

Approaching Parameters from Below*

Cedric Boeckx, Harvard University

1. Revisiting Plato's Problem

“Our ignorance of the laws of variation is profound.” Darwin’s words, taken from the *Origin of Species* (1859[1964]:167), aptly characterize the state of our knowledge of linguistic variation. At a time when some linguists feel that some *why*-questions are ripe for the asking, there is no consensus regarding why linguistic variation should exist at all, let alone why it should take the form that we think it does. There is indeed very little substantive discussion of the issue of linguistic variation in the context of the minimalist program. This may come as a surprise to some, as it is fairly common in the literature to introduce the minimalist program in the context of the Principles and Parameters approach. Specifically, it is fairly common to say that the minimalist program grew out of the perceived success of the Principles and Parameters approach, which arguably for the first time in the history of the field enabled linguists to resolve in a feasible way the tension between universal and particular aspects of language, and offered a fruitful way of thinking about how the child acquires her language (Plato’s problem).¹ There is a lot of truth to this statement, but in the absence of important qualifications (rarely spelled out in the literature) it can be highly misleading. In particular, it may give the impression that the specific implementation of the Principles and Parameters approach explored in the Government-Binding era essentially solves Plato’s Problem (‘abstractly,’ of course, since no one is under the illusion that GB-theorists got all the details right). This impression must be dispelled, for the idea that a GB-style Principles-and-Parameters architecture provides the right format for a solution to Plato’s Problem is, I think, seriously mistaken, on both empirical and conceptual grounds.² A Principles and Parameters model of the GB

* The present work benefited from comments from the participants in my advanced syntax seminar at Harvard University in the Spring of 2007, and in my intensive graduate course in Vitoria (Basque Country) in the Summer of 2007, as well as from the audiences at the Biolinguistics Meeting in Venice (June 2007), at the 9th Tokyo Conference on Psycholinguistics (March 2008), and at a talk at University of Massachusetts, Amherst (November 2007). I have been talking to many people about parameters over many years. I cannot thank them all individually here, but I would like to single out Fritz Newmeyer, Norbert Hornstein, Paul Pietroski, Noam Chomsky, Juan Uriagereka, Dennis Ott, Bridget Samuels, Hiroki Narita, Angel Gallego, Massimo Piattelli-Palmarini, Koji Sugisaki, and Luigi Rizzi.

This work is a very preliminary report. A much more detailed exposition can be found in Boeckx (in progress).

¹ See Chomsky (2005:8-9; 2007:2-3), Hornstein (2001:4), Boeckx (2006:59), among others.

² I am certainly not alone in claiming that the traditional Principles and Parameters approach is flawed (see Culicover 1999, Newmeyer 2005, Jackendoff 2002, Goldberg 2006), but unlike the authors just cited, I am not inclined to throw the baby with the bathwater (‘if minimalism depends on the success of the Principles-and-Parameters mode, and the latter fails, then ...’). It is important to bear in mind that the Principles and

style conceived of Universal Grammar as consisting of two main ingredients: principles that were truly universal, manifest in all languages, and, more importantly, principles whose formulations contained open values (parameters) that had to be fixed in the course of language acquisition. Such parametrized principles can be thought of as forming a network that is only partially wired up at the initial state, and that must await a fair amount of data processing to be fully operative. (This is the switchboard metaphor made famous by Jim Higginbotham, and adopted in Chomsky (1986:146).)

This way of thinking about the acquisition process has had undeniable success. It has led to a revival of acquisition studies, and produced some extremely interesting results, in the domains of language development and of comparative syntax. But I think that the traditional way of telling the Principles and Parameters story has outlived its usefulness. For one thing, the traditional Principles and Parameters model is no longer compatible with the way minimalists think of Universal Grammar. As I will discuss in some detail below, if minimalists are right, there cannot be any parametrized, and the notion of parametric variation must be rethought.

Second, it is fair to say that empirically the expectations of the traditional Principles and Parameters model have not been met. GB theorists expected a few points of variations each with lots of automatic repercussions throughout the grammar of individual languages ('macro-parameters'), but they found numerous, ever more fine-grained, independent micro-parameters.

In the limited space allotted to me here, I will only be able to focus on the way in which central minimalist tenets clash with the traditional Principles and Parameters approach. I will have to leave a detailed examination of the alternative offered by the minimalist program to another occasion (see Boeckx in progress). As for the empirical shortcomings of the traditional Principles and Parameters approach, I urge the reader to take a look at Newmeyer (2005:77-103) to see some of the major difficulties the standard model faces, and to appreciate the empirical task ahead (for further discussion, see Boeckx in progress).

2. Two ways of approaching UG (and parameters)

To understand the current uneasiness existing between minimalism and the standard Principles and Parameters model it is instructive to go back to an important document of the GB era: Chomsky's introduction to *Lectures on Government and Binding* (Chomsky 1981:1-16). There Chomsky outlines the Principles and Parameters approach that was pursued ever since and that Mark Baker articulated in a very accessible way in his *Atoms of Language* (Baker 2001). Chomsky makes clear that the appeal of the Principles and Parameters model is that it provides a compact way of capturing a wide range of differences. As he notes (p. 6), "[i]deally, we hope to find that complexes of properties ...

Parameters model is an approach, not a specific theory. The GB way of articulating the Principles-and-Parameters logic was just one possible way. I believe that the minimalist program offers us a different, more adequate way of exploring how principles and parameters may interact – a possibility that very few researchers in the Chomskyan tradition seem to appreciate (for a few exceptions that I have been able to find, see Raposo 2002, Hornstein in press, Gallego 2008b).

are reducible to a single parameter, fixed in one or another way.” This is clearly the ideal of Parametric Syntax. Elsewhere, Chomsky makes clear that this ideal depends on the richness of UG: “If these parameters are embedded in a theory of UG that is sufficiently rich in structure, then the languages that are determined by fixing their values one way or another will appear to be quite diverse (...)” (p. 4) The starting assumption of Government-and-Binding was this:

“What we expect to find, then, is a highly structured theory of UG [...]” (p. 3)

In a recent paper, Chomsky (2007:2) makes this very clear: “At the time of the 1974 discussions, it seemed that FL must be rich, highly structured, and substantially unique.”

As Baker (2005) insightfully observes, the traditional Principles and Parameters model takes UG to be ‘overspecified.’ This is perhaps clearest in Yang’s 2002 model, where the acquisition task is reduced to choosing one among all the fully formed languages that UG makes available. In other words, the traditional Principles and Parameters model is ultra-selectionist, guided by the slogan that learning (a little) is forgetting (a lot).

Such an approach, relying on a richly structured UG, culminates in Baker’s (2001) Parameter hierarchy (Figure 1), a (partial) characterization of the dependencies among parameters (i.e., parametrized principles).

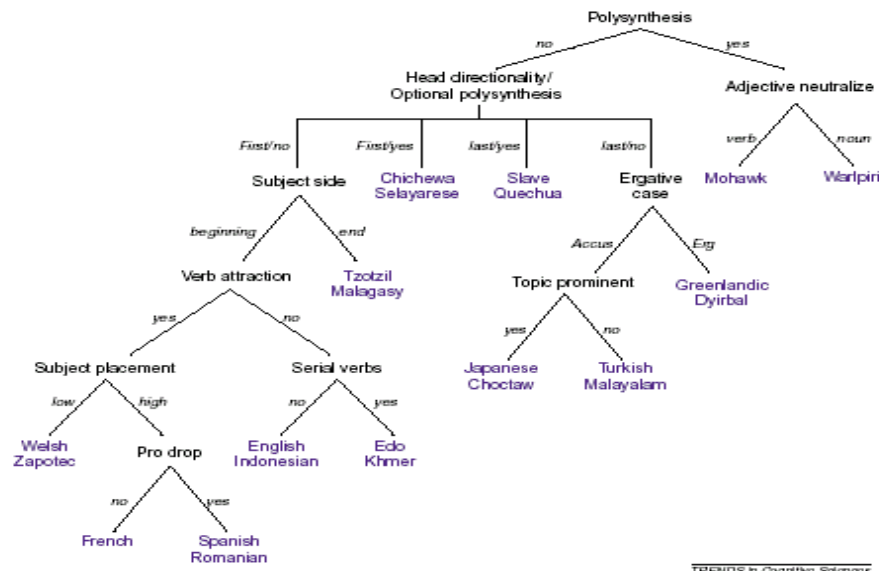


Figure 1: Baker's Parameter hierarchy (taken from Baker 2003)

The most obvious question that arises in a minimalist context, where one seeks to go beyond explanatory adequacy, is, Where does the hierarchy come from? That is, What are the design principles that would make this specific hierarchy emerge?

I do not know of any work addressing this issue. I suspect that this is due in part to the fact that Baker's hierarchy makes use of concepts (such as 'topic prominence') that have never been rigorously defined in a generative framework. The hierarchy also conceals layers of complexity (well-known to practitioners in the field) in the formulation of 'serial verbs' or 'pro-drop' that would undoubtedly render the hierarchy more intricate and elaborate. But the lack of explicit discussion of Baker's hierarchy is also due to the fact that most syntacticians working within the minimalist program have shifted their attention away from rich, complex, parametrized principles, and toward the formulation of more basic operations (such as Merge, Agree, etc.).

This is part of the shift that Chomsky (2007:4) characterizes thus:

“Throughout the modern history of generative grammar, the problem of determining the character of FL has been approached “from top down”: How much must be attributed to UG to account for language acquisition? The M[inimalist] P[rogram] seeks to approach the problem “from bottom up”: How little can be attributed to UG while still accounting for the variety of I-languages attained.”

As I write, the research program that approaches UG from below has generated a few interesting hypotheses regarding the nature of basic operations (such as Merge, Agree, and Phasal Spell-Out) that are supposed to capture the essence of the core UG principles identified in the GB era. Such research has (implicitly) abstracted away from the fact that most principles in the GB era were parametrized, and has assumed that things pertaining to linguistic variation will fall into place once we understand the nature of principles. This research strategy was made possible by the perceived success of the traditional Principles and Parameters approach as a model solution to Plato's problem. But it is important to note that the perceived success was not, or at any rate, need not be,

dependent on the exact nature of parameters. What enabled the formulation of the minimalist program is the perceived possibility of segregating the universal from the particular (suggested by the gradual abandonment of language-specific, construction-specific rules in favor of parametrized principles), not so much the formulation of parametrized principles. In this sense, the traditional Principles-and-Parameters model did not go all the way, as it kept principles and parameters intertwined in the guise of parametrized principles. I think that some syntacticians nevertheless saw in this the very possibility that one day principles and parameters could be fully segregated, which in turn warranted certain minimalist moves. Be that as it may, I believe that recent developments within minimalism make it possible to entertain the possibility of a truly universal, non-parametric syntax. Perhaps this is what Chomsky had in mind when he formulated the Uniformity hypothesis (Chomsky 2001:2), given in (1).

(1) Uniformity Hypothesis

In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances.

Chomsky's formulation is not clear enough, as it fails to make precise what "easily detectable properties of utterances" are. Here I would like to entertain a strong version of the Uniformity hypothesis – call it the Strong Uniformity Thesis (SUT) – which states that narrow syntax is not subject to variation, not even parametric variation. In other words, there is only one syntax, fully uniform, at the heart of the faculty of language, underlying all languages.

(2) Strong Uniformity Thesis

Principles of narrow syntax are not subject to parametrization; nor are they affected by lexical parameters

Under SUT, points of variation (i.e., parameters) would be confined to the margins of narrow syntax, especially the morpho-phonological component (PF).³

My reason for pursuing SUT is that if minimalist research is taken seriously, there is simply no way for principles of efficient computation to be parametrized. Contrary to what Baker and Collins 2006 seem to assume, it strikes me as implausible to entertain the possibility that a principle like 'Shortest Move' could be active in some languages, but not in others. Put differently, if minimalist research is on the right track, there can be no parameters within the statements of the general principles that shape natural language syntax. In other words, narrow syntax solves interface design specifications optimally in the same way in all languages (contra Baker 2006 and Fukui 2006). Modifying Marantz (1995:380), minimalism can be seen as the end of parametric syntax. I believe that this conclusion is a natural consequence of the claim at the heart of the generative/biolinguistic enterprise that there is only language, Human, and that this organ/faculty emerged very recently in the species, too recently for multiple solutions to design specifications to have been explored.

³ I remain silent concerning the possibility of 'pragmatic variation' of the sort discussed by Svenonius and Ramchand 2007 and Matthewson 2006, 2008.

3. Three roads to Strong Uniformity

I also believe that SUT has become more realistic as a result of three recent developments in the minimalist program, to which I would like to turn now.

The first development concerns our evolving conception of the lexicon in a minimalist grammar. Readers familiar with the literature on parameters will have no doubt noted that SUT is reminiscent of the conjecture first formulated by Hagit Borer, and endorsed by Chomsky 1995, according to which “the availability of variation [is restricted] to the possibilities which are offered by one single complement: the inflectional component.” (Borer 1984:3)

But like Chomsky’s Uniformity hypothesis above, Borer’s conjecture can be construed in more than one way:⁴ What counts as a lexical property? What is the lexicon, and where is it located with respect to other components of the grammar?

Many (including Borer herself) found Borer’s conjecture appealing because traditionally the lexicon is viewed as the repository of idiosyncracies (see Bloomfield 1933; Chomsky 1995). Since the lexicon is clearly the locus of learning, it makes sense to locate parameters there, since parameter settings must be chosen under the influence of the input the child receives. But recently, under the impetus of Hale and Keyser’s (1993, 2002) research, many aspects of lexical knowledge have been found highly principled, so much so that talk of a l(exical) syntax has appeared feasible. This trend was continued by proponents of Distributed Morphology (Halle and Marantz 1993; Marantz 2000) and by Borer herself (Borer 2005). The net result of this line of research is that much of what we took to be a ‘messy’ pre-syntactic lexical component has now been relegated to post-(narrow) syntactic areas, such as the morpho-phonological component (PF). The pre-syntactic lexicon has been considerably deflated, and very few opportunities remain for it to influence syntactic derivations in dramatic ways. (I return to these few opportunities in section 4 below.)

Such neo-constructionist or ‘epilexicalist’ approaches to the lexicon provide the first theoretical development that virtually forces SUT upon us: once parameters are confined to the lexicon, and much of the lexicon is relegated to post-syntactic components, there is little chance left for (narrow) syntax to be subject to parametric variation.

The second development I would like to highlight is the rise of the “Unrestricted Merge” model of syntax, beginning with Chomsky 2004. It has been pointed out on numerous occasions that minimalism is a research program, and that there are multiple ways of exploring it. In what can be called the early minimalist period (epitomized by Chomsky 1993, 1995), emphasis was laid on the Last Resort character of syntactic operations, as a way to reveal economy principles at work. The most productive way to conceive of syntactic operations being driven by Last Resort operations turned out to be to think of operations taking place to resolve some featural illegitimacy on specific elements (cf. principles like Greed, Attract, and the like). This led to a detailed examination of what features can be found on lexical items, how these could vary, and so on. The number, type, and properties of features become a focus of attention in the literature now known

⁴ Borer herself develops a specific implementation of her conjecture in her 1984 book, but her proposal must be rethought in light of the emergence of a minimalist architecture.

as the Cartography approach. It is also at the heart of so-called Crash-proof syntax models. More recently, a few researchers (beginning with Chomsky 2004) have begun to explore an alternative according to which syntactic operations apply ‘freely’ solely in virtue of elements bearing ‘edge features’. Such an approach vastly overgenerates, and must be supplemented with interface-internal conditions that filter out unwanted combinations (all of which, it must be stressed, are legible). In this way of looking at grammar, the edge feature boils down to the property of being a lexical item. Unlike the features that were used in the early minimalist period, which varied from lexical item to lexical item, edge features are the common denominator among lexical items; they cannot be the locus of (parametric) variation. In other words, reliance on edge features has insulated narrow syntax from parametric effects. Syntax has become immune to the sort of lexical vagaries and idiosyncracies that could not be ruled out in the early minimalist period.⁵

The third development that I want to stress in this section concerns the growing suspicion that the mapping from syntax to sense (SEM) is architecturally privileged over the mapping from syntax to sound/sign (PHON). Until recently, models of grammar treated PF and LF symmetrically (think of the symmetry of the inverted-Y model), but for a very long time (going back to remarks by traditional grammarians like Jespersen) it has been clear that there is quite a bit a variation at PF, but virtually none at LF. Thus, Chomsky (1998:123) notes that PHON/PF “yield[s] a variety of morphological systems, perhaps superficially quite different in externalization of a uniform procedure of constructing an LF representation.” As minimalists made progress in understanding the nature of operations internal to narrow syntax, it became clear that these were in some sense designed primarily to meet demands on the sense-side of the grammar (e.g., the elimination of unvalued/uninterpretable features). Chomsky (in press:3) puts it thus:

“It may be that there is a basic asymmetry in the contribution to language design of the two interface systems: the primary contribution to the structure of [the] F[aculty of] L[anguage] may be optimization of the C-I [sense] interface.”

In other words, narrow syntax is optimally designed to meet demands from the meaning side, and externalization (PF) is akin to an afterthought, or appendix. Since variation clearly exists on the sound/sign side (an unavoidable consequence of the fact that this is that aspect of language that is used for communication and learning, and communication/imitation/reproduction is a more or less, imperfect affair), but no evidence exists that it is found at the meaning side, it is not implausible to think of narrow syntax as completely uniform (meeting LF-demands), and not affected (design wise) or adapted to cope with or code for variation in the guise of (syntactic) parameters. To put it differently, the LF/PF-asymmetry naturally leads one to expect a uniform narrow syntax, designed to meet the uniform demands at the meaning side in an optimal fashion.

I should stress in closing this section that none of the three developments reviewed here provide incontrovertible evidence in favor of SUT. They simply point to the fact that

⁵ Gallego 2007 explores variation in phase-structure by exploiting Chomsky’s (in press) definition of phase-heads as *u*F-bearers, a type of variation that I consider too ‘deep’ or syntactic. For an alternative that does not rely on syntactic variation, see Gallego 2008a.

SUT emerges as the most natural hypothesis in the context of a minimalist program rigorously pursued.

4. Parameter schemata

I pointed out in the previous section that various developments in the minimalist literature conspire to make SUT plausible. Architecturally speaking, there is little room left in a minimalist conception of grammar for lexical parameters to affect narrow syntax. In this section I review a few proposals from the literature I am familiar with that characterize ways in which parameters may affect syntax. Following Giuseppe Longobardi's terminology, I will refer to these as attempts to uncover parameter schemata.⁶

Perhaps the most detailed investigation in this domain comes from Longobardi himself. In a series of works (Longobardi 2003, 2005, Gianollo, Guardiano, and Longobardi in press, Guardiano and Longobardi 2005), Longobardi argues that languages may differ in the following ways:

- (3) For a given language L,
 - a. Is F, F a functional feature, grammaticalized?
 - b. Is F, F a grammaticalized feature, checked by X, X a lexical category?
 - c. Is F, F a grammaticalized feature, spread on Y, Y a lexical category?
 - d. Is F, F a grammaticalized feature checked by X, strong?

Longobardi (2005:410-1) comments on (3) as follows. (3a) is meant to capture the fact that in some languages dimensions like definiteness are not marked on functional items like Determiners. (3b) boils down to asking whether a given functional head acts as a Probe. (3c) asks whether some feature F has an unvalued/uninterpretable variant participating in Agree-relations. (3d) is meant to capture that fact that in some languages movement is overt, while in others it is covert. (It is worth pointing out that Longobardi

⁶ Works on parameter schemata follow in the steps of GB-syntacticians like Borer (1984), Fukui (1986, 1995) and Webelhuth (1991), who sought to capture the format of parameters. Such works are sometimes advertised as trying to uncover the nature of possible parameters. But as Howard Lasnik points out (p.c.), one must exercise caution with the phrase 'possible parameter.' As he notes, in the heyday of GB, it was very popular to say that now we needed a theory of possible parameters (the way we used to need a theory of possible transformations). But the analogy was completely false. We needed a theory of possible transformations because the child had to pick out the right transformations, and if any conceivable formal operation could be a transformation, the task would be hopeless. But in a Principles and Parameters model, the child doesn't have to pick out the right parameters from a set of possible ones. Rather, the child just has to pick out the right *values* for wired-in parameters. So the true analogy would be a theory of possible parameter values. But that, of course, is totally trivial. What we really want to determine is the *actual* parameters. Works on parameter schemata is not about possible parameters, but rather, (in the absence of a complete map of all lexical items made available by UG) about what actual parameters may look like (abstractly).

(2005:411) leaves open the possibility of a fifth parameter schema pertaining to pronunciation, of the sort familiar with *pro*-drop phenomena.)

The difference between (3b) and (3c) is not entirely clear to me (for a feature to be a Probe and be checked, it must Agree, hence ‘spread’), and once eliminated, Longobardi’s schema converges with Uriagereka’s 1995 characterization of his F-parameter in (4) (posited to capture effects at the left periphery of the clause, distinguishing Germanic from Romance, and also cutting across Romance varieties):

- (4) a. Is F present in the language?
- b. Is F strong?

Roberts and Roussou 2002 also makes a similar proposal, when they propose the following parameter format:

- (5) a. Does F require PF-support?
- b. Is the support provided via Merge or Move?

Likewise, Borer 2005 puts forth that “variation within the functional domain can only be attributed to the mode in which open values are assigned range: either via direct merger of a head, or movement of a head, or by insertion of an adverb, or by Spec-head-agreement.”

As we see, there is a fair amount of consensus about the kind of parameter one finds in UG.⁷ The consensus boils down to a nested structure like (6).⁸

- (6) Is F present/active in the language? Yes/NO
- If Yes, Does F give rise to Movement, or simply Agree (/Merge)?

The nested character of (6) may prove useful in recapturing some of the effects of Baker’s (2001) parameter hierarchy. I will not explore this possibility further here, as I am skeptical regarding the adequacy of the first question in (6). I think that all languages make use of the same pool of features, and that one of the ways in which languages differ is how they express the relevant feature F. Specifically, I submit (following Fortuny 2008, and Gallego 2008b; see also Bejar 2003) that languages may choose to express f_1 and f_2 separately (analytically) or as a bundle (syncretically). This different way of thinking about the mode of parametrization available has the effect of breaking up the

⁷ The emerging consensus provides supports for Fukui’s 1995 conjecture that (i) Parametric variation outside the lexicon must be limited to ordering restrictions (“linearity”), and that (ii) inside the lexicon, only those functional elements that do not play a role at LF (such as AGR-elements) are subject to variation (in the sense that they can be absent from the lexicon of a particular language).

⁸ The formulation in (6) requires one to take head-movement to be head-merger under adjacency in PF (Harley 2004) or (if syntactic) movement to a Specifier position (Matushansky 2006, Vicente 2007). The formulation in (6) is also facilitated by the adoption of the single-cycle syntax model that takes ‘covert’ movement to be either pure Agree or Pre-Spell-Out movement following by pronunciation of a low copy.

nested dependency expressed in (6). Indeed I claim that the dependent question in (6) must be reformulated independently, viz. as: Is F viral? (i.e., does F exhibit a uF variant?). Finally, languages may differ in whether a specific (phase-)head⁹ is strong (uF -bearing) or weak (defective).

Notice that this way of formulating lexical parameters makes them interdependent; it does not single one of the parameter is the one on which all others dependent. Notice also that the lexical parameters I put forward barely affect narrow syntax. It's the same syntax whether f_1 and f_2 form a bundle or not. It is also the same syntax whether a specific feature F is viral or not. And, finally, it is the same syntax regardless of whether a phase-head is weak or strong.¹⁰

Of course, there are more ways in which languages may differ, but I contend that all other 'parametric' options arise in the post-syntactic morpho-phonological component, such as whether a head H allows its specifier to be filled by overt material, or whether the head or the tail of a chain can/must be pronounced, or whether a given head H is affixal and requires its expression in the vicinity of another head, or whether a head H precedes or follows its complements. These options arise at the point of Spell-Out, when syntactic structures must be linearized and the demands of (late-inserted) morphemes must be met.

5. Clustering effects

The preceding section outlined a very impoverished version of a parametric theory, with parametric options isolated from one another, and localized to specific heads. Let me call these nano-parameters,¹¹ to keep them distinct from the more traditional octopus-like macroparameters that GB-research made famous (beginning with Rizzi 1982 and continuing with Baker 1996, 2001, 2008).

Some researchers may find these wholly inadequate, unable as they seem to express the clusters of variation that are invariably brought up in the literature on parameters – especially the clusters that seem to align parametric values across independent heads (as the traditional head-parameter does).¹² But although such clusters of variation did much to dispel the common impression (endorsed by structuralists) that languages can vary from one another indefinitely, it is fair to say that few of the implicational statements at

⁹ Following Chomsky 2007, in press, I assume that only phase-heads probe. See Boeckx 2008, Richards 2007, and Gallego 2007 for detailed discussion.

¹⁰ The only difference will pertain to the timing of transfer, or the amount of structure transferred to the interfaces at a given time (see Boeckx 2008, Gallego 2007). This should not be characterized as a syntactic parameter, as the same issue of timing/amount of transfer arises whenever a lexical choice is made/a numeration is constructed (active v^o vs. passive v^o) in any given language.

¹¹ I use the term nano-parameter as opposed to micro-parameter to stress their low-level, 'flat' character. (I would reserve the term micro-parameter for points of variation displaying some minimal nested structure like (6).)

¹² Clusters resulting from the fact that a given head affects the surface appearance of many constructions due to the fact that this head is heavily affected by a variety of derivational options (e.g., INFL in Rizzi's 1982 *pro*-drop parameter) are much less problematic for the sort of nano-parameter I entertain.

the heart of traditional Principles-and-Parameters approach have stood the test of time. Some languages indeed appear to display the clusters the theory predicted, but many languages only display a few of the predicted clustering effects, and more often than not, languages show no clustering effects whatsoever. (For this reason alone, nano-parameters strike me as more adequate empirically.) Newmeyer 2005 is correct in stressing that the rarity of massive clustering effects takes much of the gloss away from the traditional Principles-and-Parameters model.^{13,14} Newmeyer goes on to suggest that clustering effects are just tendencies (probable, but not the only possible languages), to be captured in terms of performance effects, specifically parsing strategies.

I think Newmeyer is correct in taking parametric clusters to be tendencies, not to be accounted for in terms of UG principles, but I would like to suggest that these tendencies do not (always)¹⁵ arise due to parsing strategies. I propose that some are due to certain biases in the learning process. That is, I would like to suggest that the types of languages that parametric clusters describe act as attractors as the child acquires her language, and that only a fair amount of positive counterevidence in the data leads the child to settle on a less homogenous system of parameter settings.¹⁶

In a nutshell, the guiding intuition I would like to pursue is that clustering results from the child seeking to maximize the similarity across parameter settings, harmonize their values, and thereby economize what must be memorized (via the formation of a generalization across similar parameters). My suggestion goes in the direction of a recent trend expressing a renewed interest in the nature of learning and the interaction between nature and nurture (see Yang 2002, 2004, 2005; Pearl 2007). As Yang and Pearl have made clear (in a linguistic context; see Gallistel 1990, 2006 for more general remarks on learning modules), a proper characterization of the learning task requires paying attention to three important components: (i) a defined hypothesis space (for language, UG), (ii) data intake, and (iii) a data-update algorithm.

¹³ It is interesting to note that Nicolis's 2006 re-examination of Rizzi's classical formulation of the *pro*-drop parameter (with consequences not only for null subjects, but also post-verbal subjects and long-distance subject extraction) ends up with the claim that the only robust correlation is this: if a language allows null referential subjects, it also allows null expletive/non-referential subjects. But this is clearly a lexical correlation (null referential subjects being a (featural) superset of non-referential subjects), expressible at the level of a lexical (nano-)parameter, and not a syntactic correlation of the sort one would expect from a traditional (syntactic) parameter and of the sort Rizzi had in mind.

¹⁴ Since Newmeyer wrote his 2005 book, Baker has mounted a new challenge based on a macro-parameter regulating agreement (Baker 2008). For an alternative approach that does not require any appeal to a syntactic parameter, see Boeckx 2008.

¹⁵ Some tendencies may indeed be due to performance factors. For pertinent discussion, see Rizzi's 2007 treatment of the root-infinitive stage in Child Language, where it is suggested that some parametric values are assumed so as to reduce the computational processing load and facilitate production.

¹⁶ I predict that, all else equal, as the amount of counterevidence decreases, the shape of the language acquire will drift toward the attractors. For relevant discussion on language drift, see Roberts 2007.

Yang and Pearl have shown in detail that there are certain data intake filters, or certain learning biases that must be assumed to characterize the language learning task adequately. For example, Yang shows that the learner must be ready to tolerate a certain amount of noise within limits (which Yang takes to account for a variety of irregular, or semi-regular morphological processes). In a similar vein, Pearl has argued that the child must be ready to filter out potentially ambiguous cues for certain parameter settings. I would like to add to this a bias by which the child strives for parametric value consistency, a Superset bias with which the child processes the data, and which she abandons only if there is too much data contradicting her initial hypothesis.

(7) Superset Bias

Strive for parametric value consistency among similar parameters

For example, if the child learns that V precedes its complement and T precedes its complement, she will be inclined to hypothesize that the next head she encounters will also precede its complement, and will only reject her hypothesis if she finds enough positive counterevidence.

The Superset bias should not look too unfamiliar: something very much like it is at the heart of the evaluation metric in Chomsky and Halle 1968. It may also underlie recurring intuitions in the domain of markedness (see, e.g., Hyams 1986:chap. 6) and may help us characterize the notion of default parameter value (which Sugisaki 2007 shows does not fit well with other assumptions in a traditional Principles and Parameters setting).

Finally, if it is on the right track, the Superset bias may reveal another economy principle at work in language. But it is too soon to speculate in this direction. We must first have a better handle on the range of variation that is allowed. Only then will we be able to study the superset bias with all the care that is required to characterize how much positive counterevidence is necessary for the child to choose a more marked option, etc. The take-home message for now is that if nano-parameters are all we are left with in the context of minimalism, clustering effects – when they obtain! – will have to arise from something (or multiple things) external to narrow syntax. It seems to me that the learning process itself constitutes a natural place to look. It will certainly require closer collaboration between theoretical linguists and psycholinguists.

6. Concluding remarks, from a Biolinguistic perspective

This short paper has been very programmatic, and I ask the reader to treat it as such. Many of my suggestions are very tentative, and await much closer empirical investigation than I have been able to provide here. But one thing, I hope, is very clear: the traditional Principles-and-Parameters model does not fit snugly with minimalist intuitions. If minimalist ideas are on the right track, the standard take on what would count as a solution to Plato's Problem (parameter setting in the context of an overspecified UG) must be rethought from the ground up.

Although one finds relatively little explicit discussion of parametric variation in the minimalist literature (other than the convenient appeal to a parameter whenever two languages diverge), I hope to have shown here that one can distill a program for parameters from the minimalist literature – a minimalist program for parametric theory,

as it were. It is one that leaves very little room for parameters (i.e., points of variation) to affect narrow syntax. This is not due to the fact that empirically ‘syntactic’ effects of parameters are rare (the assumption in comparative syntax is that such effects are numerous and pervasive), but rather to the fact that syntactic parameters (i.e., parametrized principles) make no sense within minimalism.

I believe that the Strong Uniformity Thesis (SUT) in (2) emerges as a natural consequence of approaching UG from below, and with Galilean lenses. Consider, for example, Chomsky’s (2005) claim that three factors enter in the explanation of language design:¹⁷ (i) the genetic endowment (1st Factor), (ii) the environment (2nd factor), and (iii) generic ‘good design’ principles transcending the limits of genetics (3rd factor). The third factor has played a prominent role in minimalist research, and led to a much less specified view of the genetic endowment specific to language (UG). I have followed this trend here by suggesting that some effects formerly attributed to macroparameters may be due to a very general superset bias (economy guideline) relativized to the thing being learned (language). Invariably, as the importance of the 1st factor wanes, the interplay between the 2nd and the 3rd factors becomes more critical.

As I have argued elsewhere (see Boeckx 2006:chap. 4), I think that minimalist guidelines suggest an architecture of grammar that is more plausible biologically speaking than a fully specified, highly specific UG – especially considering the very little time nature had to evolve this remarkable ability that defines our species. If syntax is at the heart of what had to evolve *de novo*, syntactic parameters would have to have been part of this very late evolutionary addition. Although I confess that our intuitions pertaining to what could have evolved very rapidly are not as robust as one would like, I think that Darwin’s Problem (the logical problem of language evolution) becomes very hard to approach if a GB-style architecture is assumed. Within GB, there is no obvious answer to why variation exists, and why variation takes the form that it does. But with a considerably leaner vision of the faculty of language, variation becomes almost inevitable. If very little about language is specified genetically, and much of the core of language (narrow syntax) is the result of 3rd factor effects, variation emerges as the direct result of underspecification. It is because so little is specified about language in the genome that the varied, and ever-changing environment gives us variation in the externalized aspects of language. To take an obvious example, it is because Merge is symmetric that both head-first and head-last are possible options. Once a symmetric structure has to be broken by sheer force of the physics of speech, two options are logically available if no restriction has been pre-imposed.¹⁸ (A richer syntax could have specified a rigid head-complement order, as in Kayne 1994.)

This is not to say that we are back to the structuralists’ claim that anything goes across languages. Narrow syntax sets the limits of variation (no language will have ternary branching structures if binary branching is a 3rd factor effect; ditto for minimality, the size of phases, etc.), but within these limits, there will be variation. And there will be a lot of it, more than GB-practitioners expected, but probably as much as the detailed

¹⁷ See also Lewontin 2000 and Gould 2002 for similar views in a more general biological setting.

¹⁸ A syntactic parameter of the sort proposed by Saito and Fukui 1998 seems very redundant in the present context.

empirical work spearheaded by Richard Kayne continues to reveal (for a representative sample, see Manzini and Savoia 2005). Norbert Hornstein has pointed out (Hornstein in press) that there are many geometrical figures one can draw with a minimal set of tools like a straight edge and a compass, but there are some (e.g. triangles with 20° angles) that are impossible. The gappiness of the linguistic morphospace may have to be accounted in exactly these less specified, but no less deterministic terms.

We seem to have reached a stage where the current leading idea about how Darwin's Problem could be solved forces us to reconsider what we took to be our very best bet concerning Plato's Problem.

It is true that selectionism has now a distinguished history in biology and that parallelisms between a GB-style, richly-specified UG and current work in developmental evolutionary biology ('evo-devo') are not too hard to find.¹⁹ But one must bear in mind that the role of epigenetic factors in biology is on the rise (see Jablonka and Lamb 2005), and that the selectionism often invoked in evo-devo is in aid of understanding what Gould 1989 called disparity (fundamental design differences, as opposed to superficial diversity). One must bear in mind that in the context of language, we are dealing with a single design, a single organ, in a single species. With language, there is no disparity to speak of, only very superficial diversity. Evo-devo seeks to understand the diversity of design by revealing a design of diversity, but there is no diversity of design in language, as 50 years of hard work in generative grammar have revealed. Accordingly, the analogy between UG and the universal genome should not be pushed too far, certainly not in the direction of an overspecified UG, I think.

Inspired by recent developments in minimalism, I have sketched an approach that hopefully captures the outline of the nature of linguistic variation in a more adequate way. The line of argument advocated here provides an answer to why variation exists, and why there is so much of it (but so little in narrow syntax proper). It relates the answer to other current themes in biolinguistics that stresses the non-genomic character of nativism, and the formal simplicity of the language organ.

¹⁹ See already Piattelli-Palmarini 1986, and Jerne 1985. Perhaps the clearest parallelism can be found in the context of Sherman's 2007 Universal Genome proposal, where it is argued that descent with modification often takes the form of genetic losses). But there are other, less controversial claims in evo-devo that mirror GB-style analyses. For example, Arthur 2001 lists four ways in which developmental differences among organisms give rise to radically different adult morphologies:

- Heterochrony (different timing of gene expression)
- Heterotopy (different location of gene expression)
- Heterometry (more of the gene product being made)
- Heterotypy (change in the nature of the gene product, e.g., switch on different target genes)

It is not hard to find corresponding parameter types, such as:

- Is movement overt (pre-Spell-Out) or covert (post-Spell-Out)?
- Is the Agr-feature expressed on T or C?
- How many features does H check?/How many elements move to SpecHP?/How many heads are there in the left periphery?
- Is H strong or weak?

If the approach suggested here is on the right track, narrow syntax shows no sign of design compromise to assimilate variation, and thus provides another piece of evidence in favor of the PF-LF asymmetry recently emphasized by Chomsky. Variation is not structurally encoded (in the form of a parameter hierarchy); instead it is the result of the lack of structural encoding. It arises exactly where we expect it: not where 3rd factors reign (narrow syntax), but where the 2nd factor plays a prominent role (externalization).

Nevertheless, variation tends to show signs of optimization in its organization. As a matter of fact, in the present context, parameters emerge as a ‘mini-max’ solution (as Noam Chomsky and Massimo Piattelli-Palmarini have speculated): an attempt to navigate between the path of least genetic specification (minimal UG) and the path of least instruction (superset bias).

The few suggestions I made in this paper quickly reveal the vastness of the empirical research program that lies ahead. In many ways, we haven’t begun to solve Plato’s Problem (not even abstractly). Since the standard assumption in the field is that there is a lot of syntactic effects out there (the topic of the rich field of comparative syntax), and the main claim of this paper is that syntactic variation does not exist, many data points will have to be reanalyzed. I suspect that much of what we thought was syntax will turn out to be morpho-phonology (in the same way that the syntactic notion of timing of movement [overt vs. covert] has been replaced by a PF pronunciation algorithm that picks a certain copy in a movement chain). Like GB principles, GB parameters may turn out to be epiphenomena, to be understood as the result of deeper forces.

It is often said that minimalism has not engendered important empirical discoveries. Nothing could be further from the truth. If we find efficient design in language, that’s a surprising empirical discovery. Likewise, if we find a truly uniform syntax, if indeed it turns out that there is no syntactic variation at all, that too would be a startling empirical discovery, considering that only 50 years ago Martin Joos could write with assurance (Joos 1957) that the range of variation in this realm was virtually infinite.

References

- Arthur, W. 2001. *Biased embryos and evolution*. Cambridge: Cambridge University Press.
- Baker, M. 1996. *The polysynthesis parameter*. Oxford: Oxford University Press.
- Baker, M. 2001. *The atoms of language: the mind’s hidden rules of grammar*. New York: Basic Books.
- Baker, M. 2003. Linguistic differences and language design. *Trends in Cognitive Sciences* 7, 349-353.
- Baker, Mark. 2005. Mapping the terrain of language learning. *Language Learning and Language Development* 1, 93-124.
- Baker, M. 2008. *The syntax of agreement and concord*. Cambridge: Cambridge University Press.
- Baker, M. and Collins, C. 2006. Linkers and the internal structure of vP. *Natural Language and Linguistic Theory* 24:307-354.
- Bejar, S. 2003. Phi-syntax. Doctoral dissertation, University of Toronto.
- Bloomfield, L. 1933. *Language*. New York: Holt.

- Boeckx, C. 2006. *Linguistic minimalism: Origins, concepts, methods, and aims*. Oxford: Oxford University Press.
- Boeckx, C. 2008. Elementary syntactic structures. Ms., Harvard University.
- Boeckx, C. In progress. Linguistic Invariance and Language Variation. Ms., Harvard University.
- Borer, H. 1984. *Parametric syntax: case studies in Semitic and Romance languages*. Dordrecht: Foris.
- Borer, H. 2005. *Structuring sense (vols. 1 & 2)*. Oxford: Oxford University Press.
- Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, N. 1986. *Knowledge of Language*. New-York: Praeger.
- Chomsky, N. 1993. A minimalist program for linguistic theory. In *The view from Building 20*, ed. K. Hale and S. J. Keyser, 1-52. Cambridge, Mass.: MIT Press.
- Chomsky, N. 1995. *The minimalist program*. Cambridge, Mass.: MIT Press.
- Chomsky, N. 1998. Some observations on economy in generative grammar. In *Is the best good enough?*, ed. P. Barbosa, D. Fox, M. McGinnis, and D. Pesetsky, 115-129. Cambridge, Mass.: MIT Press.
- Chomsky, N. 2001. Derivation by phase. In *Ken Hale: a life in language*, M. Kenstowicz (ed), 1-50. Cambridge, Mass.: MIT Press.
- Chomsky, N. 2004. Beyond explanatory adequacy. In *Structures and beyond*, ed. A. Belletti, 104-131. Oxford: Oxford University Press.
- Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry* 36: 1-22.
- Chomsky, N. 2007. Approaching UG from below. In *Interfaces + Recursion = Language? Chomsky's minimalism and the view from syntax-semantics*, ed. U. Sauerland and M. Gaertner, 1-30. Mouton: de Gruyter.
- Chomsky, N. In press. On phases. In R. Freidin, C. Otero, and M.-L. Zubizarreta (eds), *Foundational issues in linguistic theory*. Cambridge, Mass.: MIT Press. [Page references to the 2005 version [Ms., MIT].]
- Chomsky, N. and M. Halle. 1968. *The sound pattern of English*. New York: Harper.
- Culicover, P. 1999. *Syntactic nuts*. Oxford: Oxford University Press.
- Darwin, C. 1859[1964]. *The origin of species*. Facsimile of 1st edition. Cambridge, Mass.: Harvard University Press.
- Fortuny, J. 2008. *The emergence of order in syntax*. Amsterdam: John Benjamins.
- Fukui, N. 1986. A theory of category projection and its applications. Doctoral dissertation, MIT.
- Fukui, N. 1995. The principles and parameters approach: a comparative syntax of English and Japanese. In *Approaches to language typology*, ed. M. Shibatani and T. Bynon, 327-372. Oxford: Oxford University Press.
- Fukui, N. 2006. *Theoretical comparative syntax*. London: Routledge.
- Gallego, A. 2007. Phase theory and parametric variation. Doctoral dissertation, Universitat Autònoma de Barcelona.
- Gallego, A. 2008a. Phases and variation. Exploring the second factor of the language faculty. Ms., Universitat Autònoma de Barcelona.
- Gallego, A. 2008b. The second factor of the language faculty: Phase architecture and linguistic variation. Talk given at the Harvard theory group, February 2008.
- Gallistel, R. 1990. *The organization of learning*. Cambridge, Mass.: MIT Press.
- Gallistel, R. 2006. Learning organs. Ms., Rutgers University.

- Gianollo, C., C. Guardiano, and G. Longobardi . In press. Historical implications of a formal theory of syntactic variation. In *Proceedings of DIGS VIII*, ed. D. Jonas and S. Anderson. Oxford: Oxford University Press.
- Goldberg, A. 2006. *Constructions at work*. Oxford: Oxford University Press.
- Gould, S. J. 1989. *Wonderful life*. New York: Norton.
- Gould, S. J. 2002. *The structure of evolutionary theory*. Cambridge, Mass.: Harvard University Press.
- Guardiano, C. and G. Longobardi. 2005. Parametric comparison and language taxonomy. In *Grammaticalization and parametric variation*, ed. M. Batllori, M.-L. Hernanz, C. Picallo, and F. Roca, 149-174. Oxford: Oxford University Press.
- Hale, K. and S. J. Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In *The View from Building 20*, ed. K. Hale and S. J. Keyser, 53-110. Cambridge, Mass.: MIT Press.
- Hale, K., and S. J. Keyser. 2002. *Prolegomena to a theory of argument structure*. Cambridge, Mass.: MIT Press.
- Halle, A. and M. Marantz. 1993. Distributed morphology and the pieces of inflection. In *The View from Building 20*, ed. K. Hale and S. J. Keyser, 111-176. Cambridge, Mass.: MIT Press.
- Harley, H. 2004. Merge, conflation, and head movement: The First Sister Principle revisited. In *Proceedings of NELS 34*, 239-254. University of Massachusetts, Amherst: GLSA.
- Hornstein, N. 2001. *Move! A minimalist theory of construal*. Oxford: Blackwell.
- Hornstein, N. To appear. *A theory of syntax*. Cambridge: Cambridge University Press.
- Hyams, N. 1986. *Language acquisition and the theory of parameters*. Dordrecht: Reidel.
- Jackendoff, R. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Jablonka, E. and M. Lamb. 2005. *Evolution in four dimensions*. Cambridge, Mass.: MIT Press.
- Jerne, N. K. 1985. The generative grammar of the immune system. *Science* 229, 1057-1059.
- Joos, M., ed. 1957. *Readings in linguistics*. New York: American Council of Learned Societies.
- Kayne, R. 1994. *The antisymmetry of syntax*. Cambridge, Mass.: MIT Press.
- Lewontin, R. 2000. *The triple helix*. Cambridge, Mass.: Harvard University Press.
- Longobardi, G. 2003. Methods in parametric linguistics and cognitive history. *Linguistic Variation Yearbook* 3, 101-138.
- Longobardi, G. 2005. A minimalist program for parametric linguistics? In *Organizing grammar*, ed. H. Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz, and J. Koster, 407-414. Berlin: Mouton/de Gruyter.
- Manzini, M. R. and L. Savoia. 2005. *I dialetti italiani e romanci (3 vols)*. Alessandria: Edizioni dell'Orso.
- Marantz, Alec. (1995). The minimalist program. In G. Webelhuth (ed), *Government and Binding and the Minimalist Program*. Oxford: Blackwell, 349-382.
- Marantz, A. 2000. Words. Ms., MIT.
- Matthewson, L. 2006. Presuppositions and cross-linguistic variation. Ms., University of British Columbia.

- Matthewson, L. 2008. Pronouns, presuppositions, and semantic variation. Paper presented at SALT 18.
- Matushansky, O. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37, 69-109.
- Newmeyer, F. 2005. *Possible and probable languages*. Oxford: Oxford University Press.
- Nicolis, M. 2006. The null subject parameter and correlating properties: The case of Creole languages. Ms., University of Siena.
- Pearl, L. 2007. Necessary bias in language learning. Doctoral dissertation, University of Maryland.
- Piattelli-Palmarini, M. 1989. Evolution, selection and cognition: From “learning” to parameter setting in biology and in the study of language. *Cognition* 31, 1-44.
- Raposo, E. 2002. Nominal gaps with prepositions modifiers in Portuguese and Spanish: A case for Quick Spell-Out. *Cuadernos de Linguística del U. Ortega y Gasset* 9, 127-144.
- Richards, M. 2007. On feature-inheritance: An argument from the Phase Impenetrability Condition. *Linguistic Inquiry* 38: 563-572.
- Rizzi, L. 1982. *Issues in Italian syntax*. Dordrecht: Foris.
- Rizzi, L. 2007. Grammatically-based target-inconsistencies in Child Language. *Proceedings of the Inaugural GALANA conference, UConn Occasional Papers in Linguistics #4*, 19-49. Cambridge, Mass.: MITWPL.
- Roberts, I. and A. Roussou. 2002. *Syntactic change*. Cambridge: Cambridge University Press.
- Roberts, I. 2007. *Diachronic syntax*. Oxford: Oxford University Press.
- Saito, M., and N. Fukui. 1998. Order and Phrase Structure and movement. *Linguistic Inquiry* 29: 439-474.
- Sherman, M. 2007. Universal Genome in the origin of Metazoa. *Cell Cycle* 6, 1873-1877.
- Sugisaki, K. 2007. A note on the default value of parameters. *Biolinguistics* 1, 114-117.
- Svenonius, P. and G. Ramchand. 2007. Mapping a parochial lexicon onto a universal semantics. Ms., University of Tromsø.
- Vicente, L. 2007. The syntax of heads and phrases. Doctoral dissertation, University of Leiden.
- Webelhuth, G. 1991. Principles of syntactic saturation. Oxford: Oxford University Press.
- Yang, C. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.
- Yang, C. 2004. Universal Grammar, statistics, or both? *Trends in Cognitive Science* 8, 451-455.
- Yang, C. 2005. On productivity. *Language Variation Yearbook* 5, 333-370.