

Chapter 1: The Minimalist Project

0. The Point of this Book

This book is an introduction to the art of minimalist analysis. What we mean by this is that it aspires to help those with an interest in minimalism to be able to “do” it. Partly, this involves becoming acquainted with the technology that is part and parcel of any specialized approach. Partly, it will involve absorbing the background assumptions that drive various aspects of the enterprise. However, in contrast to many earlier approaches to grammar, we believe that “doing minimalism” also involves developing an evaluative/aesthetic sense of what constitutes an interesting problem or analysis. In short, a good part of minimalism, in our view, is learning to appreciate what is worth doing- identifying the issues that grammatical theory in the late 1990s should address- and this is not a skill that one typically expects a text to impart. So, before we begin with the nitty gritty nuts and bolts of diverse minimalist analyses, we intend to spend some time trying to explain what we take the minimalist project to be and why its ambitions have come to prominence at this time.

1. Some Background

Since the beginning, the central task of generative grammar has been to explain how it is that children are able to acquire grammatical competence despite the impoverished nature of the data that is input to this process. How children are able to do this, dubbed “Plato’s Problem” (Chomsky 1986), can, in retrospect, be seen as driving research in modern generative linguistics since its beginnings in the mid 1950s.

The problem can be characterized abstractly as follows. Mature native speakers of a natural language have internalized a set of rules, a grammar, that is able to generate an unbounded number of grammatical structures. This process of grammar acquisition is clearly influenced by the linguistic data that the native speaker was exposed to as a child. It is patently obvious to the most casual observer that there is strong relation between growing up in Belo Horizonte and speaking (a variety of) Brazilian Portuguese and growing up in Montreal and speaking (a variety of) English. However, slightly less casual inspection also reveals that the grammatical information that can be gleaned from the data that the child has access to, the Primary Linguistic Data (PLD), is insufficient to explain the details of the linguistic competence that the mature native speaker manifests. In other words, the complexity of the attained capacity, the speakers grammatical competence, vastly exceeds that of the input to the process, the linguistic information that the child absorbs from the environment in which it grows up.

To bridge this gap, generative grammarians have postulated that children come biologically equipped with an innate dedicated capacity to acquire language -are born with a Language Faculty. The last 40 years of research can be seen as trying to provide a description of this faculty that responds to two salient facts about human natural language; its apparent surface diversity and the ease with which it is typically acquired despite the above noted poverty of the linguistic stimulus. In the last decade or so the following consensus description of the language faculty has emerged which, it is generally believed, adequately addresses these twin properties.

Kids come biologically equipped with a set of principles for constructing grammars - principles of “Universal Grammar” (UG). These principles can be thought of as functions that take PLD as input and deliver particular grammars (e.g. of English, Brazilian Portuguese etc) as

output. Thus, the principles of UG can be thought of as general conditions on grammars with open parameters that can be set in a variety of ways. Specific grammars arise once values are assigned to these parameters. Acquiring a natural language thus amounts to assigning values to these open parameters, i.e. “setting” these parameters. It is assumed that children set parameter values on the basis of the primary linguistic data that they have access to in their linguistic environments.

Observe two important features of this proposal. First, on this view, the acquisition process is sensitive to the details of the linguistic/environmental input as it is the PLD that provides the information on the basis of which parameter values are fixed. Second, the shape of the knowledge attained is not restricted to whatever information can be garnered from the PLD as the latter exercises its influence against a rich backdrop of fixed general principles that UG makes available.

Observe further that each characteristic of this model responds to one of the two basic features noted above. The fact that particular grammars are the result of setting parameter values in response to properties of the PLD allows for considerable diversity among NLS. This is especially so if the principles of UG have a tight deductive structure for in this case even a small change in the value of a particular parameter can have considerable ramifications for the structure of any particular output grammar. Second, on this view, learning is restricted to setting parameter values and so the fine details of a native speaker's linguistic competence is not limited to whatever structural information can be gleaned from the PLD. A speaker's linguistic capacities are a joint function of the environmental input AND the principles of UG. These latter can be quite complex. However, by assumption, they need not be learned as they form part of the innately endowed language faculty. Thus a native speaker's manifest competence is not restricted to the information born by the relatively impoverished linguistic input.

This picture of the structure of the language faculty has been dubbed the “Principles and Parameters” (P&P) theory. To repeat, it now constitutes the consensus view of how best to describe the overall structure of the language faculty. The minimalist program adopts this consensus view. It, thus, incorporates the claim that the initial state of the language faculty should be characterised as a set of general principles with open parameters and that specific grammars can be identified with a set of values for these parameters, values set by the primary linguistic data.

This is a very substantive assumption. It's reasonableness stems from the fact that, to date, only Principles and Parameters (P&P) architectures have provided a way of addressing Plato's problem in the domain of language. Minimalists accept that any adequate theory of UG must be of the P&P variety. In effect, minimalism assumes that having a P&P architecture is a boundary condition on any adequate characterization of UG.

Adopting this assumption has one particularly noteworthy consequence. It changes both the sorts of questions it is worthwhile focusing on and the principles in terms of which competing proposals should be evaluated. Let us explain.

As in any other domain, proposals in linguistics are evaluated along several dimensions: naturalness, parsimony, simplicity, elegance, and explanatoriness. Though all these measures are always in play, in practice, some dominate others during particular periods. In retrospect, it is fair to say that explanatory adequacy, i.e. the ability to cast some light on the linguistic version of Plato's problem, carried the greatest weight when alternatives were considered. In effect, given the centrality of Plato's problem, research in the last 15 years or so has focused on finding grammatical constraints of the right sort. By “right sort” we mean tight enough to permit grammars to be acquired on the basis of PLD yet flexible enough to allow for the observed variation across natural languages. In short, finding a suitable answer to Plato's problem has been the primary research engine and proposals have been primarily evaluated in terms of

whether they satisfactorily met its demands. This does not mean to say that other methodological standards have been irrelevant. Simplicity, parsimony, naturalness etc. have also played a role. However, as a practical matter, these benchmarks of theory evaluation have been rather weak and have been swamped by the need to develop accounts with a reasonable stab at addressing Plato's problem.

The consensus that P&P style theories offer a solution to Plato's problem necessarily effects how one will rank competing proposals from here on in. Put baldly, if P&P theories are assumed to solve Plato's problem the issue now becomes which of the conceivable P&P models is best and this question is resolved using conventional criteria of theory evaluation. In other words, once explanatory adequacy is bracketed, as happens when only accounts that have P&P architectures are considered, an opening is created for simplicity, elegance and naturalness to emerge from the long shadow cast by Plato's Problem and become the critical measures of theoretical adequacy. The minimalist program aims to consider the properties of UG in light of these evaluative criteria. But, this is no easy task. To be able to move in this direction, minimalism must address how to concretize these evaluative notions -simplicity, naturalness, elegance, parsimony- in the research setting that currently obtains. Put another way, the problem is to find a way of taking the platitudes that simpler, more elegant, more natural theories are best and giving them some empirical bite.

To recap. Once P&P theories are adopted as boundary conditions on theoretical adequacy, the benchmarks of evaluation shift to more conventional criteria such as elegance, parsimony etc. The research problem then becomes figuring out how to interpret these general evaluative measures in the particular domain of linguistic research. As we concentrate on syntax in what follows, one important item on the minimalist agenda is to find ways of understanding what constitutes a more or less natural, more or less parsimonious, more or less elegant etc syntactic account. Note that there is little reason to believe that there is only one way (or even just a small number of ways) of putting linguistic flesh on these methodological bones. There may be many alternative ways of empirically realizing these notions. If so, there will be no unique minimalist approach but a family of minimalist programs each animated by similar general concerns but developing accounts that respond to different specific criteria of evaluation or even to different weightings of the same criteria.

It would be very exciting if minimalism did in fact promote a research environment in which various alternative equally minimalist yet substantially different theories of grammar thrived as it would then be possible to play these alternatives off against one other to the undoubted benefit of each. This possibility is worth emphasizing for it highlights an important feature of minimalism. Minimalism is not a theory so much as a program for research. The program is successful just in case trying to work out its leading ideas leads to the development of interesting analyses and theory. In this sense there is no unique minimalist theory though there may be a family of approaches that gain inspiration from similar sources. Theories are true or false. Programs are fecund or sterile. Minimalism aims to see whether it is possible to interpret the general methodological benchmarks of theory evaluation in the particular setting of current syntactic research in ways that lead to fruitful and interesting theories. The current problem, however, is not to choose among competing implementations of these methodological yardsticks but to develop even a single non-trivial variant.

One last point, there is no a priori reason to think that approaching grammatical issues in this way is guaranteed of success. It is possible that the language faculty is just "ugly", "inelegant", "profligate", "unnatural" and massively redundant. If so, the minimalist project will fail. However, one cannot know if this is so before one tries.

2. Some Economy Notions

The question before us now is how implement notions like elegance, beauty, parsimony, naturalness etc in the current linguistic context. One way into this question is to recruit those facts about language, the “big facts” that any theory worthy of consideration must address. We can then place these as further boundary conditions on theoretical adequacy. We already have one such big fact, viz. that the theory have a P&P architecture.

Other “obvious” features of language and linguistic competence afford additional boundary conditions to structure a minimalist inquiry into the properties of UG. Here are six others.

- (1) Sentences are basic linguistic units.
- (2) Sentences are pairings of sound and meaning.
- (3) There is no upper bound to the number of sentences in any given natural language.
- (4) Sentences show displacement properties in the sense that expressions pronounced in one position are interpreted in another.
- (5) Sentences are composed of smaller expressions (words or morphemes).
- (6) These smaller units are composed into units with hierarchical structure, i.e. phrases, larger than words and smaller than sentences.

These properties are completely uncontroversial. They are properties that students of grammar have long observed characterize natural language. Moreover, as we will see, these facts suggest a variety of minimalist projects when coupled with the following two economy conditions.

The first type of economy considerations are the familiar methodological “Ocham’s Razor” sort that relate to theoretical parsimony and simplicity: all things being equal, two primitive relations are worse than one, three theoretical entities are better than four, four modules is better than five. In short, more is worse, fewer is better. Let’s call these types of considerations principles of ‘methodological economy’.

There is a second set of minimalist measures. Let’s dub these principles of ‘substantive economy.’ Here a premium is placed on least effort notions as natural sources for grammatical principles. The idea is that locality conditions and well formedness filters are reflections of the fact that grammars are organized frugally to maximize resources. Short steps preclude long strides, derivations where fewer rules apply are preferred to those where more do, movement only applies when it must, no expressions occur idly in grammatical representations (i.e. full interpretation holds). These substantive economy notions generalize themes that have consistently arisen in grammatical research. Just think of the A-over-A condition (Chomsky 19xx), the Principle of Minimal Distance (Rosenbaum 196x), the Superiority Condition (Chomsky 1973), the Minimality Condition (Rizzi 1990) and the Minimal Binding Requirement (Aoun and Li 1993). It is natural to reconceptualize these in least effort terms. Minimalism proposes to conceptually unify all grammatical operations along these lines.

The two economy notions coupled with the seven “big facts” listed above promote a specific research strategy: look for the simplest theory whose operations have a least effort flavor and that accommodates the big facts noted above. This proposal actually has considerable weight. Consider some illustrative examples of how they interact to suggest various minimalist projects.

The fact that the number of sentences in any given natural language are essentially infinite (fact (3)) implies that grammars exist, i.e. rules that can apply again and again to yield an unbounded number of different structures. The fact that sentences have both sound and meaning properties implies that the outputs of grammars “interface” with systems responsible for the

articulatory and phonetic (AP) features of a sentence and those that provide a conceptual and intentional (CI) interpretation for these objects. More specifically, if one is considering a theory with levels, e.g. a GB-style theory, this implies that there must exist grammatical levels that interface with the cognitive systems responsible for AP and CI properties. In effect, the levels LF and PF must exist if any levels exist at all. In this sense, their existence is conceptually necessary. Given this, methodological economy states that there is a premium on grammatical theories that can make do with these two levels alone. In other words, one minimalist project would be to show that levels other than LF and PF can be dispensed with.

In the context of a GB style theory for example, this would amount to showing that D-structure and S-structure are in principle eliminable without any significant empirical loss. This in turn requires reconsidering (and possibly reanalyzing) the evidence for these levels. For example, in GB style theories recursion is a defining characteristic of DS. Given that a mechanism for recursion must be part of any grammar given (3), if DS is eliminated, this requires rethinking how recursion is to be incorporated into grammars.

Consider a second minimalist project. The above considerations lead to the conclusion that grammars must interface with CI and AP components. Given this, there is a premium on grammatical principles that originate in this fact. For example, if some sorts of grammatical objects are uninterpretable by the CI or AP interfaces then grammatical structures (e.g. phrase markers) that contain these will be illegible by these interfaces unless these wayward objects are dispatched before interpretation. In effect, the interfaces impose “Bare Output Conditions” that the objects which the grammar generates must conform to. Given least effort criteria favored by principles of substantive economy, the favored accounts will exploit Bare Output Conditions in limiting grammatical structures.

Methodological economy further prompts us to consider how derivations proceed. We would be looking for those that have a least effort flavor in producing the objects that the interfaces interpret. For example, we should consider theories that employ natural locality conditions (e.g. require that derivations be short, or movements be local or operations be simple). In effect, minimalism favors accounts that implement least effort themes by promoting derivational economies of various sorts.

Thus, minimalist commitments will lead one to develop accounts that rest on various least effort notions. This could mean accounts that rest on how strings are generated, in which case economy of derivational resources is the key, or simplest to interpret, in which case economy of representational resources, i.e. Full Interpretation, are highlighted. In sum, given the general setting noted above, we will begin to look for two kinds of conditions on grammars; ones that correspond to the filtering effects of the interfaces, (“Bare Output Conditions”) or those that correspond to the derivational features of the grammar (“Economy Conditions”). Filtering mechanisms that resist interpretation in one of these ways are less favored.

Consider another set of questions minimalist considerations lead to. What are the basic primitives of the system; the basic relations, objects and operations? If phrases exist (i.e. (6) above) then if phrases are organized in roughly X' terms, as standardly assumed, then a set of privileged relations are provided. In X'-theory, phrases have (at least) three parts -heads, complements and specifiers- and invoke two relations -head/complement and specifier/head. Given the obvious fact that natural languages contain phrases, UG requires these objects and relations whatever else it needs. Therefore, parsimony counsels that at most these objects and relations should be part of UG. This implies, for example, that sentences be analyzed as types of phrases rather than as having an idiosyncratic structure. This is essentially the conclusion GB has already drawn. Labelling sentences as IPs, TPs and CPs embodies this consensus.

The recognition that phrases are a minimally necessary part of any theory of grammar further suggests that we re-examine whether we need government among the inventory of basic

grammatical relations. Methodological simplicity urges doing without this extra notion given that we already have two others. All things being equal, we should adopt government only if the X'-theoretic relations we already have prove empirically inadequate.

We can of course go further still. We can reconsider the status of X' theory itself. How natural is this? The fact that phrases exist does not imply that they have an X' structure. What features of phrasal organization follow from the mere fact that they exist and which require more elaborate justification. For example, are bar levels basic features of phrases or simply the reflections of something more basic? Is the fact that heads take maximal projections as complements and specifiers a primitive principle or the reflection of something more primitive? How much of X' theory needs to be assumed axiomatically and how much results from the fact that phrases must be constructed and interpreted?

Consider one last illustrative example. (4) above notes that displacement is an obvious fact about natural languages. Assume, for sake of argument, that displacement is due to the fact that grammars have movement rules like those assumed in typical GB accounts. We can then ask how much of the GB theory of movement is motivated on minimalist grounds. In a GB theory, for example, movement is defined as an operation that leaves traces. Are traces conceptually required? In part perhaps, insofar as they simply model displacement and provide a mechanism for coding the fact that expressions can be interpreted as if in positions distinct from the ones they overtly appear in. Does the simple fact of displacement motivate the GB view that traces are indexed categories without lexical contents i.e. '[XP e]_i'? Or does the existence of displacement phenomena suffice to ground the claim that traces are subject to special licensing conditions (such as the ECP) that do not apply to lexical items more generally? This is far less clear.

Traces in GB are grammar internal constructs with very special requirements that regulate their distribution. Historically, the main motivation for traces was their role in constraining overgeneration in the context of a theory where movement was free, not in providing vehicles for interpretation. The primary service traces and the conditions on them provided was to filter unwanted derivations that resulted from a grammar based on a rule like 'Move alpha'. Why assume that such entities exist, especially in the context of minimalist theory in which it is often assumed that movement is not free (as it is in GB) but only occurs if it must, occurs only if needed to produce an object that the interpretive interfaces can read? Methodologically we should resist postulating traces as grammatical formatives unless strong empirical reasons force this conclusion. On conceptual grounds traces are of dubious standing.

What could replace traces? Well, we independently need words and phrases (see (5) and (6)). Why not assume that they are used by the grammar to accommodate displacement? In other words, assume that traces are not new kinds of expressions but they are copies of expressions that are already conceptually required. This seems simpler than postulating a novel construct if one's main goal is to accommodate displacement. In short, GB traces must earn their keep empirically and all things being equal a copy theory of traces is preferable.

What holds for traces holds for other grammar internal formatives as well; PRO, 0-operators and chains to name three more. It also brings into question the value of modules like the ECP, control theory and predication whose purpose is to monitor and regulate the distribution of these null (grammar internal) expressions. None of this means that the best theory of UG won't contain such entities or principles. However, minimalist reasoning suggests that they be adopted only if there is strong empirical motivation for doing so. On conceptual grounds, the burden of proof is on those who propose them. At the very least, minimalist scruples force us to reconsider the empirical basis of these constructs and to judge whether their empirical payoffs are worth the methodological price.

These sorts of considerations can be easily amplified, as we will see when we get into details in the chapters that follow. This suggests that the “big facts” in tandem with the principles of methodological and substantive economy can in fact be used to generate interesting research projects. These considerations prove more fruitful still when the proposals they prompt are contrasted with an interesting foil. Government Binding Theory proves to be an admirable straight man to the minimalist jokster.

3. Using GB

GB is the most successful P&P theory elaborated to date. It affords a useful subject on which to implement the minimalist methodological concerns outlined above. In what follows, we will constantly be assuming (one of) the standard GB approaches to a particular problem and asking whether we can do better. In effect, the GB story will set the mark that any competing reanalysis will have to beat.

By contrasting the alternatives with the GB account we are able to accomplish two things. First, we allow ourselves the option of rethinking the empirical bases of various modules of UG: e.g. what data lies behind the binding theory or theta criterion. Second, we can ask whether the GB implementation of some particular grammatical approach, the leading idea as well as technical implementation, is really for the best that we can come up with.

Let us put this another way. Section 2 shows that minimalist concerns can be used to raise many questions. These questions are sharpened when asked in a GB context. For example, consider the fact that sentences pair sound and meaning and that this fact is accommodated within GB by having PF and LF phrase markers. A reasonable minimalist question given GB as a starting point is whether the other two GB levels, DS and SS are dispensible and if not why not? Observe that even if we come to the conclusion that one or the other or both of these levels must be retained, we will have a far better understanding about what justifies them if we go through this minimalistically inspired process. Of course, it is always possible that we might discover that DS and SS are convenient but not really necessary. This discovery would, in turn, prompt us to see whether certain technical alternatives might allow us to get the results for which we postulated these levels but without having levels at all. Chomsky 1993 attempts this and suggests that perhaps our acceptance of a four level theory (as in GB) was somewhat hasty.

At any rate, a second important feature of what follows is our always beginning with the GB account of any given phenomenon and then asking whether there is anything minimalistically undesirable about it, e.g. does it use undesirable primitives, does it rely on undesirable operations or levels, etc. The third step is to consider alternatives that might do better. In other words, in what follows we emphasize the process of minimalist analysis and try to see how it leads to novel hypothesis.

It is worth keeping in mind that the fact that an analysis is minimalistically suspect does NOT imply that it is incorrect. To repeat, minimalism is a project: to see just how well designed the faculty of language is given the obvious facts that we know about it. The answer could be that it has design flaws. We could come up with this answer empirically: the best account of the grammar suffers from a certain redundancy or inelegance. It is, after all, conceivable that GB is roughly right and that when all the relevant facts are considered it is the best theory of grammar we can devise. Even this conclusion would be interesting. For it would indicate that starting from different initial considerations we end up with the conclusion that GB is roughly right. In what follows you will see that this is not the conclusion that many have come to. However, it could have been and still could be. This does not remove the interest of analyzing GB accounts

in minimalist terms. For what minimalism does is afford us the opportunity of rethinking the empirical and theoretical bases of our claims and this is always worth doing.

This said, the reader will observe that grammars that arise from minimalist reflection have a very different “look” from the standard GB varieties. One aim of what follows is to move readers through the complexities of some current speculations that fly under the minimalist flag.

4. The Basic Story Line

The Minimalist Program explores the hypothesis that the language faculty is the optimal realization of interface conditions. In other words, it is a nonredundant and optimal system in the sense that particular phenomena are not overdetermined by linguistic principles and that the linguistic system is subject to economy restrictions with a least effort flavor. The program also addresses the question of what conditions are imposed on the linguistic system in virtue of its interaction with performance systems (“the bare output conditions”).

Earlier versions of the Principles and Parameters Theory worked with the hypothesis that the linguistic system has several levels of representation encoding systematic information about linguistic expressions. Some of these levels are conceptually necessary, since their output is the input to performance systems which interact with the linguistic system. The Minimalist Program restricts the class of possible linguistic levels of representation to only the ones which are required by conceptual necessity, namely, the ones which interface with performance systems.

Chomsky (1993, 1994, 1995:chap. 4) takes these performance systems to be the Articulatory-Perceptual System (A-P) and the Conceptual-Intentional System (C-I). The linguistic levels which interface with A-P and C-I are PF and LF, respectively. Assuming that these are the only interface levels, PF and LF can be conceived of as the parts of the linguistic system which provide instructions to the performance systems. Under the Minimalist perspective, all principles and parameters of the linguistic system should be stated in either LF or PF terms, perhaps as modes of interpretation by the performance systems. Furthermore, all principles and parameters should make reference only to elements that function at the interface levels and to local relations among them. Linguistic expressions are then taken to be optimal realizations of interface conditions, where optimality is determined by economy conditions specified by Universal Grammar (UG).

Another assumption is that the language faculty is comprised of a lexicon and a computational system, which is strictly derivational (see Chomsky (1994:5-6, 1995:223-224)). The lexicon specifies the items which enter into the computational system and their idiosyncratic properties, excluding whatever is predictable by principles of UG or properties of the language in question. The computational system arranges these items in a way to form a pair (p, l) , where p is a PF object and l is an LF object. If l and p are legitimate objects (i.e. they satisfy Full Interpretation in the sense of Chomsky (1986b, 1993)), the derivation is said to converge at LF and at PF, respectively. If either l or p does not satisfy Full Interpretation, the derivation is said to crash at the relevant level. A derivation is taken to converge if and only if it converges at both LF and PF.

If D is the set of permissible derivations which yield a pair (p, l) , the set of convergent derivations C is thus the subset of D whose members satisfy Full Interpretation at LF and at PF. Finally, the set of admissible derivations A constitute the subset of C which is selected by economy considerations. In other words, the derivations which reach the performance systems are only the ones which converge in an optimal way.

In the chapters that follow, we discuss specific aspects of the Minimalist Program, as formulated in these general terms.