominative case licensing (e.g., *He seems to himself to like Mary* versus **It seems to himself that Fred likes Mary*), and if accusative case is structurally analogous to nominative case, then we expect the same to be true for accusative case. In particular, we expect accusative case to be available if movement to the relevant spec is licensed and we expect this movement to widen the scope of the accusative-marked nominal. Consider an illustration of each claim.

If accusative case is structurally analogous to nominative case, then the head that case-licenses it is distinct from the head that assigns it its theta role. For concreteness, we can consider the accusative licensing head to sit outside of VP, where "objects" are typically assigned their theta roles. Thus, we obtain a configuration like that in example 17*a* with a movement of the nominal DP as in example 17*b*. This nominal might be the thematic object, the indirect object in a double-object construction. or an embedded subject in an exceptional case marking (ECM) construction. In all instances, the DP moves to the VP-external case position for licensing, in analogy with nominative case:

(17a) $[_{\text{case}}\ \text{Case}^0\ [_{\text{VP}} \ldots \text{V} \ldots \text{DP} \ldots$

(17b) $[_{\text{case}}\ \text{DP}\ [_{\text{case}}\ \text{Case}^0\ [_{\text{VP}} \ldots \text{V} \ldots \text{DP} \ldots$

If this assumption is roughly correct, then, first, we can expect accusative licensing to track the possibility of A-movement. And to a first approximation, it does. Consider the following paradigm:[19]

(18a) John believes him to be tall.

(18b) *John believes him is tall.

(18c) John was believed t to be tall.

(18d) *John was believed t is tall.

Just as A-movement/raising from the subject position of a finite clause is prohibited, so too is accusative case. Why? Because accusative case requires movement of *him* in example 18*b* to the case head that sits above *believe*, and this kind of movement is prohibited, as example 18*d* illustrates.

[18]Given something like bare phrase structure (see Chomsky 1995, chapter 4), the label of a phrase is simply the head of that phrase, so merging with a labeled projection of a head is, in effect, merging with the head itself.

[19]Examples 18*a–d* are data, not structures. "t" is inserted here as a friendly reminder to the reader that *John* has moved from this position. A reviewer has observed that given the copy theory of movement (which, recall, follows from Merge) MP requires operations that "delete" some copies and retain others. There is a small industry devoted to specifying how this should be done. My favorite is Nunes (2004).

Second, we expect that accusative case will widen the scope domain of the accusative nominal. This is also attested.[20] Examples 19*a* and 19*b* indicate that it is possible for an embedded ECM subject to bind an anaphor in a matrix adjunct whereas a nominative embedded subject cannot:

(19a) The lawyer proved [the men_1 to be guilty] during each $other_1$'s trials.

(19b) The lawyer proved [the men_1 were guilty] during each $other_1$'s trials.

Example 19*a* supports the sensible reading in which the *during* phrase modifies the matrix predicate *proved*. This reading is unavailable in example 19*b*, where the only reading is the silly one in which the *during* phrase modifies *were guilty*. This is expected if licensing accusative case on the ECM subject requires moving it to the edge of the higher matrix VP (as in example 17*b*). In contrast, licensing nominative leaves *the men* in the embedded spec T and hence leaves the matrix *during* phrase outside its c-command domain, prohibiting binding of the reciprocal.

Finally, if case dependencies are mediated via (A-)movement, then we expect case licensers to always c-command their case licensees. This follows if case is licensed under Merge, case licensing is a nonlocal dependency (i.e., it is mediated by I-merge, that is, Movement), and Merge is subject to the NTC. These assumptions are evident for nominative case. That nominative case assigners c-command their nominal assignees follows from the predicate-internal subject hypothesis. The revamped theory of accusative case in example 17 fits accusative case to the same mold. If case is a nonlocal dependency mediated by I-merge, then this is what we should expect to find, as is the dependency between scope and case value that we seem to find.

The confluence of Case and Binding provides evidence that case dependencies involve A-movement. This, in turn, supports unifying the Case and Movement modules, thereby reducing the internal modularity of FL.

4.2.2. Control and movement. Let us now consider unifying construal and movement. Before going into detail, I note that one nice property of construal dependencies follows at once if they are formed by movement; properties 6*j* and 6*k* are necessary features of construal dependencies. How so? Unifying construal and movement treats antecedence as a chain relation. The head "binds" links that are lower down. Thus analyzed, binders must c-command their bindees, given that I-merge is subject to Extension.[21] Thus, the c-command condition on anaphoric dependencies (i.e., properties 6*j* and 6*k*) follows.[22]

Before discussing how the unification operates with respect to (Obligatory) Control,[23] let us review some history. Until MP, GG had sharply distinguished movement from construal (e.g., Raising from Control or Passive from Reflexivization). The distinction rested on identifying which positions are filled in underlying [i.e., D(eep)-structure] Structure (DS) and which are filled transformationally. The basic idea is crystallized in the GB slogan that DS is the "pure representations of GF-θ," meaning that all and only theta positions are occupied in DS. This conception of DS, coupled with the requirement that DS feed the transformational component (i.e., no DS operations follow transformational ones) and the observation that construal dependencies involve multiple thematically marked nominals, results in a system in which construal dependencies cannot be

[20] Data are from Lasnik & Saito (1991) following original observations by Paul Postal.

[21] This abstracts away from possible morphophonological spell-out processes.

[22] Note that this conception also allows us to dispense with c-command as a primitive of FL. It is a descriptive useful notion, but not fundamental. c-Command is simply the relation that arises in a Merge-based conception of structure building and displacement coupled with the assumption that anaphoric dependencies are chains.

[23] GG has long distinguished Obligatory from Nonobligatory Control. For discussion and a review of the history, see Boeckx et al. (2010).

unified with movement dependencies (i.e., a system in which we need both movement rules and construal rules).

MP's elimination of DS as a level deflects this line of reasoning. If DS does not exist and structure building and transformational rules can interleave, then there is no reason, in principle, why construal dependencies cannot be treated as a species of movement (i.e., I-merge). Note that the elimination of DS is a central (and venerable) minimalist tenet (Chomsky 1995, pp. 167–218). Thus, unifying construal and movement builds on one of the deeper MP intuitions concerning the structure of FL. Can this unification be effected, and if so, is it empirically viable? Here are a few reasons to believe that the answer to both questions is yes.[24]

If Control is a species of A-movement [i.e., is generated by the same operation (I-merge) that generates Passive and Raising constructions], then we would expect PROs to have the properties of A-traces.[25] There is evidence that they do. First, their distribution is similar. Both Raising and Control are licit only from the subject position of nonfinite clauses:

(20a) John seems t to like Mary.

(20b) *John seems t will like Mary.

(20c) John expects PRO to like Mary.

(20d) *John expects PRO will like Mary.

Second, in languages such as Brazilian Portuguese that allow Raising from (apparently) finite embedded clauses like example 20*b*, as in example 22, Control is also possible in cases like example 20*d*, as in example 21 (where *ec* stands for empty category). This is what we expect if the same operation (I-merge) generates both Raising and Control constructions. Whatever prohibits the application if I-merge in one should prohibit it in the other:[26]

(21)	[O Pedro]$_1$	disse	[que	[[o	irmão	d[o	João$_2$]]$_3$	estava	achando
	the Pedro	said	that	the	brother	of	João	was	thinking
	[que	$ec_{*1/*2/*3/*4}$	deveria	ganhar	medalha]]].				
	that		should	receive	a medal				

"Pedro said that João's brother$_1$ was thinking that he$_1$ should receive a medal."

(22)	[Os estudiantes]$_1$	parecem	que	ec_1	vijaram	mais	cedo.
	the students	seem-3pl	that		traveled-3pl	more	early

"The students seem to have traveled earlier."

Third, as example 21 indicates, the locality condition that a raised expression has to its trace also characterizes that between a controller and the empty category it controls. So, for example, I-merge is always to a c-commanding position. Thus, in example 21 only *o irmão do João* is a licit antecedent for the embedded empty category because it alone locally c-commands this position. These locality restrictions parallel those typical of A-movement, and we expect to see them regulate control dependencies if control is a species of A-movement. And if A-movement

[24]For a fuller defense, see Boeckx et al. (2010).

[25]Note: this does not mean that Control is Raising. It means that the operations that generate Control configurations are A-movements (i.e., I-merges).

[26]For discussion of Finite Control and why Brazilian Portuguese allows it and Raising from finite subjects, see Boeckx et al. (2010, chapter 4).

is simply an instance of I-merge, then the c-command restriction that explains the antecedence possibilities in example 21 follows from the NTC.

Fourth, consider one last parallel between Raising and Control. Both tolerate *wanna* contraction, in contrast to *wh*-traces. Thus, if we assume that both Control and Raising are the products of A-movement (i.e., I-merge), then we can explain why if one allows contraction, the other does too:[27]

(23a) They want PRO to kiss Mary. → They wanna kiss Mary.

(23b) They used t to live in the attic. → They usta live in the attic.

(23c) Who do they want t to vanish from the party? → *Who do they wanna vanish from the party

In summary, the distribution, licensing conditions, and phonological properties of A-traces and (Obligatory Control) PRO are all of a piece. This trivially follows if Control, like Raising, is a species of A-movement generated by I-merge.

4.2.3. A-anaphors. Reflexive dependencies also have many of the signature properties of A-chains, as explicitly noted by Chomsky (1982) and accounted for in GB accounts by classifying A-traces as anaphors subject to principle A of Binding Theory. For this reason, A-traces and reflexives are (il)licit in the same configurations:

(24a) *John seems [t is intelligent].

(24b) *John believes [himself is intelligent].

(24c) John seems [to be intelligent].

(24d) John believes [himself to be intelligent].

(24e) *John seems it was told t that Sue is intelligent.

(24f) *John wants Mary to tell himself that Sue is intelligent.

These data follow directly if reflexives "live on" A-chains. Given standard assumptions concerning I-merge, this could be theoretically accommodated if copies can convert to reflexive in certain configurations:[28]

(25) [John believes [John (→ himself) to be intelligent]].[29]

5. CONCLUSION

MP is now about 25 years old. The conventional wisdom seems to be that it has failed. I have argued that it has not. The dissatisfaction with MP, I believe, comes from misunderstanding the empirical and theoretical point of the project. MP does not primarily aim to explain the properties of particular constructions or language particular Gs; rather, it aims to explain the UG properties of FL. The MP question is: Why does FL have the kinds of properties noted in

[27]Note that this parallel behavior is not expected on any other extant theory of control/raising. If PRO is not a residue of A-movement, then why it acts like one in this rather subtle regard is left mysterious.

[28]This partly resurrects the old Lees–Klima theory of Reflexivization, but without many of the problems. For discussion, see Lees & Klima (1963), Lidz & Idsardi (1997), and Hornstein (2001).

[29]Interestingly, there exist copy reflexive languages in which Reflexivization takes the form of repeating a nominal copy. Thus, *John likes himself* surfaces as *John likes John*. This observation supports the old Lees–Klima suggestion that reflexives morphemes are simply the outer morphological clothing of an underlying copy (Lees & Klima 1963).

list 6, above? This project relies on the premise that earlier GG research got things (roughly) right.

As I have argued in Section 3, above, over the last 25 years, MP has provided some reasonable answers to this question. The answers rest on two kinds of claims: that something like Merge is the basic computational operation of FL and that all grammatical dependencies are mediated by Merge. The aim has been to illustrate the explanatory mileage that we can get out of these assumptions. The careful reader will have noted that the above has not explained all of the properties in list 6, and the educated linguist will have noted that this list does not exhaust the kinds of properties that GG researchers have reason to believe characterize FL. What follows from these gaps? In my opinion, which I have tried to support here, the right conclusion is that there is still lots of work left to do and that this should build on the evident success of the Merge-based approach.

As Chomsky has been fond of saying, Minimalism is not a theory but a program. Programs are not right or wrong but rather fecund or sterile. The relevant questions to ask in evaluating MP research, then, are whether the project asks interesting questions and whether it has been successful in addressing them. I have argued that the questions are indeed interesting and follow seamlessly from the basic GG project outlined more than 60 years ago. I have also argued that substantive line of minimalist theory based on Merge and aimed at unifying the GB modules has produced real insights (insights that even have decent empirical support). If this be failure, then let's hope there is much more to come.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Boeckx C, Hornstein N, Nunes J. 2010. *Control as Movement*. Cambridge, UK: Cambridge Univ. Press

Chomsky N. 1964. *Current Issues in Linguistic Theory*. The Hague: Mouton

Chomsky N. 1977. On *wh*-movement. In *Formal Syntax*, ed. PW Culicover, T Wasow, A Akmajian, pp. 71–132. New York: Academic

Chomsky N. 1982. *Lectures on Government and Binding*. Dordrecht, Neth.: Foris

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

Hornstein N. 2001. *Move! A Minimalist Theory of Construal*. Malden, MA/Oxford, UK: Blackwell

Lidz J, Idsardi W. 1997. Chains and phono-logical form. *Univ. Pa. Work. Pap. Linguist.* 8:109–25

Lasnik H, Saito M. 1991. On the subject of infinitives. In *Papers from the 27th Regional Meeting of Chicago, Linguistic Society 1991*, Part 1: *The General Session*, ed. LK Dobrin, L Nichols, RM Rodriguez, pp. 324–43. Chicago: Chicago Linguist. Soc.

Lees R, Klima E. 1963. Rules for English pronominalization. *Language* 39:17–28

Nunes J. 2004. *Linearization of Chains and Sideward Movement*. Cambridge, MA: MIT Press