

John Benjamins Publishing Company



This is a contribution from *Linguistic Variation Yearbook 4*
© 2004. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Against a parameter-setting approach to typological variation

Frederick J. Newmeyer
University of Washington

The dominant position among generative grammarians with respect to typological variation is that it should be captured by parameters, which are either directly tied to principles of Universal Grammar (UG) or to functional projections provided by UG. Parameter-setting approaches, however, have failed to live up to their promise. They should be replaced by a model in which language-particular rules take over the work of parameter settings and in which most typological variation follows from independently-needed principles of performance. In such a model, UG specifies the class of possible languages, but not the set of probable languages.

Keywords: parameter, processing, rule, typology, Universal Grammar

1. The parametric approach to language variation

My goal in this paper to contrast two approaches to typological variation in grammar within the general envelope of formal theory. In one approach, variation is captured largely by means of parameters, either directly tied to principles of Universal Grammar (UG) or to functional projections provided by UG. In the other approach, variation is captured by means of extragrammatical principles. I will argue that the second approach is better supported than the first.

The remainder of §1 presents the principal arguments that have been adduced in support of the idea of parameters. Section 2 makes the case that these arguments are defective. Put simply, there is no evidence that parametric approaches are any more successful than rule-based approaches. In §3 I argue that the burden of accounting for crosslinguistic generalizations should be shifted away from competence and to performance. Section 4 is a brief conclusion.

1.1 Two contrasting approaches for capturing typological variation in language

Since Chomsky (1981), the dominant position among generative grammarians with respect to the overall structure of linguistic theory has been something along the lines of (1):

- (1) Some central features of (currently predominant) linguistic theory:
 - a. Principles of Universal Grammar (or, more recently, a set of functional projections provided by UG), which have
 - b. different parameter settings for different languages (thereby accounting for language-particular differences).
 - c. By means of a. and b., typological variation is accounted for.
 - d. A residue of marked (language-particular) morphosyntactic properties.

The following two quotes from Chomsky combine to illustrate (1a–d). The first maintains that the principles of UG and their possible settings are innate, while allowing for a (presumably learned) residue of marked exceptions. The second makes it clear that accounting for typological variation among languages is part of the role of UG.

[W]hat we ‘know innately’ are the principles of the various subsystems of S_0 [= the initial state of the language faculty — FJN] and the manner of their interaction, and the parameters associated with these principles. What we learn are the values of these parameters and the elements of the periphery ... The language that we then know is a system of principles with parameters fixed, along with a periphery of marked exceptions. (Chomsky 1986: 150–151)

There has also been very productive study of generalizations that are more directly observable: generalizations about the word orders we actually see, for example. The work of Joseph Greenberg has been particularly instructive and influential in this regard. *These universals are probably descriptive generalizations that should be derived from principles of UG.* (Chomsky 1998: 33; emphasis added)

An alternative position is presented in (2):

- (2) Some central features of an alternative way of looking at linguistic theory:
 - a. Unparameterized principles of Universal Grammar.
 - b. Language-particular rules.
 - c. Extragrammatical principles that account for typological variation.

Under position (2), neither UG principles nor functional categories are associated with a range of parameter settings. Rather, language-particular differences are captured by differences in language-particular rules. Accompanying this view, the burden of accounting for typological generalizations is shifted from competence to performance.

The purpose of this paper is to defend position (2), while pointing out the deficiencies of position (1). But before proceeding further, it is crucial that I specify what I mean by 'language-particular rule'. I have two rather different types of devices in mind. First, and less interestingly, the notion encompasses the formal mechanisms devised to handle phenomena whose explanation seems to lie outside the scope of the core principles of Principles-and-Parameters syntax (from whatever period). In a GB approach, then, language-particular rules would encompass processes assigned to the marked periphery like Exceptional Case Marking, deletions like VP Ellipsis, 'stylistic' and phonology-sensitive processes like Heavy-NP-Shift, the residue of idiosyncratic filters, and so on. Given the still-inchoate state of much minimalist theorizing, it is more difficult to characterize what a language-particular rule would be in that general approach. Candidates include XP adjunctions such as extraposition, right-node raising, VP-adjunction, and scrambling and processes that might apply in the 'P-syntax', such as those sensitive to linear order and those sensitive to phonological conditions.

The second sense of 'language-particular rule' is unique to the present work. Essentially, they are parameter-settings 'detached' from the parameters themselves (which are hypothesized not to exist). So it is frequently assumed that UG provides a 'Head Parameter' like (3), with settings (also provided by UG) like (3a) and (3b) or else (in the MP) linearization statements, parameterized as in (4a–b):

- (3) HEAD PARAMETER: Complements are to the left or to the right of the head.
 - a. HEAD-LEFT (English, Swahili, ...)
 - b. HEAD-RIGHT (Japanese, Lakhota, ...)
- (4) a. For a syntactic object $\{\alpha, \beta\}$, if α is a head and β is a complement, then α precedes β . (English, Swahili, ...)
- b. For a syntactic object $\{\alpha, \beta\}$, if α is a head and β is a complement, then α follows β . (Japanese, Lakhota, ...)

In my alternative view, UG would still specify that phrases are headed (and hence in the process of acquisition children would still be driven to identify the

position of the head). However, what they would acquire for each of the four languages mentioned would simply be a rule along the lines of (5a–d):

- (5) a. English: Complements are to the right of the head.
- b. Swahili: Complements are to the right of the head.
- c. Japanese: Complements are to the left of the head.
- d. Lakhota: Complements are to the left of the head.

Now consider a broad UG principle like Subjacency. In the traditional GB view, Subjacency is stated something along the lines of (6), with parameter settings as in (6a–c):

- (6) SUBJACENCY: No moved element may cross two or more bounding nodes, with the choice of bounding nodes parameterized as follows:
 - a. English chooses S.
 - b. Italian chooses S'.
 - c. Russian chooses S and S'.

In my alternative way of looking at things, Subjacency could still exist as a UG principle, but would be stated much more broadly, as in (7):

- (7) SUBJACENCY': There are limits on the structural distance between the launching and landing site of a moved element.

Move- α , then, (if we make a set of assumptions that otherwise characterize GB), would be formulated somewhat differently in English, Italian, and Russian. However, in no language would Move- α be totally 'free'. That is, given the UG principle of Subjacency, Move- α will always be bounded to some degree or other. Similarly, given the approach to bounding compatible with minimalist assumptions, namely that such effects are derived from the number and properties of the specifiers of CP and IP (as in Richards 2001), in my alternative conception the component of the grammar specifying phrase structure would simply have different rules expanding CP and IP.

To take another example, it has been assumed since the early 1980s that an important parameter separating languages is whether *Wh*-Movement takes place in the (narrow) syntax or in LF (after Spell-Out). In the view adopted here, grammars are simply formulated so that some contain a rule of overt movement and others a rule of covert movement. If the choice of overt or covert movement can be derived from some other feature of the language, then well and good, but, again, there is no UG-provided specification that languages have some global property, or some local property tied to a feature of a functional head, which divides them into overt-movers and covert-movers.

One's first reaction to these remarks might well be that I am advocating a 'less constrained' grammatical theory than advocates of Principles-and-Parameters syntax have generally assumed, given that in my alternative view, UG has a diminished role to play. To a certain extent, such is true. However, I have two points to make in defense. The first is that the degree to which a theory can be constrained is itself constrained by empirical reality. And empirical reality, as I see it, dictates that the hopeful vision of UG as providing a small number of principles each admitting of a small number of parameter settings is simply not workable. The variation that one finds among grammars is far too complex for such a vision to be realized. The second is that the degree of grammatical variation is in fact highly constrained, but much more by performance factors than by UG. Hopefully, the truth of both points will become apparent as the paper unfolds.

1.2 The 'standard story' on the superiority of the parametric approach

It should prove useful to frame the discussion by outlining the standard story on the differences between parameters and rules, as in (8):

- (8) The standard story on parameters (and how they contrast with rules):
- a. Parameters are descriptively simple, whereas rules are (generally) not.
 - b. Parameters have binary settings (an idea which is inapplicable to rules).
 - c. Parameters are small in number; the number of rules is open-ended.
 - d. Parameters are hierarchically/implicationally organized, thereby accounting for both order of first language acquisition and typological generalizations (there is nothing comparable for rules).
 - e. Parameters are abstract entities with a rich deductive structure, making possible the prediction of (unexpected) clusterings of morphosyntactic properties.
 - f. Parameters and the set of their possible settings are innate (and therefore universal). Rules are not (normally) assumed to be drawn from an innate set.
 - g. Parameter settings are easily learned, while rules are learned with much greater difficulty.
 - h. Parametric change is markedly different from rule-based change (such as grammaticalization and morphological change).

Descriptive simplicity, (8a), is illustrated by the Head Parameter and by Subjacency, referred to above, which have been said to require no more than terse

elegant statements such as (3) or (4) and (6). Other parameters, say, those fixing whether a language is null subject or not, configurational or not, and so on, have been thought to be equally simple.

Binarity, claim (8b), is generally assumed by practitioners of Principles-and-Parameters syntax, or is at least held up as a desideratum: ‘The values of each parameter form no continuum and in the ideal case just amount to two’ (Longobardi 2003: 108). As far as smallness of number, (8c), is concerned, I am aware of no estimate that exceeds a few dozen. Along these lines, Kayne remarks that ‘the number of independent binary-valued syntactic parameters needed to allow for 5 billion syntactically distinct grammars [i.e., a different one for each person on earth — FJN] is only 33 (2 raised to the 33rd power is about 8.5 billion) ... it seems plausible that the child is capable of setting at least that many syntactic parameters’ (Kayne 2000: 8). Smallness of number serves more than just aesthetic purposes. Fodor (2001b: 734) observes that ‘it is standardly assumed that there are fewer parameters than there are possible rules in a rule based framework; otherwise, it would be less obvious that the amount of learning to be done is reduced in a parametric framework.’

A parameter-setting model suggests a way of incorporating typological generalizations directly into grammatical theory (8d). Since there is a general tendency for languages to place complements on the same side of the head, all it takes is one parameter to capture this typological generalization, namely (3) or (4), in which all complements either uniformly precede or uniformly follow their heads. Other typological generalizations have been argued to follow from implicational relations between parameter settings (see especially Baker 2001). For example, verb-final languages tend not to allow *wh*-fronting (Dryer 1991). To capture this fact, it has been assumed that an implicational statement like (9a) is part of UG, or perhaps a more abstract implication along the lines of (9b):

- (9) a. HEAD-RIGHT \supset NO *WH*-FRONTING
 b. HEAD-RIGHT \supset ‘PARAMETER X’, (where the consequence of this parameter is that *wh*-fronting is impossible)

Languages in which all heads do not uniformly precede or uniformly follow their complements of course do exist, as do languages with heads on the right and *wh*-fronting. Therefore, any system of parameterized principles has to be flexible enough to allow for them. The usual assumption has been that typological rarity is a reflex of ‘less-than-fully-optimal’ grammars. That is, typologically rare grammars are more complex than grammars of typologically

consistent languages, perhaps by requiring special marked parameter settings, by violating an implicational statement among parameter settings, or whatever. For example, Chinese is consistently head final except in the rule expanding X' to X^0 (if the head is verbal it precedes the complement). Note that Chinese manifests the ordering V–NP, but NP–N:

- (10) a. you sange ren mai-le shu
 HAVE three man buy-ASP book
 ‘Three men bought books’ (Huang 1982)
 b. Zhangsan de sanben shu
 Zhangsan DE three book
 ‘Zhangsan’s three books’

Travis (1989) suggests that Chinese has a marked parameter setting for word order. Normally, if a language is head final, it assigns Case and Theta-Role to the left, as in (11a). However Chinese has a special setting that violates this default ordering, namely (11b):

- (11) a. Unmarked setting: HEAD-RIGHT \supset THETA-ASSIGNMENT TO LEFT & CASE-ASSIGNMENT TO LEFT
 b. Marked setting (Chinese): HEAD-RIGHT & THETA-ASSIGNMENT TO RIGHT & CASE-ASSIGNMENT TO RIGHT

As far as language acquisition is concerned, a common (but by no means universal) assumption has been that children are born with the unmarked combination of parameter settings: ‘We would expect the order of appearance of structures in language acquisition to reflect the structure of markedness in some respects...’ (Chomsky 1981: 9; see also Lebeaux 1987, Williams 1981: 25; Rizzi 1986). Learning a typologically rare language might therefore involve extra time, extra effort, or both.

The putatively rich deductive structure of parameters (8e) has been most investigated with respect to the Null Subject Parameter. In the original formulations of this parameter, those found in Chomsky (1981), it was predicted that null subject languages should manifest the (not obviously related) properties in (12) (null subject Spanish is contrasted with non-null subject English):

- (12) a. missing subjects (*Llueve* ‘It is raining’)
 b. free inversion in simple sentences: *Leyó el libro María* ‘Read the book Mary’
 c. long *wh*-movement of subject across *wh*-islands (*el hombre que me pregunto a quién vio* ‘the man who I wonder whom (he) saw’)

- d. empty resumptive pronouns in embedded clauses (*ésta es la muchacha [que me pregunto [quién cree [que ___ puede ...]]]* ‘this is the girl who I wonder who thinks that (she) may ...’)
- e. apparent violations of the *that*-trace filter (*¿Quién dijiste que salió temprano?* ‘Who did you say (*that) left early?’)

The idea that parameters and their set of possible settings are innate, (3f), is intrinsic to the parametric view of language variation. Here is a recent and particularly clear statement to that effect:

Children instinctively work their way down the [parameter] hierarchy, taking advantage of its logical structure to avoid agonizing over needless decisions. ...The hierarchy provides a logical flowchart that children use in the process of language acquisition (Baker 2001: 192; 195)

Going along with such an idea is (8g), the easy learnability of the settings themselves for a particular language, which in principle allow language learning to be a ‘simple quiz’ (Fodor 2001b: 734).

And, finally, in a number of publications, Lightfoot (1991, 1999) has made claims about parametric change (8h) that distinguish it from rule-based change. The former involves changes of parameter settings and the latter includes grammaticalization, morphological changes, such as the loss of gender markers (Jones 1988), the reduction in verbal desinences, and the loss of subjunctive mood. He points to six features that are characteristic of parametric change: each new grammatical property is manifested by a cluster of new phenomena; the new property sets off a ‘chain reaction’ of further changes; it spreads rapidly and manifests an S-curve; an earlier structural property becomes obsolescent; significant change in meaning occurs; and (as a prior condition) the change was triggered by shifts in unembedded data only. Among examples of parametric change in the history of English, Lightfoot includes the change from OV to VO order; the ability of the infinitival *to* marker to transmit case-marking and head-government properties of a governing verb; the loss of inherent D-structure oblique case; the reanalysis allowing the stranding of prepositions; the recategorization of the premodal verbs; and the loss of the ability of verbs to move to a governing INFL position.

One can readily understand the appeal of the conjunction of hypotheses (8a–h). If they are correct, then the parametric view of language is far-ranging in explanatory power. However, the remainder of this paper will argue that they are not correct. Taking these hypotheses in order, I will argue that the facts are at least as supportive of a rule-based account of variation (2a–c) as of a parametric account (1a–d).

2. The deficiencies of parameter-setting models

This section will take each argument in turn that has been presented in support of parameter-setting model and point out its deficiencies. I will attempt to demonstrate that in all cases, a rule-based account is either more adequate than a parameter-based one or that, when all the facts are taken into account, they are empirically indistinguishable.

2.1 Descriptive simplicity

Let us begin by looking at the question of parameters and rules simply from the point of view of formal description. I take it as uncontroversial that parameters are motivated only to the extent that they lead overall to more formal simplicity. If as many parameter settings are needed in a parameter-based model as rules would be needed in a rule-based one and the former turn out to be as complex as the latter, then clearly nothing is gained by opting for parameters.

In fact, I see no reason to think that there is any gain in descriptive simplicity with parameters. Let us begin with the question of adjective-noun ordering. Consider the following differences between French and English (as presented in Bouchard 2003):

- (13) a. un **gros** ballon **rouge**
b. a **big** red ball
- (14) a. un tissu **anglais** **cher**
b. an **expensive** English fabric
- (15) a. an **old** friend (= friend who is aged or friend for a long time)
b. une **vieille** amie (= friend for a long time)
c. une amie **vieille** (= friend who is aged)

In the account presented in Cinque (1994), the facts of (13)–(15) result from the following parametric differences between French and English:

- (16) a. French has postnominal adjectives (as in 13a) because of a parametric difference with English that allows N-movement in the former language, but not in the latter.
- b. *Cher* has scope over *anglais* in (14a) because French has a parametric difference with English that triggers movement of a N-ADJ constituent.
- c. In (15), the two positions for *vieille* in French, but only one for *old* in English, results from a parametric difference between the two

languages regarding the feature attraction possibilities of functional categories in the two languages.

The problem with such an account, as noted by Bouchard, is that the word ‘parameter’ is used as nothing more than a synonym for the word ‘rule’. There is no increase in descriptive elegance, economy, or whatever in Cinque’s account over an account which does no more than say that English and French have different rules of adjective placement.

In any number of cases a parametric approach seems simpler than a rule-based one only because no one has ever attempted to capture the full complexity of any one language by means of them. Consider English word order. English is not uniquely SVO. As we can see in (17), other orders of subject, verb, and object are possible:

- (17) a. The last lecture Mary really hated. (OSV)
- b. Drink the whole bottle, John never would. (VOS)
- c. Away ran John. (VS)

In a rule-based model, presumably English would have an underlying SVO order and the sentences of (17) would be generated by means of optional movement rules. But the point is that simply characterizing English as manifesting the parameter ‘head-first’ is not sufficient. Either a separate ‘parameter’ would have to be posited to allow for (17a–c), effectively using the word ‘parameter’ as a not very enlightening synonym for the word ‘rule’, or their derivation would have to be consigned to the ‘marked periphery’. Neither alternative results in any gain in overall simplicity.

In general, parametric accounts have been vague on which phenomena in the language under discussion need to be handled by separate marked rules. An exception is the discussion of Hixkaryana in Kayne (1994) and Baker (2001). This language for the most part manifests OVS word order, an order which is extremely rare typologically:

- (18) kanawa yano toto
 canoe took person
 ‘the man took the canoe’ Hixkaryana (Derbyshire 1985)

One’s first thought might be that what is needed is a parameter allowing for OVS order. But in fact Kayne and Baker reject the idea that a special word order parameter is involved here. Rather Hixkaryana is (parametrically) SOV and allows the fronting of VP by a movement rule (the language does in fact have SOV as variant order):

(19) S[OV] → [OV]S

In other words, in this account word order is determined both by a parameter and a language-specific rule. Such a treatment undercuts the attractiveness of a parametric approach. That is, it is no longer a matter of comparing a theory with parameters (and all their virtues) with a theory with rules (and their lack of virtues). Rather, it is a matter of comparing a theory with parameter settings *and* rules versus one with rules alone. On the basis of Occam's razor, one would be forced to renounce the idea of any a priori desirability of a parametric theory.

The dilemma here is that if the scope of parameters is expanded (to include the English and Hixkaryana cases), then 'parameter' has simply become a synonym for 'rule' (a danger noted long ago by Safir 1987: 78). If parameters are kept simple and few in number, then grammatical theory needs to characterize the properties of rules, just as it would have to in a purely rule-based approach.¹

2.2 Binariness

In general, where features have been posited in morphosyntax, they have been assumed to be binary, just as in phonology. But binarity in phonology has a conceptual basis, in that it is rooted in categorical perception (Richard Wright, personal communication). It goes without saying that such an account does not generalize to the morphological or syntactic properties of language.

In fact, there is little evident binarity in morphosyntax. The number of genders, numbers, and cases in a particular language might be two, but it might be more or less than that (one thinks of the heroic, but failed, attempt in Jakobson (1936/1971) to reduce the cases of Russian to sets of binary distinctions). Along the same lines, there is little reason to think of grammatical relations, thematic roles, and so on as lending themselves to characterization in terms of binary oppositions. Where binarity has been posited in syntax, it has not been very successful. It is generally assumed that the categories N, A, P, and V are decomposed into binary features, as in Table 1.²

Table 1. Categories and features (Chomsky 1970)

	+N	-N
+V	A	V
-V	N	P

But as Baker (2003: 2) has astutely observed, ‘this theory is widely recognized to have almost no content in practice. The feature system is not well integrated into the framework as a whole, in that there are few or no principles that refer to these features or their values’. Indeed, there is little reason to believe that this or any pair of binary feature specifications for the four major categories could be well motivated. To begin with, *any* such breakdown necessarily makes the very weak claim that four of the six combinations involving two categories are ‘natural’; in the system illustrated in Table 1, only the pairs A-P and V-N are unnatural. And yet, it is difficult to see what might be unnatural about either of them. In the course of the history of English, for example, we have seen adjectives become prepositions (e.g. the word *like*), and, as pointed out in Baker (2003), APs and PPs can be appended to a transitive clause to express the goal or result of the action, but NPs and VPs cannot:

- (20) a. John pounded the metal flat. (AP)
 b. John threw the ball into the barrel. (PP)
 c. *John pounded the metal a sword. (NP)
 d. *John polished the table shine. (VP)

Likewise, gerunds crosslinguistically manifest both N-related properties and V-related properties to varying degrees. Even more problematically, there are numerous attested generalizations involving three out of the four categories: P, A, and V are Case assigners in some languages, and, in English, N, A, and P, but not V, allow QP specifiers (for more examples along these lines, see Reuland 1986 and Muysken and van Riemsdijk 1986 and for detailed criticism of the feature system for categories, see Stuurman 1985 and Déchaine 1993):

- (21) a. Whose [_Nbook] is that?
 b. How [_Ared] was the sunset?
 c. How much [_Pover] the limit are we?
 d. *How fast [_Vcycling] was Lance Armstrong?

As far as parameters are concerned, some have indeed been assumed to have binary settings, such as those that determine whether a language is configurational or not, whether it has (overt) *wh*-fronting or not, and so on. But many others are not (evidently) binary. So for example, the possible binding domains across languages have been argued to be in a subset relation with respect to each other, rather than contrasting in binary fashion:

- (22) Domains for binding of anaphoric elements (Manzini and Wexler 1987)
 γ is a governing category for α iff

- γ is the minimal category that contains α and a governor for α and
- a. can have a subject or, for α anaphoric, has a subject β , $\beta \neq \alpha$; or
 - b. has an Infl; or
 - c. has a Tense; or
 - d. has a 'referential' Tense; or
 - e. has a 'root' Tense;
- if, for α anaphoric, the subject of γ is accessible to α .

Likewise, there is nothing (evidently) binary about the possible bounding nodes for Subjacency, about the number of specifiers a DP can take, or about any number of other syntactic properties that might distinguish one language from another.

In a sense, binarity was built into the early Minimalist Program (MP) by means of the (binary) distinction between strong and weak features. Languages differ in terms of the binary feature specifications of functional heads; if the relevant feature of the head is strong, there is overt movement, if the feature is weak, then there is movement in LF. But the problem is that it has never been clear how to relate many parametric differences among languages to the distinction between strong and weak features (one thinks of binding and bounding domains again, for example). And thinking of parametric choices as binary choices is less evident in current instantiations of the MP, where notions like 'interpretable' and 'uninterpretable' have replaced 'strong' and 'weak'. Perhaps each functional head is associated with two features, one interpretable and one uninterpretable, but it is not obvious why such should be the case.

In short, parameters do not clearly have binary settings, and in that respect parameter settings do not differ from rules.

2.3 Smallness of number

How many parameters are there? The answer depends, of course on what is in the marked periphery. Certainly hundreds have been proposed since the notion was introduced around 1980. David Lightfoot has written that 30–40 are a reasonable number from the perspective of language acquisition, but goes on to make the disconcerting comment that 'a single issue of *Linguistic Inquiry* may contain 30–40 proposed parameters' (Lightfoot 1999: 259). Such has been the norm since the introduction of parametric theory. Consider the papers collected in *NELS-15* in 1985, several of which propose parameters of variation. Some of these in fact do seem fairly general, including one distinguishing nominative/accusative languages from ergative/absolute languages.

But others have the appearance of being uncomfortably language-particular, including one that states that Finnish is immune to the Case Filter; one which has *Wh*-Movement pass through INFL in Yoruba; and one that states that a preposition must be properly governed in Dutch in order to be a proper governor itself.

If we can (for purposes of argument) make the assumption that there is one binary parameter setting for each functional head, we need 'simply' count the number of functional heads to determine the number of parameters. And how many functional heads are there? If Cinque (1999) is right, there are at least 32 functional heads in the IP domain alone (possibly multiplied by the number of verb forms, i.e., finite/infinitive/past participle, etc.):

(23) Functional heads in the IP domain (Cinque 1999):

- | | | |
|-----------------------------------|--------------------------------------|---------------------------------------|
| a. Mood _{Speech Act} | b. Mood _{Evaluative} | c. Mood _{Evidential} |
| d. Mood _{Epistemic} | e. T(Past) | f. T(Future) |
| g. Mood _{Irealis} | h. Mod _{Necessity} | i. Mod _{Possibility} |
| j. Mod _{Volitional} | k. Mod _{Obligation} | l. Mod _{Ability/permission} |
| m. Asp _{Habitual} | n. Asp _{Repetitive(I)} | o. Asp _{Frequentative(I)} |
| p. Asp _{Celerative(I)} | q. T(Anterior) | r. Asp _{Terminative} |
| s. Asp _{Continuative} | t. Asp _{Perfect(?)} | u. Asp _{Retrospective} |
| v. Asp _{Proximative} | w. Asp _{Durative} | x. Asp _{Generic/progressive} |
| y. Asp _{Prospective} | z. Asp _{SgCompletive(I)} | aa. Asp _{PlCompletive} |
| bb. Voice | cc. Asp _{Celerative(II)} | dd. Asp _{SgCompletive(II)} |
| ee. Asp _{Repetitive(II)} | ff. Asp _{Frequentative(II)} | |

On the basis of a look at fifteen languages, fourteen of them Indo-European (from only four subfamilies), Longobardi (2003) proposes 30 binary parameters for DP. Cinque (1994) divides Adjective Phrase into at least five separate maximal projections encoding Quality, Size, Shape, Color, and Nationality. Beghelli and Stowell (1997) break down Quantifier Phrase into projections headed by *Wh*, Neg, Distributive, Referential, and Share. CP has also been split into a dozen or more projections, including ForceP, FocusP, and an indefinite number of Topic Phrases (Rizzi 1997). Facts pertaining to clitic inversion and related phenomena in some northern dialects of Italian have led to the positing of Left Dislocation Phrase, Number Phrase, Hearer Phrase, and Speaker Phrase (Poletto 2000). Damonte (2004) proposes projections corresponding to the set of thematic roles, including Reciprocal, Benefactive, Instrumental, Causative, Comitative, and Reversive Phrases. We have seen Verb Phrase split into two projections, one headed by V and the other by 'v' (Chomsky 1995). Zanuttini

(2001) posits four distinct Negative Phrase projections for Romance alone and McCloskey (1997) argues that at least three subject positions are needed.

Other proposals have led to a potentially exponential increase in the number of functional projections and their interrelationships, and hence in the number of parameters. For example, Giorgi and Pianesi (1997) have mooted the possibility of 'syncretic categories', that is, those that conflate two or more otherwise independent ones, as, for example, TP/AgrP. Along similar lines, Bobaljik (1995); Thráinsson (1996); and Bobaljik and Thráinsson (1998) suggest that languages differ not only by the settings of their parameters, but by the very existence in a particular language of the functional categories in which they are situated (see also Fukui 1995). Such a proposal leads to at least a ternary value for each parameter: positive, negative, or not applicable. Complicating things still further, Ouhalla (1991a) argues that an important dimension of parametric variation among languages is the relative ordering of embedding of functional categories. So for example, in his analysis, in Berber and Chamorro, the AgrP projection is below the TnsP projection, while in English and Welsh, TnsP is below AgrP.

One might, of course, argue along with Cinque and contra Ouhalla that the ordering among functional categories is universal. In that view, languages would differ parametrically in their lexicalization possibilities, some functional categories being lexicalized in some languages, but not in others. However, transferring the parametric choice to the lexicon neither decreases the number of potential parameters nor gives them an edge over rules. First, the number of parameters is not reduced, since the burden of specifying whether a functional category is present in a particular language or not has merely been transferred to the lexicon. Second, the statement that some language makes the parametric choice that lexical item L licenses functional projection P is indistinguishable from the statement that there is a language-particular rule involving L that specifies P.

Kayne, as noted above, has written that, given the existence of microparameters which might even differentiate the grammars of two individuals within the same speech community, 'some number substantially greater than 5 billion' grammars might exist in the world (Kayne 2000: 8). To handle this number, Kayne calculates that only 33 binary-valued parameters would be needed. His math may be right, but from that fact it does not follow that only 33 parameters would be needed to capture all of the microvariation that one finds in the world's languages and dialects. In principle, the goal of a parametric approach is to capture the set of *possible* human languages, not the set (however large)

of actually existing ones. One can only speculate that the number of such languages is in the trillions or quadrillions. In any event, Kayne's own work suggests that the number of parameters is vastly higher than 33. Depending on precisely what counts as a parameter (Kayne is not always clear on that point), just to characterize the difference among the Romance dialects discussed in the first part of Kayne (2000) with respect to clitic behavior, null subjects, verb movement, and participle agreement would require several dozen distinct parameters. It is hard to avoid the conclusion that characterizing just a few more differences among the dialects would lead to dozens of new parameters.

If the number of parameters needed to handle the different grammars of the world's languages, dialects, and (possibly) idiolects is in the thousands (or, worse, millions), then ascribing them to an innate UG to my mind loses all semblance of plausibility. True, we are not yet at the point of being able to 'prove' that the child is not innately equipped with 7846 (or 7,846,938) parameters, each of whose settings is fixed by some relevant triggering experience. I would put my money, however, on the fact that evolution has not endowed human beings in such an exuberant fashion.

Nobody is in a position say how many parameters are necessary in a parametric account, any more than one is in a position to say how many rules are necessary for any given language in a rule-based approach. But I see no reason to conclude that the number of the former is significantly less than the number of the latter.

2.4 Hierarchical/implicational organization

Given the importance of parameters in the P&P approach, one might assume that the pages of the generative-oriented journals would be filled with articles devoted to working out the relevant implicational relations among parameters with the ultimate goal of deriving the more robust generalizations that have been uncovered in the past few decades of typological investigation. Nothing could be farther from the truth, however. Many articles and books posit a particular parameter (and associated settings) to distinguish one language or dialect from another and most introductions to transformational syntax devote a few pages (but rarely more) to how crosslinguistic generalizations might be captured. But with one exception, researchers have not attempted a comprehensive treatment of parameters and their settings. That exception is Mark Baker's book *The Atoms of Language: The Mind's Hidden Rules of Grammar* (Baker 2001).³ Baker takes seriously the P&P program for typology and proposes an

intricate ‘Parameter Hierarchy’ (PH) in which implicational relations between parameters and their settings are made explicit. Figure 1 presents his final version of that hierarchy.⁴

The PH is to be interpreted as follows. Each of the bold-faced expressions are particular parameters, while their settings are in small capitals. If Parameter X has logical priority over Parameter Y, then X is written higher than Y and is connected to Y by a downward slanting line. If two parameters are logically independent of each other, then they are written on the same line and separated by a dash. Such is the case only for the Head Directionality Parameter

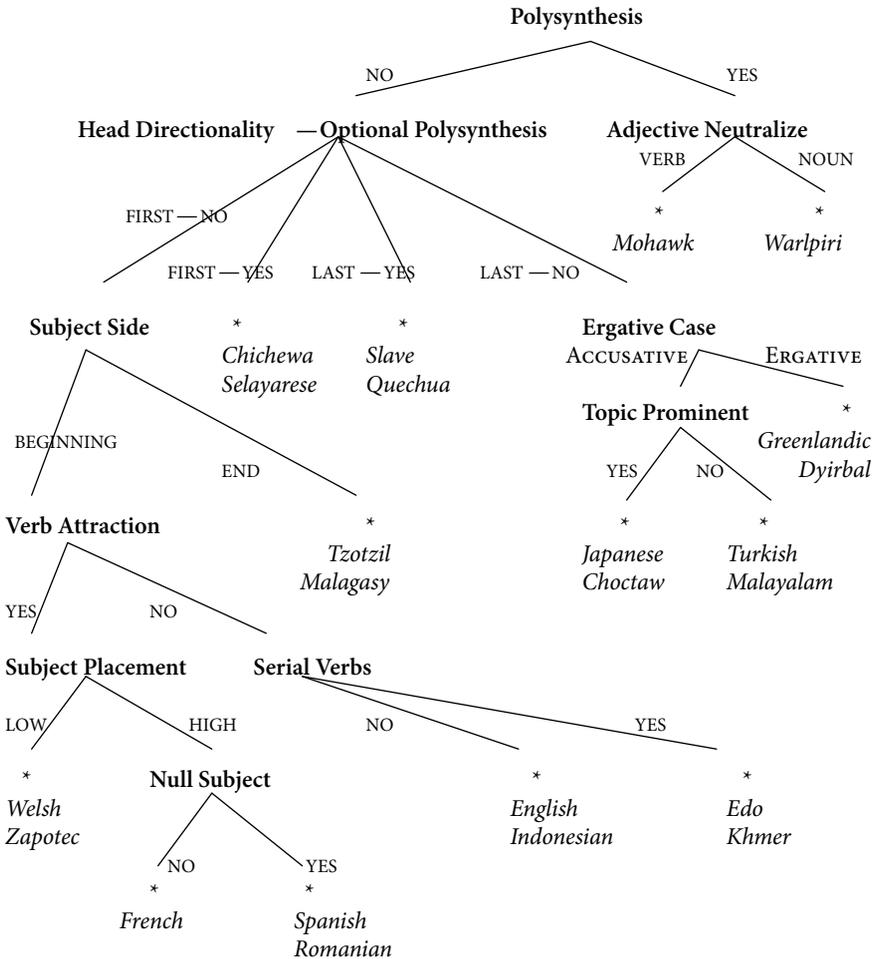


Figure 1. The Parameter Hierarchy (Baker 2001: 183)

(HDP) and the Optional Polysynthesis Parameter (OPP). The logical independence of these two parameters leads to four possible ‘choices,’ each represented by a branching line: ‘head first’ for the HDP and ‘no’ optional polysynthesis for the OPP; ‘head first’ for the HDP and ‘yes’ optional polysynthesis for the OPP; ‘head last’ for the HDP and ‘yes’ optional polysynthesis for the OPP; and ‘head last’ for the HDP and ‘no’ optional polysynthesis for the OPP. If there are no further parametric choices to be made, given a particular setting of a particular parameter, then the branch ends in a terminal symbol *. Beneath the asterisk, languages are listed that have this combination of parameter settings. As a consequence, structurally similar languages should end up being close on the diagram, and dissimilar languages far apart.

In Baker’s account, the clustering of typological features is a consequence of the formulation of the parameters themselves and the hierarchical relations among them. To take a simple case, VO languages tend to be prepositional because the notion ‘head’ enters into the definition of the Head Directionality Parameter and verbs and prepositions are heads of their respective phrases. More subtly, all polysynthetic languages are predicted to be nonconfigurational (in fact, Baker rejects the idea of a separate ‘Configurationality Parameter’ as in Hale 1983), since a positive value for the Polysynthesis Parameter treats full arguments as mere adjuncts, with corresponding freedom of occurrence. And head-initial languages are claimed never to be either ergative or topic prominent, since the branch leading to the Ergative Case Parameter originates from the setting ‘last’ for the Head Directionality Parameter and the Topic Prominent Parameter branches from the Ergative Case Parameter. In a similar manner, the nonexistence is predicted of a polysynthetic language that has serial verbs, a subject-final language without a distinct category of ‘adjective,’ and a subject-final language with optional polysynthesis.

Baker’s account of why certain typological features are more common than others is more indirect. Essentially, the more ‘choices’ a language learner needs to make, the rarer the language type is claimed to be. As far as VO versus OV is concerned:

Since the difference between English-style and Japanese-style word order is attributable to a single parameter [Head Directionality], there is only one decision to make by coin flip: heads, heads are initial; tails, heads are final. So we expect roughly equal numbers of English-type and Japanese-type languages. (Baker 2001: 134)

Why are VSO languages so much rarer than SVO languages, then? Because two more parameters enter into the characterization of the former than of the latter:

Within the head-initial languages, however, it requires two further decisions [the value for the Subject Placement Parameter and the value for the Verb Attraction Parameter] to get a verb-initial, Welsh-type language: Subjects must be added early and tense auxiliaries must host verbs. If either of these decisions is made in the opposite way, then subject-verb-object order will still emerge. If the decisions were made by coin flips, we would predict that about 25 percent of the head-initial languages would be of the Welsh type and 75 percent of the English type. This too is approximately correct ... (Baker 2001: 134)

In other words, one-way implications fall out naturally from the PH. If a language has serial verbs, then it must have the setting 'no' for the Verb Attraction Parameter. But a 'no' setting for this parameter does not entail the presence of serial verbs. Rather it predicts a 50–50 chance of serial verbs.

The great potential appeal of the PH is its architectural simplicity. All choices branch from a single parametric node, namely the Polysynthesis Parameter, which is the most basic parametric choice that language learners need to make. Furthermore, with one exception (the PH puts the Head Directionality and Optional Polysynthesis parameters at the same level), all branching (and hence all choices) are binary. One binary choice leads inexorably to another, with parameters on collateral branches playing, in principle, no role in any particular choice.

Unfortunately, the typological evidence argues against a model of parametric choice with properties remotely that simple. Take the Ergative Case Parameter as an example. In the PH, this parameter comes into play only for head-final languages without optional polysynthesis. But what Baker really wants to say, I think, is that only languages with these typological properties can be *ergative*. Somehow, speakers of head-initial languages have to know (or come to know) that their language is accusative. Nothing on the PH conveys the information that accusativity is, in essence, the 'default'. Along the same lines, nothing in the PH conveys information about whether ergative languages can have serial verbs, whether languages that neutralize adjectives can be verb-subject, or whether topic prominent languages can have null subjects. Recording this information in the PH would considerably complicate its architecture, since doing so would require that branches cross each other or some equivalent notational device.

There are serious problems as well with the idea that the rarity of a language type is positively correlated with the number of 'decisions' (i. e. parametric choices) that a language learner has to make. Baker's discussion of verb-initial languages implies that for each parameter there should be a roughly equal

number of languages with positive and negative settings. That cannot possibly be right. There are many more nonpolysynthetic languages than polysynthetic ones, despite the fact that the choice of one or the other is a matter of a yes-no choice. The same point could be made for subject-initial head-first languages vis-à-vis subject-last ones and nonoptional polysynthesis languages vis-à-vis optional polysynthetic ones. Most problematically of all, the Null Subject Parameter is the lowest of all in the PH, implying that null subject languages should be rare, indeed, rarer than verb-initial languages. However, according to Gilligan (1987), a solid majority of the world's languages are null subject.⁵

The PH is also rife with purely empirical problems. To 'start at the top', Baker assigns a positive value of the Polysynthesis Parameter to both Mohawk and Warlpiri, despite extreme typological differences between them. Among other things, Mohawk makes heavy use of incorporation and has no overt case marking, while Warlpiri has rich case marking. The problem with distinguishing the two languages by means of a case marking parameter is that Baker wants case marking to fall out from the Head Directionality Parameter, since most head-final languages have case marking. But as Table 2 shows, a sizeable percentage of head-first languages have case marking, while 36% of head-last languages lack it. None of these languages would appear to have any place in the PH.⁶

Table 2. Percent of languages of each type with explicit dependent (case) marking (Siewierska and Bakker 1996)

V-initial	V-medial	V-final
42	30	64

Furthermore, the PH posits that a positive value for the Polysynthesis Parameter automatically suggests adjective neutralization, that is, adjectives belonging to the class of nouns or verbs, rather than forming a category of their own. But it is far from being the case that adjective neutralization is limited to polysynthetic languages. According to Dixon (1977), in Chinese, Thai, and in many Austronesian languages, adjectives belong to the class of verbs, while in Arabic, Tagalog, and in many Dravidian and Bantu languages, adjectives belong to the class of nouns. Again, there is no place for such languages in the PH.

Moving further down the hierarchy, one finds more curious features. The Ergative Case Parameter applies only to head-final languages and indeed the great majority of languages with ergative case are head-final. The problem is that agreement morphology can also be ergative (as Baker himself notes, p. 181). However, such languages tend to be either verb-initial (as in Chamorro and Sahaptin) or verb-final (as in Abkhaz and Canela-Kraho) (see Nichols

1992). Since the parameters that determine verb-finality and verb-initiality could not be farther apart on the PH, it is far from clear how the two subcases of ergative agreement marking could be treated in a unified fashion and how both could be unified parameter-wise with ergative case marking (should that be desirable). To cite a different example, the Serial Verb Parameter is placed to allow only SVO languages to manifest this phenomenon, even though Schiller (1990) gives examples of SOV and VOS languages with serial verbs. In fact, Crowley (2002: xi) has offered the opinion that ‘OV [is the] order of most serializing languages.’ And only a subset of SVO languages are permitted to have a positive value for the Null Subject Parameter, even though null subject languages can be SOV (Turkish) and VSO (Irish).

The PH represents an attempt to revive the ‘holistic typologies’ that dominated the field in the 19th and early 20th centuries. The idea was to divide the languages of the world into a small number of basic ‘types’ (in the 19th century ‘agglutinating’, ‘isolating’, etc.), from which other typological properties were predicted to follow automatically. Most typologists have abandoned holistic approaches, simply because they do not work. Most seriously, typological properties tend to crossclassify with each other, rather than being organized hierarchically. It is clearly the case that not all problems with the PH are ‘structural’ in the sense that the very architecture of the hierarchy prevents an adequate statement of the relevant generalization. Indeed, Baker is to be commended for pushing the UG-parametric approach as far as he has — and certainly much farther than anyone else has done. One would naturally expect empirical problems in a pioneering work of this sort. But the bulk of problems are crucially structural. No hierarchy of the general form of the PH is capable of representing the parametric choices that the child is hypothesized to make. Needless to say, from that fact we have no right to conclude that *no* parametric approach of any sort can succeed. But the burden, I feel, is on those who have an alternative that equals Baker’s PH in detail and coverage, while at the same time providing an empirically more adequate model.

2.5 Clustering

From one parameter setting is it possible to derive diverse, seemingly unrelated, properties that hold across many unrelated languages? This subsection is devoted to arguing that there is little reason to think so. I begin by pointing out that clustering within individual languages can be handled by a rule-based model as elegantly as by a parametric model (§2.5.1). Moreover, the clustering

predicted to follow from the Null Subject Parameter simply does not exist (§2.5.2) and there is little reason to believe that certain differences between English and French that have been based on a single difference in parameter setting generalize beyond those languages (§2.5.3). Indeed, there is not that much robust clustering in general (§2.5.4). I go on to discuss some consequences of the Lexical Parameterization Hypothesis (§2.5.5). Section 2.5.6 is a summary.

2.5.1 *'Clustering' in a single language*

To begin, many studies that point to parameters and clustering give evidence from only a *single* language, or a few very closely related languages or dialects. For example, Roberts (1993b) argues that a single parametric change in the history of French triggered loss of simple inversion in questioning, the loss of V2 order, and the change of French from null subjecthood to obligatory subject status. But he provides no evidence that the properties attributed to this parameter are linked crosslinguistically. Likewise, the centerpiece of Lightfoot (1991) is the claim that a wide-ranging clustering of properties resulted from six new parameter settings in the history of English. But again, the prediction that particular parameter settings have particular clustering effects is not tested with respect to other languages.

In fact, rule-based models are also able to handle clustering effects within a single language. Consider the following phrase-structure and transformational rules proposed in Chomsky (1957):⁷

- (24) a. AUX → TNS (M) (*have + en*) (*be + ing*)
b. affixal element + verbal element → verbal element + affixal element

From these simple rules, a host of seemingly unrelated properties follows: the fact that on the surface the perfect and progressive in English are overlapping and discontinuous, the impossibility of double modals, the site for the insertion of supportive *do*, some of the peculiarities of contraction, and more. Nothing can be concluded from (24) about the clustering of those properties in any other language, but, of course, no claims have ever been made to that effect.

2.5.2 *The Null Subject Parameter*

The very few claims of parametric clustering across a wide variety of languages have not been borne out. Consider the best studied parameter in the Principles-and-Parameters approach — the Null Subject Parameter. Four of the typological features that have been claimed to be associated with this parameter

are the possibility of null thematic subjects in tensed clauses, null nonthematic (expletive) subjects, subject inversion, and *that*-trace violations. One early study of the phenomenon, Rizzi (1982), predicts the following possible clustering of features:⁸

(25)	NULL TS	NULL NTS	SI	THAT-T
	yes	yes	yes	yes
	no	yes	yes	yes
	no	no	no	no

But still other language types exist, or at least appear to. In particular, we find languages such as Brazilian Portuguese (Chao 1981) and Chinese (Huang 1982, 1984) that have null subjects, but not subject inversion. Taking such language types into account, Safir (1985) broke the Null Subject Parameter into three parts, dissociating null nonthematic subjects, null thematic subjects, and subject inversion, thereby predicting a wider set of languages than did Rizzi, namely the following:

(26)	NULL TS	NULL NTS	SI	THAT-T
	yes	yes	yes	yes
	yes	yes	no	no
	no	yes	yes	yes
	no	no	yes	yes
	no	no	no	no

Rizzi's and Safir's predictions were put to the test by Gilligan (1987), who worked with a 100 language sample, which he attempted to correct for areal and genetic bias.⁹ Gilligan devotes many pages of discussion of the problems involved in determining whether a language manifests one of the four properties or not. His final determination was often based on the results of then-current generative analyses, rather than on mere surface facts about the language in question. For example, he excluded Chinese, Thai, Indonesian, Burmese and other languages that lack agreement morphology from the ranks of those permitting null thematic subjects on the basis of the analysis of Chinese in Huang (1984), which takes the empty subject in that language to be a null topic, rather than a pro. Gilligan found the following correlations of properties in his sample (languages for which there was not sufficient data are excluded):

(27)	yes–yes	yes–no	no–yes	no–no
NULL TS – NULL NTS	24	0	15	2
NULL TS – SI	22	49	11	15
NULL TS – THAT-T	5	3	2	1
NULL NTS – SI	14	25	1	1
NULL NTS – THAT-T	7	2	0	1
SI – THAT-T	4	0	3	4

According to Gilligan, the data in (27) reveal that the only robust correlations among the four features are the following:

- (28) a. NULL TS → NULL NTS
 b. SI → NULL NTS
 c. SI → THAT-T
 d. THAT-T → NULL NTS

These results are not very heartening for either Rizzi's theory nor for Safir's, nor, indeed, for any which sees in null subject phenomena a rich clustering of properties. In three of the four correlations, null nonthematic subjects are entailed, but that is obviously a simple consequence of the virtual nonexistence of languages that manifest *overt* nonthematic subjects. Even worse, five language types are attested whose existence neither theory predicts.

Jaeggli and Safir (1989), which was published after Gilligan's study, revived the idea of the correlation between the possibility of null subjects and the existence of rich agreement. Spanish provides the best example of the correlation between the two:

- (29) habl-o 1s (Spanish)
 habl-as 2s
 habl-a 3s
 habl-amos 1pl
 habl-áis 2pl
 habl-an 3pl

German has no null thematic subjects, though it has null expletive subjects:

- (30) Es ist möglich, dass — getanzt wurde
 it is possible that danced was

German is inflected for person, number, and tense and so on that basis one might predict that it should be (fully) null subject:

(31)	(ich) arbeit-e	1s
	(du) arbeit-est	2s
	(er/sie) arbeit-et	3s
	(wir) arbeit-en	1pl
	(ihr) arbeit-et	2pl
	(sie) arbeit-en	3pl

Jaeggli and Safir noted that some forms in the German inflectional paradigm are identical and suggested tentatively that for that reason German is not fully null subject. But they go on to point to Irish (McCloskey and Hale 1984), which has synthetic forms inflected for tense, person, and number (32a) and analytic forms inflected only for tense (32b). Null subjects are licensed in the synthetic paradigm, even though some forms are identical:

(32)	a. 'would put'	b. 'put'
	1s chuirf-inn	cuir-im
	2s chuirf-ea	cuir-eann
	3s chuirf-eadh	cuir-eann
	1p chuirf-imis	cuir-eann
	2pl chuirf-eadh	cuir-eann
	3pl chuirf-eadh	cuir-eann

On the other hand, Japanese and Chinese have no number-person inflection, yet are null subject. Jaeggli and Safir (1989) conclude with the suggestion that null subjects are permitted in all and only languages with morphologically uniform inflectional paradigms, that is, in all languages in which all forms in the paradigm are complex, or none are (see also Jaeggli and Hyams 1987). So we have:

- (33) a. No complex forms: Chinese, Japanese
 b. All complex forms: Spanish, Italian, Irish, German
 c. Some complex forms: English (*talk*, but *talk-s*); French (*parl*, but *parl-ō*)

Jaeggli and Safir go to provide an account of the fact that German (and Icelandic) are null subject with expletives, not with thematic subjects.

Jaeggli and Safir's proposal has been seemingly disconfirmed by the facts obtaining in Old French (Roberts 1993b), Swedish and Russian (Speas 1994), Brazilian Portuguese (Rohrbacher 1994), and the Häme dialect of Colloquial Finnish (Vainikka 1989). Roberts (1993b); Speas (1994; 2000); and Vainikka and Levy (1999) propose more empirically adequate approaches than Jaeggli

and Safir's, though at the same time ones involving many more abstract mechanisms (and therefore difficult to evaluate without at the same time making a host of controversial subsidiary assumptions). For example, Speas (1994) bases her analysis on whether Spec, AgrP is required to satisfy any grammatical conditions: in Italian-type languages, the AGR head contains phonetic material, so the specifier is not called upon to fulfill conditions, while Chinese-type languages have no AGR at all. Hence these two language types are null subject. English-type languages cannot omit subjects, since no agreement morpheme is base-generated in AGR. Somewhat similarly, Vainikka and Levy posit that UG allows for a particular agreement-related feature bundle to occupy either the AGR position or the subject position. If the features occupy subject position, the language is null subject; otherwise not.

It is very difficult for me to imagine how the child could ever come to set this particular parameter, given any of the proposals discussed in the above paragraphs. Presumably it would have to have its inflectional morphology fully in place before it could decide whether it was safe to omit subjects or not. That order of development seems to me to be quite implausible. Significantly for our interests, Jaeggli and Safir and the subsequent work mentioned above all but omit discussion of the rich constellation of syntactic properties, whose unification under the aegis of a single parameter only a few years earlier had been heralded as the great success of the parametric approach to typology.

2.5.3 *Some differences between English and French*

To take another example of a failed prediction of clustering within the envelope of parametric theory, Kayne (1984) links parametrically the following four properties of French, all of which differ from their English counterparts:

- (34) a. The assigning of oblique case by prepositions (as opposed to objective case) (*avec lui/*le*)
- b. The impossibility of P-stranding (**Qui as-tu parlé à?*)
- c. The impossibility of Exceptional Case Marking (**Je crois Jean être sage*)
- d. The impossibility of Dative Shift (**J'ai donné Marie un livre*)

But Kayne's parameter appears to make incorrect predictions crosslinguistically. For example, according to Derek Bickerton (personal communication) few English-based creoles have stranding (35a–b illustrate with data from Sranan):

- (35) a. Nanga san u koti a brede?
with what you cut the bread

- b. *San u koti a brede nanga?
 what you cut the bread with

Yet, in such creoles there is no evidence for distinguishing objective from oblique case. Also, such an account does not distinguish elegantly between Icelandic, a case-rich stranding language, from German and some Slavic languages, also case-rich, but nonstranding. Chinese and Indonesian have Dative Shift, but no stranding, while Prince Edward Island French has stranding but no Exceptional Case Marking. Also, there is experimental work by Karin Stromswold (1988, 1989) that shows that acquisitional data do not bear out the idea that one parameter is implicated in these processes.

Now, one must be clear that the facts discussed in this and in the previous subsection do not in and of themselves 'logically' refute either Rizzi or Safir on the one hand or Kayne on the other. It is possible that the cases of nonpredicted subject inversion, for example, are the result of something other than the Null Subject Parameter. As Gilligan himself points out: 'Perhaps the Rizzi hypothesis is correct but its effects are obscured in [Brazilian Portuguese and Mandarin — two languages with a cluster of properties predicted not to exist] because of some as yet unanalyzed aspect of these languages' (p. 90). Or, perhaps we have a sampling problem, which, if corrected, would bear out Rizzi or Safir. Or perhaps, as suggested in Pica (2001), the Null Subject Parameter itself is an epiphenomenon, whose effects are to be attributed to the interaction of other parameters. Nevertheless, the fact that even the most extensively investigated generative parameters appear to lack typological support makes one wonder what the status would be of the myriad of others, were they put to a similar test.

2.5.4 *On the robustness of clustering in general*

An interesting question is how robust the crosslinguistic clustering of any set of particular morphosyntactic features is *in general*. The most thoroughly investigated set of typological features are the Greenbergian correlations, relating basic order of subject, verb, and object to other morphosyntactic properties. Table 3, gleaned from Dryer (1991), presents the correlates of verb-finality, SVO order, and verb-initiality for thirteen properties.

What we observe is that SVO languages are in general intermediate in their properties between those in which the verb is on the right margin or on the left margin. But the extent to which they are intermediate differs significantly for each property. A simple 'Head Parameter' would therefore make crashingly incorrect predictions, among other things attributing to a majority of V-final

Table 3. Percentage of V-final, SVO, and V-initial languages manifesting particular properties (Dryer 1991)

Property	V-final	SVO	V-initial
Postpositional	96	14	09
Relative-Noun	43	01	00
Standard of comparison-Adjective	82	02	00
Predicate-Copula	85	26	39
Subordinate clause-Subordinator	70	06	06
Noun-Plural word	100	24	13
Adpositional phrase-Verb	90	01	00
Manner Adverb-Verb	91	25	17
Verb-Tense/aspect aux verb	94	21	13
Verb-Negative auxiliary	88	13	00
Genitive-Noun	89	59	28
Sentence-Question particle	73	30	13
Wh-in situ	71	42	16

languages the incorrect ordering of relative and noun and failing to explain why V-initial languages are more likely than SVO languages to manifest predicate-copula order. Needless to say, one cannot rule out a priori the possibility that there exists a complex set of abstract parameters from which the percentages in Table 3 can be derived. However, I would say that skepticism is in order here and that the radically different percentages suggest that the different properties in the table are best handled by distinct rules.

As pointed out in Pintzuk, Tsoulas and Warner (2000), the MP notion that parameters are expressed as features on functional entities within the lexicon entails that 'the grammatical coverage of a single parameter is therefore constrained and may be less wide-ranging in nature' (p. 7). They suggest that it is 'not yet clear' whether the reduction in the scope of the notion 'parametric difference' will impede the effort to apply parameter theory to capturing clusters of surface differences. It seems to me, on the other hand, that it is completely clear that this aspect of the MP makes it all but impossible to predict any significant degree of clustering. The evident *lack* of clustering, then, might be taken as an argument in favor of the MP (over GB), though at the same time it leads inevitably to scaling back the arguments for a parameter-based approach over a rule-based approach.

2.5.5 *Some consequences of the Lexical Parameterization Hypothesis*

Not only have there been few examples of demonstrable crosslinguistic clustering resulting from a particular abstract parameter setting, but it has become increasingly clear that one must abandon the idea that the notion ‘setting for a particular parameter’ holds for individual languages. The original vision of parameters was an extremely attractive one, in that the set of their settings was conceived of as a checklist for a language as a whole. But the Lexical Parameterization Hypothesis (LPH; see Borer 1984; Manzini and Wexler 1987), the idea that values of a parameter are associated not with particular grammars, but with particular lexical items, has put an end to this vision. The great promise of parametric theory was its seeming ability to provide a generative approach to language typology, that is, to be able to characterize the difference from one language to the next by means of differences in parameter settings. The LPH (no matter how well motivated it might be) dashes all hope that this promise might be fulfilled. As Safir has observed:

If lexical entries within a language L can have different parameter values for different lexical items in L, then parametric theory does not compare L and another language L', but rather the lexical items of L' are compared with lexical items of L. To put it another way, it is not so meaningful to talk about a language typology so much as a crosslinguistic lexical typology. (Safir 1987: 80)

Puzzlingly from my point of view, the relocation of the site of parametric variation from grammars of entire languages to functional heads is often portrayed as a major step forward. For example, Pica (2001: vi) writes that this move ‘allows a radical simplification of the nature and design of UG yielding a model where the notion of “level” is largely reduced to S-structure whose semantics and phonological properties are built derivationally’. But the price paid for this ‘radical simplification’ is both an explosion in the number of functional categories needed to be posited within UG and, more seriously, the transfer of the burden for accounting for language-particular differences from properties of UG per se to idiosyncratic properties of lexical entries in particular languages.¹⁰ In earlier versions of P&P syntax (and in current versions such as that of Baker 2001), a given language L was posited to have a particular setting for the Head Directionality Parameter, the Serial Verb Parameter, and so on. But now, in principle, it is individual lexical items in L that need to be specified as to how they relate to head directionality, serial verbs, and so on.¹¹ I certainly agree that ‘twenty years of intensive descriptive and theoretical research has shown [that] metaparameters do not exist’ (Pica 2001: v–vi). But we must regard that conclusion as a cause for disappointment, not rejoicing.

The idea of parameter settings applying to grammars of entire languages is further challenged by the fact that a putative language-wide parameter might have to be set differently for different structures in the same language. Consider, for example, the distinction between strong and weak islands (Szabolcsi and den Dikken 1999). English is practically as permissive as Italian with respect to extraction when the extracted *wh*-element is a direct object, especially if the lowest clause is nonfinite (36a–d) and Grimshaw (1986) has called attention to a dialect of English where the facts seem to be identical to those which led Rizzi to propose *S'* as a bounding node for Italian (37a–d):

- (36) a. This is the car that I don't know how to fix.
 b. ?This is the car that I don't know when they fixed.
 c. *Tell me how you wonder which car they fixed.
 d. *Tell me how you wonder which car to fix.
- (37) a. Which book did the students forget who wrote?
 b. Which book did the TAs tell the students that they shouldn't forget who wrote?
 c. *Which book did the student forget who told them that Dorothy Sayers wrote?
 d. *Which book did the students forget who asked/told them who wrote?

It is by no means clear how to handle examples like (36) and (37) in any approach, though they seem to provide fairly conclusive evidence against a meta-parameter like Subjacency. Indeed, that is now Rizzi's opinion as well (see also Rizzi 2004):

The *S/S'* case is of some historical significance, as it was the first concrete instance of language variation treated in terms of the parametric approach. ... But seen in retrospect, this case looks quite atypical. First, no other major property seems to be clearly related to the extractability from *wh*-islands ... so one does not find the clustering of properties ... Second, judgments vary considerably among speakers ... Third, unlike the major familiar cases of parameters, this one does not seem to be reducible to a property of a head. (Rizzi 1989: 355)

Again, the conclusion that we have to abandon metaparameters like the one for Subjacency is regrettable, rather than laudable.¹² Be that as it may, however, their abandonment effectively negates the argument against the positing of language-particular rules. I see no empirical difference between the statement 'In language L some head H has the parametric property of licensing functional

category F (with feature content F)' and the statement 'Language L has a rule R specifying the existence of functional category F (with feature content F)'.

2.5.6 Summary

In summary, the phenomena discussed above suggest that the original vision of (meta)parameters was overly ambitious. A parametric theory would receive enormous support if unexpected correlations of properties followed from the positing of an abstract parameter. In such a situation, parameters would differ fundamentally from language-particular rules. Unfortunately, however, two decades of intensive research has failed to reveal the existence of the hoped for correlations.¹³ Given the LPH, on the other hand, parameter settings seem to differ little in their degree of generality from language-particular rules. Arguments against a model incorporating (2b) — one in which language-particular rules are central — lose their force when one focuses on microvariation in the properties of functional heads.

2.6 Innateness/Universality

It is important to stress that a rejection of parametric theory does not in and of itself entail a rejection of innate UG principles per se. The poverty of the stimulus based-arguments for the innateness of one or another constraint still need to be evaluated on their own merits. Let's take the extraction of *wh*-elements, for example and, for purposes of argument, hypothesize that typological differences among languages are characterized in terms of clearly-defined metaparameters. Now how might UG play a role in explaining how languages may differ in their bounding possibilities? As we have seen, GB assumed that UG provides the set of possible bounding nodes (i. e., S and NP; S' and NP; S, S', and NP; and so on), each set corresponding to a parametric choice that the child makes based on positive evidence. Later work has pointed to the specifier system of CP and IP as the crucial factor in determining how bounding works crosslinguistically. But the assumption that the set of possible bounding nodes or the set of the possible number of specifiers of CP and IP is provided in advance by an innate UG seems entirely gratuitous. If, as we have hypothesized, children know innately that general conditions on extraction exist and, as surely must be the case, positive evidence is available to pin down the specific conditions for any particular language, what reason have we for the additional assumption that UG pre-selects the set of possible bounding nodes or the number of possible specifier positions? I have never seen a poverty of the

stimulus-based argument for such an assumption and doubt that one can be constructed.

In other words, replacing parameter-settings by rules has no effect one way or the other on the innateness of grammatical principles.

2.7 Learnability

Most work in the generative tradition takes as a starting assumption the idea that first language acquisition is shaped by UG, an idea that I feel is amply supported by empirical evidence. However, most work goes on to make the additional assumption that the principal mechanism by which UG-shaped acquisition takes place is parameter-setting. That is, it takes parameters as a given and raises questions such as: 'Do parameters have default values?', 'Can parameters be reset in the course of acquisition?', and so on. Yet a number of factors suggest that a parameter-setting strategy for first language acquisition is far from the simple task that it is portrayed in much of the literature.

Several of the problems for parameters result from the fact that what the child hears are sentences (or, more correctly, utterances), rather than structures. But any given utterance is likely to massively underdetermine the particular structural property that the child needs to set some particular parameter. The greater the number of parameters to be set, the greater the problem, particularly given that few of the parameter settings appear to have unambiguous triggers. How might a child come to know, for example, in which of the 32 IP projections some particular adverb might find its proper place? Citing Clark (1994), Janet Fodor points out that there is an 'exponential explosion from the parameters to the number of learning steps to set them ... If so, the learner might just as well check out each grammar, one by one, against the input; nothing has been gained by the parameterization. ... [to] set one parameter could cost the learner thousands or millions of input sentences' (Fodor 2001b: 736). What makes the problem even more serious is the fact that children are obviously not born with the ability to recognize triggers for any one particular language. English-speaking, Chinese-speaking, and Japanese-speaking children all need to arrive at a negative setting for the Ergativity Parameter, given its existence, but it is by no means obvious what feature common to the three languages would lead the very young child to arrive at that particular setting.

In other words, the fundamental problem is that parameter-setting presupposes some non-negligible degree of prior structural assignment. To illustrate the problem, Hyams (1986) speculates that a child sets the Null Subject

Parameter with a negative value when it hears an expletive. But how can the child know what an 'expletive' is without already having a syntax in place? 'Expletive' is not an a priori construct available to the newborn, but is interpreted only with respect to an already existing grammar. But if the grammar is already in place, then why do we need parameters at all?¹⁴

Given the hypothesis that parameters are complemented by rules in the marked periphery, the learner's task is not simplified by the positing of parameters. As pointed out by Foley and Van Valin (1984: 20), it is made more complicated. Since learners have to acquire rules anyway, they have a double burden: acquiring both rules and parameter settings and figuring out which phenomena are handled by which. And one would assume (along with Culicover 1999: 16) that any learning mechanism sophisticated enough to acquire the 'hard stuff' in the periphery would have no trouble acquiring the 'easy stuff' at the core, thereby rendering the notion 'parameter' superfluous.

Along the same lines, there is no evidence that 'peripheral' knowledge is stored and/or used any differently from that provided by the system of principles and parameters per se. When head-directionality or V2-ness are at stake, do German speakers perform more slowly in reaction time experiments than do speakers of head-consistent non-V2 languages? Do they make more mistakes in everyday speech, say by substituting unmarked constructions for marked ones? Do the marked forms pose comprehension difficulties? In fact, is there any evidence whatsoever that such knowledge is dissociable in some way from more 'core' knowledge? As far as I am aware, the answers to all of these questions are 'no'. As Janet Fodor has stressed: 'The idea that there are two sharply different syntax learning mechanisms at work receives no clear support that I know of from theoretical, psychological, or neurological studies of language' (Fodor 2001a: 371).

Finally, there is little credence to the idea that there is a robust correlation between the order of acquisition of some feature and its typological status, a fact which casts into doubt the idea that parameters are organized in an implicational hierarchy. Some late-acquired features are indeed typologically relatively rare, as appears to be the case for the verb-raising that derives VSO order from SVO order (see Guilfoyle 1990 for Irish; Radford 1994 for Welsh; Ouhalla 1991b for Arabic). But other grammatical features appear to be acquired relatively late, without being typologically rare (see Eisenbeiss 1994 on scrambling in German).

In general, however, children acquire the relevant structures of their language quite early, regardless of how common that structure is crosslinguistically.

Hence English-speaking children acquire P-stranding before pied-piping (Karin Stromswold, personal communication). French-speaking children have verb-raising from the earliest multi-word utterances (Déprez and Pierce 1993; Pierce 1992; Meisel and Müller 1992; Verrips and Weissenborn 1992). English-speaking children never manifest verb-raising (Stromswold 1990; Harris and Wexler 1996). There is no period during which the grammars of German-speaking children lack V2 (Meisel 1990; Clahsen and Penke 1992; Verrips and Weissenborn 1992; Meisel and Müller 1992; Poeppel and Wexler 1993). Furthermore, children figure out very early whether their language is null subject or not (Valian 1991) and children acquiring English, German, and French evidence strong knowledge of locality in *wh*-extraction domains at early ages (Roeper and De Villiers 1994).¹⁵

From all of the above we conclude that a parameter-setting model provides no more insight into the process of language acquisition than a rule-based model.

2.8 Parametric change

Turning to other claims that have been made about parameters and change, Lightfoot's six diagnostics distinguishing parametric change from rule-based change (see (8h) above) have not been found reliably to partition the class of attested diachronic developments (Harris and Campbell 1995: 37–45). But equally seriously from our perspective, his diagnostics are not easily reconciled with claims that have been made by other generative-oriented historical linguists about language change. In particular, grammaticalization-related changes have been analyzed as falling squarely within the orbit of parametric theory (see especially Roberts 1993a, Roberts and Roussou 1999). Furthermore, some of the particular changes that Lightfoot attributes to parametric change seem incompatible with current views on the nature of parameters. For example, he cites as an instance of parametric change the reanalysis of a preposition with an adjacent verb, thereby creating a complex verb that can properly govern the trace of movement (and thereby licensing stranding of the preposition), as shown in (38a–b):

- (38) a. You talked [_{pp} to who] > You [_v talked to] who > Who_i did you [_v talk to] e_i?
- b. e was spoken [_{pp} to Mary] > e was [_v spoken to] Mary > Mary_i was [_v spoken to] e_i

It is not at all obvious how such a change could be located in some functional projection, as is required by most current approaches to parameters.

The MP conception that parameters are lexically-based makes it harder to distinguish the two types of changes (and their typological implications) that have been at the center of Lightfoot's historical work. After all, grammaticalization-based changes, morphological changes, and so on are examples of lexical change *par excellence*. It is not clear that the MP provides enough in the way of theoretical machinery to characterize any clear-cut distinction (if indeed, such a distinction exists) between parametric and nonparametric change. Perhaps for this reason 'much recent work in historical syntax has moved away from a focus on abrupt parametric change toward an emphasis on synchronic syntactic variation and the implications of gradual syntactic change as they can be interpreted within the Principles-and-Parameters paradigm' (Pintzuk, Tsoulas and Warner 2000: 10).

2.9 Summary

Despite the importance attributed to them in the past two decades of work in generative grammar, there is little reason to believe that parameterized principles play a role in the theory of UG.

3. Hawkins's processing-based explanation of crosslinguistic variation

This section will be as positive in its outlook as the previous ones were negative. I have argued at length that it is not the task of formal grammar to account for the typological variation that we find across languages. This section, on a more positive note, outlines how a theory of performance — in particular, that aspect of performance devoted to on-line processing — is well designed to handle this variation. Here I rely to a considerable extent on the work of John A. Hawkins (in particular, Hawkins 1994, 2004). Since I will be presenting the easily-accessible work of another scholar, that is to say, work that I cannot claim as my original contribution to the field, I will summarize his results, but not go into great detail on the mechanisms that he proposes.

Section 3.1 sketches the hypothesis, central to Hawkins's approach, that much of grammar can be thought of as a conventionalization of performance preferences and §3.2 sketches those principles, which are derived from performance pressure to maximize efficiency and reduce complexity.

3.1 Grammar as a conventionalization of performance preferences

The central organizing principle of Hawkins (2004) (and earlier work) is the 'Performance-Grammar Correspondence Hypothesis':

- (39) Performance-Grammar Correspondence Hypothesis (PGCH)
Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by distributional patterns of selection in corpora and by ease of processing in psycholinguistic experiments. (Hawkins 2004: 3)

The PGCH is manifest in a number of ways. For example, one might find a language or languages in which speakers have different structural means for expressing the same content. It is very often the case that one of the alternatives is more frequently chosen than the other. If the PGCH is correct, then languages *without* these structural alternatives will more often than not grammaticalize the preferred alternative. Or one notes that, given a group of semantically-related constructs, speakers more often make reference to some of these constructs than to others. The PGCH predicts that more languages will grammaticalize the more frequently appealed to constructs than the less frequently appealed to ones. To be more explicit, the PGCH makes the following predictions:

- (40) Grammatical predictions of the PGCH:
- a. If a structure A is preferred over an A' of the same structural type in performance, then A will be more productively grammaticalized, in proportion to its degree of preference; if A and A' are more equally preferred, then A and A' will both be productive in grammars.
 - b. If there is a preference ranking A>B>C>D among structures of a common type in performance, then there will be a corresponding hierarchy of grammatical conventions (with cut-off points and declining frequencies of languages).
 - c. If two preferences P and P' are in (partial) opposition, then there will be variation in performance and grammars, with both P and P' being realized, each in proportion to its degree of motivation in a given language structure. (Hawkins 2004: 6)

(40a) can be illustrated by the tendency of heads consistently to precede complements or to follow complements. One might be tempted to simply declare a head parameter provided by Universal Grammar and leave it at that. Descriptively adequate grammars certainly must express this generalization in

languages in which it applies, even if a UG-provided head parameter does not enter into the picture. But there is a lot more to the story. Consider a VO language like English, where heads typically precede complements:

- (41) V-NP, P-NP, A-of-NP, N-of-NP

In each case a 'lighter' head precedes a 'heavier' complement. But the light-before-heavy tendency in the grammar involves far more than the head-complement relation. For example, the canonical order of VP constituents is relentlessly lighter-to-heavier:

- (42) $_{VP}$ [V - NP - PP - CP] (*convince my students of the fact that all grammars leak*)

Also notice that single adjectives and participles can appear in pre-head position:

- (43) a. a silly proposal
b. the ticking clock

But if these adjectives and participles themselves have complements, the complements have to appear in post-head position:

- (44) a. *a sillier than any I've ever seen proposal
b. a proposal sillier than any I've ever seen
- (45) a. *the ticking away the hours clock
b. the clock ticking away the hours

Many more examples of this phenomenon could be provided. These generalizations reflect the PGCH.¹⁶ Evidence for its performance basis is simple. Where speakers have a choice in a VO-type language, they tend to put shorter before longer constituents. So, except for cases in which there is a strong lexical relation between V and P, PP's can typically occur in any order after the verb:

- (46) a. Mary talked to John about Sue.
b. Mary talked to Sue about John.

But all other things being equal, the greater the length differential between the two PP's, the more likely speakers will be to put the shorter one first.¹⁷ Interestingly, Hawkins's approach makes precisely the opposite length and ordering predictions for head-final languages. And to be sure, there is a heavy-before-light effect in those languages, both in language use and in the grammar itself.

So there is no question in my mind that grammars have been shaped by processing considerations — that is, by language in use.

(40a) can also be illustrated with an example provided by Givón (1979). English allows referential indefinites both in subject position and in the existential construction, as in (47a–b):

- (47) a. A man in the yard is asking for you.
b. There's a man in the yard (who's) asking for you.

However, speakers are vastly more likely to say something like (47b) than (47a). And as suggested by the PGCH, in the majority of languages referential indefinite subjects are banned outright.

To illustrate (40b), text counts show that speakers more often make reference to single entities than to multiple entities and more often to multiple entities than to precisely two entities. In keeping with the PGCH, there is an implicational hierarchy: if a language has a distinctive dual form, then it will have a distinctive plural form. And it is virtually never the case that a language requires more morphosyntactic complexity to express singularity than plurality (Greenberg 1966).

Now let us turn to (40c). Speakers have two measurable preferences: P (48a) and P' (48b):

- (48) a. P: Verbs should be adjacent to their complements.
b. P': Verbs should be adjacent to their adjuncts.

These preferences exist because there are combinatorial and/or dependency relations between both verbs and their complements and between verbs and their adjuncts and there is a measurable performance preference for structures where such relations are stated between adjacent elements. Assuming that complements and adjuncts are on the same side of the verb (a typical state of affairs whose explanation we can pass over for the sake of the exposition), P and P' are in opposition for any case in which a verb occurs with at least one complement and one adjunct. However, in the greater percentage of cases, it is the complement that is closer than the adjunct. Why should this be? Because there are *more* combinatorial and/or dependency relations that link complements to their heads than link adjuncts to their heads. Complicating matters, another motivation for grammatical structure is the reduction of the amount of phrase structure that needs to be processed on line. In a VO language, such reduction translates as a preference for short constituents preceding long constituents. Hence the prediction that languages that allow adjuncts to intervene

between verbs and their object complements will be more likely to do so for short adjuncts than for long adjuncts. As we can see below, the prediction is fulfilled for French:

- (49) a. J'admire souvent la gloire de mon père.
 'I admire often the glory of my father'
 b. *J'admire la gloire de mon père souvent
 'I admire the glory of my father often'
- (50) a. *J'admire quand je regarde à la télé des films sur la Seconde Guerre mondiale la gloire de mon père.
 'I admire when I watch on the TV films about the Second World War the glory of my father'
 b. J'admire la gloire de mon père quand je regarde à la télé des films sur la Seconde Guerre mondiale.
 'I admire the glory of my father when I watch on the TV films about the Second World War'

3.2 Efficiency and complexity in grammar

The central question, of course, is *why* speakers should prefer one structure over another. If we look at individual speakers in individual speech situations, then there can be a myriad of reasons. However, if we confine ourselves to the big picture, one factor greatly overrides all others in importance: Speakers attempt to increase efficiency by reducing structural complexity. Efficiency can be increased in three ways: first, by minimizing the domains (i.e. the sequences of linguistic forms and their conventionally associated properties) within which certain properties are assigned; second, by minimizing the linguistic forms themselves (phonemes, morphemes, etc.) that are to be processed, and by reducing their conventionally associated properties, maximizing in the process the role of contextual information (broadly construed), including frequency effects and various inferences; and third, by selecting and arranging linguistic forms so as to provide the earliest possible access to as much of the ultimate syntactic and semantic representation as possible.

Concretely, Hawkins proposes three efficiency principles:

- (51) Minimize Domains (MiD)
 The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or

dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain. (Hawkins 2004: 32)

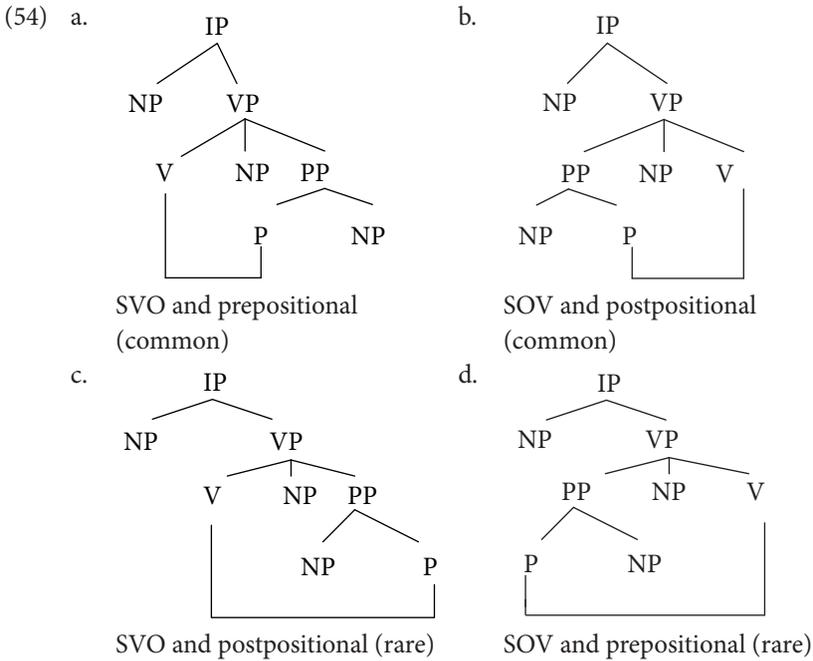
(52) Minimize Forms (MiF)

The human processor prefers to minimize the formal complexity of each linguistic form *F* (its phoneme, morpheme, word or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property *P* can be assigned in processing to a given *F*. (Hawkins 2004: 38)

(53) Maximize On-line Processing (MaOP)

The human processor prefers to maximize the set of properties that are assignable to each item *X* as *X* is processed, thereby increasing $O(n\text{-line}) P(\text{roperty})$ to $U(\text{ltimate}) P(\text{roperty})$ ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are misassigned or unassigned to *X* in a structure/sequence *S*, compared with the number in an alternative. (Hawkins 2004: 51)

Let us now have a glimpse at how some interesting typological prediction follows from each of these principles, beginning with MiD. This principle encompasses that of Early Immediate Constituents, which formed the centerpiece of Hawkins (1994) and was given an overview presentation in Newmeyer (1998: Ch. 3, §4.2.2). The basic insight of MiD is that the processor prefers shorter processing domains than longer ones, given combinatorial and/or dependency relations between two elements within a particular domain. The more such relations, the greater the pressure for adjacency. For example, MiD explains why SVO languages tend to be prepositional and SOV languages postpositional. There are four logical possibilities, illustrated in (54a–d): SVO and prepositional (54a); SOV and postpositional (54b); SVO and postpositional (54c); and SOV and prepositional (54d):



Notice that the domain necessary to identify the constituents of the VP in (54a) and (54b) — the common orderings — is the distance from P to V, with only the object NP intervening. But in (54c) and (54d) — the uncommon orderings — the object of the preposition intervenes as well. In other words, (54c) and (54d) are rarer because they are harder to process.

Let us consider another robust typological generalization, what Hawkins (1983) has called the ‘Prepositional Noun Modifier Hierarchy’:

- (55) Prepositional Noun Modifier Hierarchy (PrNMH; Hawkins 1983)
If a language is prepositional, then if RelN then GenN, if GenN then AdjN, and if AdjN then DemN.

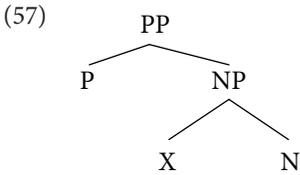
In short, if a language allows structurally complex categories to intervene between a preposition and its object, then it allows categories of less structural complexity. This hierarchy predicts the possibility of prepositional phrases with the structures depicted in (56) (along with an exemplifying language):

- (56) a. $pp[P_{NP}[__N\dots]]$ (Arabic, Thai)
b. $pp[P_{NP}[__N\dots]; pp[P_{NP}[\text{Dem } N\dots]]$ (Masai, Spanish)
c. $pp[P_{NP}[__N\dots]; pp[P_{NP}[\text{Dem } N\dots]]; pp[P_{NP}[\text{Adj } N\dots]]$ (Greek, Maya)

- d. $PP[P_{NP}[___N\dots]; PP[P_{NP}[\text{Dem } N\dots]; PP[P_{NP}[\text{Adj } N\dots];$
 $PP[P_{NP}[\text{PossP } N\dots]]$ (Maung)
- e. $PP[P_{NP}[___N\dots]; PP[P_{NP}[\text{Dem } N\dots]; PP[P_{NP}[\text{Adj } N\dots];$
 $PP[P_{NP}[\text{PossP } N\dots]; PP[P_{NP}[\text{Rel } N\dots]]$ (Amharic)

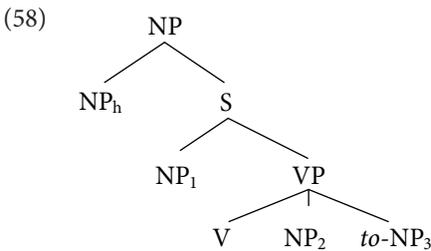
However, no language allows, say, a relative clause to intercede between a preposition and its noun complement, but not an adjective.

The PrNMH follows straightforwardly from MiD. The longer the distance between the P and the N in a structure like (57), the longer it takes to recognize all the constituents of the PP:



Given the idea that grammars are organized so as to reduce constituent recognition time, the hierarchy follows. Since relative clauses tend to be longer than possessive phrases, which tend to be longer than adjectives, which tend to be longer than demonstratives, which are always longer than ‘silence’, the hierarchy is predicted on parsing grounds.¹⁸ It is not by any means clear how the generalization captured by the PrNMH could be handled by means of parameters.

For another example supporting MiD, consider the Accessibility Hierarchy for relativization (Keenan and Comrie 1977). If a language can relativize a direct object (NP₂ in the tree below), it can relativize the subject (NP₁). If it can relativize the indirect object (NP₃) it can relativize both the subject and the direct object. And so on for more ‘oblique’ grammatical relations:



Note that the domain encompassing the head noun (NP_h) and NP₁ is shorter/less structurally complex than the NP_h-NP₂ domain, which in turn is shorter/

less structurally complex than the NP_h - NP_2 domain. In other words, the more processing involved, the rarer the structure.¹⁹

MiF embodies the insight that there is an inverse relationship between familiarity and/or frequency and complexity. Hence, as mentioned above, it is almost unknown for a language to have a form expressing plurality that is shorter or less complex than the singular form. Another example is illustrated by what Givón (1991) and earlier work (Givón 1983b, 1985) calls the ‘quantity principle’.²⁰ He argues that speakers will choose longer or more prominently stressed structures to encode ‘information that is either semantically larger, less predictable, or more important’ Givón (1991: 87). Hence, zero anaphora will be chosen when a referent is fully predictable, unstressed lexical pronouns when it is somewhat less so, followed by stressed lexical pronouns, definite NPs, and modified definite NPs. As the contributions to Givón (1983a) indicate, this generalization appears to hold crosslinguistically.

MaOP predicts that a structure or sequence will be dispreferred in proportion to the number of properties that are unassignable to it on line. Put simply, as far as processing is concerned, the sooner the better. This principle accounts for a wide variety of well-known typological generalizations involving left-right asymmetries, among which are the following:

- (59) Asymmetries predicted by MaOP:
- a. Fillers tend to precede gaps
 - i. *Wh*-questions
 - ii. Relative clauses
 - iii. Control structures
 - iv. A wide variety of ‘deletion’ constructions
 - b. Antecedents tend to precede anaphors
 - c. Topics tend to precede predications (cf. Japanese *wa*)
 - d. Restrictive relative clauses tend to precede appositives
 - e. Agents tend to precede patients
 - f. quantifiers/operators tend to precede elements within their scope

Why should *wh*-phrases tend crosslinguistically to precede, rather than follow, their extraction sites? MaOP provides an answer. Fodor (1983) observed that given an obvious filler (say, a *wh*-phrase in noncanonical position), the hearer is primed to search for a coreferential gap, but a gap is simply the absence of something — its existence could easily go unobserved by the hearer. MaOP provides a concretization of this insight. Note that more *properties* are immediately assignable to a lexical *wh*-phrase than to a gap. The other asymmetries outlined in (59) are explained by MaOP in parallel fashion.

Consider now the correlation expressed in (9) in parametric terms, namely that verb-final languages tend not to have *wh*-fronting. MaOP provides a straightforward explanation. Heads, in general, are the best identifiers of their subcategorized arguments. If one hears the verb *give*, for example, one is primed to expect two associated internal arguments, one representing a recipient and the other an object undergoing transfer. On the other hand, a human NP might or might not be a recipient and an inanimate NP might or not be an object undergoing transfer. Hence, if arguments precede their heads, as they do in SOV languages, extra cues are useful to identify their thematic status. Such can be accomplished by keeping them contiguous to the head (that is, by restricting their movement possibilities) and / or by endowing them with case marking that uniquely identifies their thematic role or helps to narrow down the possibilities.

But not every relation between grammatical elements is an asymmetrical one. For example, there are about equal numbers of VO and OV languages. Such symmetries occur when the two elements depend on each other for property assignments. So the direct object depends on the verb for case, thematic role, and for the construction of the VP mother node. And the verb depends on the object for selection of the intended syntactic and semantic cooccurrence frame (e.g. transitive vs. intransitive *run* as in *John ran* vs. *John ran the race*), and for the intended semantics of V from among ambiguous or polysemous alternatives (*ran the race/the water/the advertisement*).

Again, it is not my intention here (nor would it be appropriate) to present and motivate the dozens, if not hundreds, of typological generalizations that are explained by the processing principles in Hawkins (1994) and Hawkins (2004). Suffice it to say that they provide a convincing alternative to the idea that such generalizations should be accounted for internally to grammatical theory. If such is correct, there is no need for UG-provided parameters.

4. Conclusion

This paper has argued that it is not the job of generative theory to account for typological generalizations. Attempts to do so by means of parameterized principles have been failures. Such generalizations belong to the domain of performance, rather than to the domain of formal grammar and, as a consequence, Universal Grammar itself can be relieved of the responsibility of accounting for them. While I am certain that some formal linguists will take this conclusion

negatively, in my view it is an entirely positive development. The principles proposed by Hawkins have an obvious basis in pressure for efficiency in language processing and it is always a good thing to be able to derive the unknown from the known. Stefan Frisch has expressed my sentiments on this issue better than I could in my own words:²¹

For the traditional formalist, it is actually desirable for some linguistic patterns, especially those that are gradient, to be explained by functional principles. The remainder, once language processing influences are factored out, might be a simpler, cleaner, and more accurate picture of the nature of the innate language faculty and its role in delimiting the set of possible human languages. (Frisch 1999: 600)

Despite the importance attributed to them in the past two decades of work in generative grammar, there is little reason to believe that parameterized principles play a role in the theory of UG. The bulk of their work should be taken over by the interplay of unparameterized UG principles, language particular rules, and processing principles sensitive to structural relations holding among grammatical elements.

Acknowledgement

I would like to thank Pierre Pica and Johan Rooryck for their helpful comments on this paper, without thereby meaning to imply that they are in agreement with its conclusions. Much of what is written here will appear in expanded form in Newmeyer (2005).

Notes

1. Interestingly, Chomsky has remarked with respect to the Principles-and-Parameters approach that '[t]here remains a derivative sense in which a language *L* is a 'rule system' of a kind: namely, the rules of *L* are the principles of UG as parameterized for *L*.' (Chomsky 1991: 417)
2. To be accurate, prepositions were not incorporated into the feature system until Jackendoff (1977), though Jackendoff's system of cross-classification differed from Chomsky's. Aside from the fact that adding prepositions completes the symmetry of the system, the strongest argument for their inclusion that I am aware of is found in Lefebvre and Muysken (1988). They argue that if [-V, -N] is taken to represent the unspecification of features (or their underspecification), one is on the way to explaining why N, A, and V are often grammaticalized as P over time.

3. *The Atoms of Language*, however, presents an unusual challenge to the critical reader. It is not a 'research monograph' in the usual sense of the term, but rather, as the dust cover puts it, a 'book for a general audience'. Very little knowledge is presupposed about the intricacies of grammatical theory. Baker's book, then, is possibly unique in the annals of science publishing, in that it is a popularization of research results that were never argued for in the scholarly literature in their full technically elaborated form. Unfortunately I see no alternative but to regard and evaluate the book as if it presents research results, even though the claims that it makes are typically presented in an extremely informal manner, given the intended audience.

4. Baker also points to three other parameters, one governing *wh*-fronting and two governing domains for anaphors, that he cannot place in the hierarchy.

5. Baker (2001: 109) suggests that '[w]hen the study of null subject phenomena is taken out of its original, Romance-centric context, we may discover that the Null Subject Parameter is essentially the same thing as the Optional Polysynthesis Parameter'. Unfortunately, his rationale for this speculation is not presented in sufficient detail to allow adequate evaluation.

6. Baker (2001: 177) remarks that the difference between Mohawk-type languages and Warlpiri-type languages is 'induced by the Adjective Neutralization Parameter', but I do not understand how.

7. I have changed some of the category labels for clarity of exposition.

8. In this and in the following examples, the following abbreviations are used: NULL TS = Null thematic subjects; NULL NTS = Null nonthematic subjects; SI = subject inversion; THAT-T = the possibility of *that*-trace filter violations.

9. The most extensive published discussion of Gilligan's work that I am aware of is found in Croft (2003: 80–84).

10. It should be noted in passing that in no framework ever proposed by Chomsky has the lexicon been as important as it is in the MP. Yet in no framework proposed by Chomsky have the properties of the lexicon been as poorly investigated.

11. One is reminded of early versions of transformational grammar, in which each verb was marked for those transformations that it 'governed'.

12. Rizzi (2004), while acknowledging the problems inherent in a metaparametric approach to bounding, has remarked that the distinction between meta- and microparameters bears no theoretical import and has offered the opinion that there is 'no sound basis' for rejecting the idea that 'a single parameter may be responsible for a complex cluster or properties' (p. 332).

13. Analogously, the hypothesized subset relationship among binding domains (Manzini and Wexler 1987) referred to above in §2.2 is apparently spurious (Bondre 1993). There is no hope for interpreting binding domains in terms of a complex parameter, whereby it would follow that the conditions for long-distance domains are fulfilled automatically if the conditions for smaller domains are fulfilled.

14. For similar arguments, see Nishigauchi and Roeper (1987); Haider (1993), Valian (1990); and Mazuka (1996).
15. On the other hand, a number of language acquisition researchers continue to provide evidence for the idea, first articulated in Borer and Wexler (1987), that principles of grammar 'mature' with age (see, for example, Babyonyshev, Ganger, Pesetsky and Wexler 2001). For an interesting, albeit brief, overview of the issues involved, see Smith and Cormack (2002).
16. English allows 'heavy' sequences of adjectives in prenominal position: *the big, black, hairy, ferocious dog*. In fact, Hawkins's approach predicts that the short-before-long effect will be much less pronounced if the elements that contribute to pre-head heaviness are all of the same grammatical category. Seeming pre-head phrases such as *easy to please* in *an easy to please client* are actually complex lexical items (Nanni 1978; Roeper and Siegel 1978).
17. The discourse status of the elements involved also plays a role in ordering (see Arnold, Wasow, Losongco and Ginstrom 2000).
18. Importantly, there is no possibility of abandoning structural categories such as PP, N, etc. in the statement of the PrNMH in favor of merely quantitative notions such as 'length' or 'weight'. Long adjectives position in the hierarchy the same as short adjectives, while short possessive phrases position themselves in the same way as long possessive phrases.
19. For a largely compatible processing-based account of the Accessibility Hierarchy, see Kirby (1997).
20. The quantity principle seems to me to be an instantiation of Grice's maxim of quantity: 'Make your contribution as informative as required (for the current purposes of the exchange)' Grice (1975:54). Hawkins notes that this maxim is subsumed under MiF. For other approaches to the grammatical correlates of predictability or familiarity of information content, see Prince (1981, 1985); and Gundel, Hedberg and Zacharski (1990).
21. For a similar sentiment expressed over 25 years ago, see Kuno (1978).

References

- Arnold, Jennifer E., Thomas Wasow, Anthony Losongco and Ryan Ginstrom. 2000. "Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering". *Language*, 76.28–55.
- Babyonyshev, Maria, Jennifer Ganger, David M. Pesetsky and Ken Wexler. 2001. "The maturation of grammatical principles: Evidence from Russian unaccusatives". *Linguistic Inquiry*, 32.1–43.
- Baker, Mark C. 2001. *The atoms of language: The mind's hidden rules of grammar*. New York: Basic Books.
- Baker, Mark C. 2003. *Lexical categories: Verbs, nouns, and adjectives*. Cambridge Studies in Linguistics, vol. 102. Cambridge: Cambridge University Press.

- Beghelli, Filippo and Timothy A. Stowell. 1997. "Distributivity and negation: The syntax of *each* and *every*". *Ways of scope taking*, ed. by Szabolcsi, Anna, 71–107. Dordrecht: Kluwer.
- Bobaljik, Jonathan D. 1995. *Morphosyntax: The syntax of verbal inflection*: MIT Ph. D. dissertation.
- Bobaljik, Jonathan D. and Höskuldur Thráinsson. 1998. "Two heads aren't always better than one". *Syntax*, 1.37–71.
- Bondre, Priya. 1993. "Parameter setting and the binding theory: No subset problem". *The parametrization of universal grammar*, ed. by Fanselow, Gisbert, 17–35. Amsterdam: John Benjamins.
- Borer, Hagit. 1984. *Parametric syntax: Case studies in Semitic and Romance languages*. Studies in Generative Grammar, vol. 13. Dordrecht: Foris.
- Borer, Hagit and Kenneth Wexler. 1987. "The maturation of syntax". *Parameter setting*, ed. by Roeper, Thomas and Edwin Williams, 123–172. Dordrecht: Reidel.
- Bouchard, Denis. 2003. "The origins of language variation". *Language Variation Yearbook*, 3.1–41.
- Chao, Wynn. 1981. "PRO-drop languages and nonobligatory control". *University of Massachusetts Occasional Papers*, 6.46–74.
- Chomsky, Noam. 1957. *Syntactic structures*. Janua Linguarum Series Minor, vol. 4. The Hague: Mouton.
- Chomsky, Noam. 1970. "Remarks on nominalization". *Readings in English transformational grammar*, ed. by Jacobs, Roderick and Peter Rosenbaum, 184–221. Waltham, MA: Ginn.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Studies in Generative Grammar, vol. 9. Dordrecht: Foris.
- Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Chomsky, Noam. 1991. "Some notes on economy of derivation and representation". *Principles and parameters in comparative grammar*, ed. by Freidin, Robert, 417–454. Cambridge, MA: MIT Press. [Reprinted in *The minimalist program* by Noam Chomsky, pp. 129–166. Cambridge, MA: MIT Press, 1995].
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1998. "Noam Chomsky's minimalist program and the philosophy of mind. An interview [with] Camilo J. Cela-Conde and Gisèle Marty". *Syntax*, 1.19–36.
- Cinque, Guglielmo. 1994. "On the evidence for partial N movement in the Romance DP". *Paths towards universal grammar*, ed. by Cinque, Guglielmo, Jan Koster, Jean-Yves Pollock, Luigi Rizzi and Raffaella Zanuttini, 85–110. Washington: Georgetown University Press.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.
- Clahsen, Harald and Martina Penke. 1992. "The acquisition of agreement morphology and its syntactic consequences: New evidence on German child language from the Simone-corpus". *The acquisition of verb placement: Functional categories and V2 phenomena in language acquisition*, ed. by Meisel, Jürgen, 181–224. Dordrecht: Kluwer.

- Clark, Robin. 1994. "Finitude, boundedness, and complexity". *Syntactic theory and first language acquisition: Cross-linguistic perspectives*, ed. by Lust, Barbara, Gabriella Hermon and Jaklin Kornfilt, 473–489. Hillsdale, NJ: Erlbaum.
- Croft, William. 2003. *Typology and universals: Second edition*. Cambridge: Cambridge University Press.
- Crowley, Terry. 2002. *Serial verbs in Oceanic: A descriptive typology*. Oxford: Oxford University Press.
- Culicover, Peter W. 1999. *Syntactic nuts: Hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press.
- Damonte, Federico. 2004. *The thematic field: The syntax of valency-enriching morphology*: University of Padua Ph. D. dissertation.
- Déchaine, Rose-Marie A. 1993. *Predicates across categories: Towards a category-neutral syntax*: University of Massachusetts Ph. D. dissertation.
- Déprez, Viviane and Amy Pierce. 1993. "Negation and functional projections in early grammar". *Linguistic Inquiry*, 24.25–67.
- Derbyshire, Desmond C. 1985. *Hixkaryana and linguistic typology*. Arlington, TX: Summer Institute of Linguistics.
- Dixon, R. M. W. 1977. "Where have all the adjectives gone?" *Studies in Language*, 1.1–80.
- Dryer, Matthew S. 1991. "SVO languages and the OV:VO typology". *Journal of Linguistics*, 27.443–482.
- Eisenbeiss, Sonia. 1994. "Kasus und Wortstellungsvariation im deutschen Mittelfeld". *Was determiniert Wortstellungsvariation?*, ed. by Haftka, B., 277–298. Opladen: Westdeutscher Verlag.
- Fodor, Janet D. 1983. "Phrase structure parsing and the island constraints". *Linguistics and Philosophy*, 6.163–223.
- Fodor, Janet D. 2001a. "Parameters and the periphery: Reflections on syntactic nuts". *Journal of Linguistics*, 37.367–392.
- Fodor, Janet D. 2001b. "Setting syntactic parameters". *The handbook of contemporary syntactic theory*, ed. by Baltin, Mark and Chris Collins, 730–767. Oxford: Blackwell.
- Foley, William A. and Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge Studies in Linguistics, vol. 38. Cambridge: Cambridge University Press.
- Frisch, Stefan. 1999. Review of *Linguistic structure and change*, ed. by Berg, Thomas. *Journal of Linguistics*, 35.597–601.
- Fukui, Naoki. 1995. *Theory of projection in syntax*. Studies in Japanese Linguistics, vol. 4. Stanford, CA: CSLI Publications.
- Gilligan, Gary M. 1987. *A cross-linguistic approach to the pro-drop parameter*: University of Southern California Ph. D. dissertation.
- Giorgi, Alessandra and Fabio Pianesi. 1997. *Tense and aspect: From semantics to morphosyntax*. Oxford: Oxford University Press.
- Givón, Talmy. 1979. *On understanding grammar*. New York: Academic Press.
- Givón, Talmy, ed. 1983a. *Topic continuity in discourse: A quantitative cross-language study*. Typological Studies in Language, vol. 3. Amsterdam: John Benjamins.
- Givón, Talmy. 1983b. "Topic continuity in discourse: An introduction". *Topic continuity in discourse*, ed. by Givón, Talmy, 1–42. Amsterdam: John Benjamins.

- Givón, Talmy. 1985. "Iconicity, isomorphism, and non-arbitrary coding in syntax". *Iconicity in syntax*, ed. by Haiman, John, 187–220. Amsterdam: John Benjamins.
- Givón, Talmy. 1991. "Isomorphism in the grammatical code: Cognitive and biological considerations". *Studies in Language*, 15.85–114.
- Greenberg, Joseph H. 1966. *Language universals, with special reference to feature hierarchies*. The Hague: Mouton.
- Grice, H. P. 1975. "Logic and conversation". *Syntax and semantics, vol. 3: Speech acts*, ed. by Cole, Peter and Jerry Morgan, 41–58. New York: Academic Press.
- Grimshaw, Jane. 1986. "Subjacency and the S/S' parameter". *Linguistic Inquiry*, 17.364–369.
- Guilfoyle, Eithne. 1990. *Functional categories and phrase structure parameters*: McGill University Ph. D dissertation.
- Gundel, Jeanette K., Nancy Hedberg and Ron Zacharski. 1990. Givenness, implicature, and the form of referring expressions in discourse. Berkeley Linguistics Society 16.442–453.
- Haider, Hubert. 1993. "Principled variability: Parameterization without parameter fixing". *The parametrization of universal grammar*, ed. by Fanselow, Gisbert, 1–16. Amsterdam: John Benjamins.
- Hale, Kenneth. 1983. "Warlpiri and the grammar of nonconfigurational languages". *Natural Language and Linguistic Theory*, 1.5–47.
- Harris, Alice C. and Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge Studies in Linguistics, vol. 74. Cambridge: Cambridge University Press.
- Harris, Tony and Ken Wexler. 1996. "The optional-infinitive stage in Child English: Evidence from negation". *Generative perspectives on language acquisition: Empirical findings, theoretical considerations, and crosslinguistic comparisons*, ed. by Clahsen, Harald, 1–42. Amsterdam: John Benjamins.
- Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.
- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge Studies in Linguistics, vol. 73. Cambridge: Cambridge University Press.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Huang, C.-T. James. 1982. *Logical relations in Chinese and the theory of grammar*: MIT unpublished Ph. D. dissertation.
- Huang, C.-T. James. 1984. "On the distribution and reference of empty pronouns". *Linguistic Inquiry*, 15.531–574.
- Hyams, Nina M. 1986. *Language acquisition and the theory of parameters*. Dordrecht: Reidel.
- Jackendoff, Ray. 1977. *X-bar syntax: A study of phrase structure*. Cambridge, MA: MIT Press.
- Jaeggli, Osvaldo and Nina Hyams. 1987. Morphological uniformity and the setting of the Null Subject Parameter. *NELS* 18.238–253.
- Jaeggli, Osvaldo and Kenneth J. Safir. 1989. "The null subject parameter and parametric theory". *The null subject parameter*, ed. by Jaeggli, Osvaldo and Kenneth J. Safir, 1–44. Dordrecht: Reidel.
- Jakobson, Roman. 1936/1971. "Beitrag zur allgemeinen Kasuslehre". *Selected Writings*, ed. by Jakobson, R., 23–71. The Hague: Mouton.

- Jones, Charles. 1988. *Grammatical gender in English: 950 to 1250*. London: Croom Helm.
- Kayne, Richard S. 1984. *Connectedness and binary branching*. Studies in Generative Grammar, vol. 16. Dordrecht: Foris.
- Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kayne, Richard S. 2000. *Parameters and universals*. Oxford: Oxford University Press.
- Keenan, Edward L. and Bernard Comrie. 1977. "Noun phrase accessibility and universal grammar". *Linguistic Inquiry*, 8,63–99.
- Kirby, Simon. 1997. "Competing motivations and emergence: Explaining implicational hierarchies". *Linguistic Typology*, 1,5–31.
- Kuno, Susumu. 1978. "Generative discourse analysis in America". *Current trends in textlinguistics*, ed. by Dressler, Wolfgang, 275–294. New York: Walter de Gruyter.
- Lebeaux, David. 1987. "Comments on Hyams". *Parameter setting*, ed. by Roeper, Thomas and Edwin Williams, 23–39. Dordrecht: Reidel.
- Lefebvre, Claire and Pieter Muysken. 1988. *Mixed Categories: Nominalizations in Quechua*. Studies in Natural Language and Linguistic Theory. Dordrecht: Kluwer.
- Lightfoot, David W. 1991. *How to set parameters: Arguments from language change*. Cambridge, MA: MIT Press.
- Lightfoot, David W. 1999. *The development of language: Acquisition, change, and evolution*. Blackwell/Maryland Lectures in Language and Cognition, vol. 1. Oxford: Blackwell.
- Longobardi, Giuseppe. 2003. "Methods in parametric linguistics and cognitive history". *Language Variation Yearbook*, 3,101–138.
- Manzini, M. Rita and Kenneth Wexler. 1987. "Parameters, binding, and learning theory". *Linguistic Inquiry*, 18,413–444.
- Mazuka, Reiko. 1996. "Can a grammatical parameter be set before the first word? Prosodic contributions to early setting of a grammatical parameter". *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, ed. by Morgan, Jerry L. and Katherine Demuth, 313–330. Mahwah, NJ: Erlbaum.
- McCloskey, James. 1997. "Subjecthood and subject positions". *A handbook of theoretical syntax*, ed. by Haegeman, Liliane, 197–236. Dordrecht: Kluwer.
- McCloskey, James and Kenneth Hale. 1984. "On the syntax of person-number inflection in Modern Irish". *Natural Language and Linguistic Theory*, 4,245–281.
- Meisel, Jürgen. 1990. "INFL-ction, subjects, and subject-verb agreement". *Two first languages: Early grammatical development in bilingual children*, ed. by Meisel, Jürgen, 237–298. Dordrecht: Foris.
- Meisel, Jürgen and N. Müller. 1992. "Finiteness and verb placement in early child grammars". *The acquisition of verb placement: Functional categories and V2 phenomena in language acquisition*, ed. by Meisel, Jürgen, 109–138. Dordrecht: Kluwer.
- Muysken, Pieter and Henk van Riemsdijk. 1986. "Projecting features and featuring projections". *Features and projections*, ed. by Muysken, Pieter and Henk van Riemsdijk, 1–30. Dordrecht: Foris.
- Nanni, Deborah L. 1978. *The 'easy' class of adjectives in English*: University of Massachusetts Ph. D. dissertation.
- Newmeyer, Frederick J. 1998. *Language form and language function*. Cambridge, MA: MIT Press.

- Newmeyer, Frederick J. 2005. *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. Oxford: Oxford University Press.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Nishigauchi, Taisuke and Thomas Roeper. 1987. "Deductive parameters and the growth of empty categories." *Parameter setting*, ed. by Roeper, Thomas and Edwin Williams, 91–121. Dordrecht: Reidel.
- Ouhalla, Jamal. 1991a. *Functional categories and parametric variation*. London: Routledge.
- Ouhalla, Jamal. 1991b. Functional categories and the head parameter. Paper presented at the 14th GLOW Colloquium
- Pica, Pierre. 2001. "Introduction." *Linguistic Variation Yearbook*, 1.v–xii.
- Pierce, Amy. 1992. *Language acquisition and syntactic theory: A comparative analysis of French and English child grammars*. Dordrecht: Kluwer.
- Pintzuk, Susan, George Tsoulas and Anthony Warner. 2000. "Syntactic change: Theory and method." *Diachronic syntax: Models and mechanisms*, ed. by Pintzuk, Susan, George Tsoulas and Anthony Warner, 1–22. Oxford: Oxford University Press.
- Poeppl, David and Kenneth Wexler. 1993. "The full competence hypothesis of clause structure in early German." *Language*, 69.1–33.
- Poletto, Cecilia. 2000. *The higher functional field: Evidence from Northern Italian dialects*. Oxford: Oxford University Press.
- Prince, Ellen F. 1981. "Toward a taxonomy of given-new information." *Radical pragmatics*, ed. by Cole, Peter, 223–256. New York: Academic Press.
- Prince, Ellen F. 1985. "Fancy syntax and 'shared knowledge'." *Journal of Pragmatics*, 9.65–82.
- Radford, Andrew. 1994. "Clausal projections in early child grammars." *Essex Research Reports in Linguistics*, 3.32–72.
- Reuland, Eric J. 1986. "A feature system for the set of categorial heads." *Features and projections*, ed. by Muysken, Pieter and Henk van Riemsdijk, 41–88. Dordrecht: Foris.
- Richards, Norvin. 2001. *Movement in language: Interactions and architectures*. Oxford: Oxford University Press.
- Rizzi, Luigi. 1982. *Issues in Italian syntax*. Studies in Generative Grammar, vol. 11. Dordrecht: Foris.
- Rizzi, Luigi. 1986. "Null objects in Italian and the theory of *pro*." *Linguistic Inquiry*, 17.501–557.
- Rizzi, Luigi. 1989. "On the format for parameters." *Behavioral and Brain Sciences*, 12.355–356.
- Rizzi, Luigi. 1997. "The fine structure of the left periphery." *Elements of grammar: Handbook of generative syntax*, ed. by Haegeman, Liliane, 281–337. Dordrecht: Kluwer.
- Rizzi, Luigi. 2004. "On the study of the language faculty: results, developments and perspectives." *Linguistic Review*, 21.323–344.
- Roberts, Ian. 1993a. "A formal account of grammaticalization in the history of Romance futures." *Folia Linguistica Historica*, 13.219–258.
- Roberts, Ian. 1993b. *Verbs and diachronic syntax: A comparative history of English and French*. Studies in Natural Language and Linguistic Theory, vol. 28. Dordrecht: Kluwer.

- Roberts, Ian and Anna Roussou. 1999. "A formal approach to 'grammaticalization'". *Linguistics*, 37.1011–1041.
- Roeper, Thomas and Jill De Villiers. 1994. "Lexical links in the *wh*-chain". *Syntactic theory and first language acquisition: Cross-linguistic perspectives*, ed. by Lust, Barbara, Gabriella Hermon and Jaklin Kornfilt, 357–390. Hillsdale, NJ: Erlbaum.
- Roeper, Thomas and Muffy Siegel. 1978. "A lexical transformation for verbal compounds". *Linguistic Inquiry*, 9.199–260.
- Rohrbacher, Bernhard. 1994. *The Germanic VO languages and the full paradigm: A theory of V to I raising*: University of Massachusetts Ph. D. dissertation.
- Safir, Kenneth J. 1985. *Syntactic chains*. Cambridge: Cambridge University Press.
- Safir, Kenneth J. 1987. "Comments on Wexler and Manzini". *Parameter setting*, ed. by Roeper, Thomas and Edwin Williams, 77–89. Dordrecht: Reidel.
- Schiller, Eric. 1990. The typology of serial verb constructions. *Chicago Linguistic Society* 26.393–406.
- Siewierska, Anna and Dik Bakker. 1996. "The distribution of subject and object agreement and word order type". *Studies in Language*, 20.115–161.
- Smith, Neil and Annabel Cormack. 2002. "Parametric poverty". *Glott International*, 6.285–287.
- Speas, Margaret. 1994. "Null arguments in a theory of economy of projection". *University of Massachusetts Occasional Papers*, 17.179–208.
- Speas, Margaret. 2000. "Constraints on null pronouns". *Optimality-theoretic syntax*, ed. by Legendre, Géraldine, Jane Grimshaw and Sten Vikner, 393–425. Cambridge, MA: MIT Press.
- Stromswold, Karin. 1988. "The acquisitional implications of Kayne's theory of prepositions". Unpublished ms., MIT.
- Stromswold, Karin. 1989. "Using naturalistic data: Methodological and theoretical issues (or how to lie with naturalistic data)". *Paper presented at the 14th Annual Boston University Child Language Conference, October 13–15, 1989*.
- Stromswold, Karin. 1990. *Learnability and the acquisition of auxiliaries*: MIT Ph D thesis.
- Stuurman, Frits. 1985. *Phrase structure theory in generative grammar*. Dordrecht: Foris.
- Szabolcsi, Anna and Marcel den Dikken. 1999. "Islands". *Glott International*, 4–6.3–8.
- Thráinsson, Höskuldur. 1996. "On the (non)-universality of functional categories". *Minimal ideas: Syntactic studies in the minimalist framework*, ed. by Abraham, Werner, Samuel David Epstein, Höskuldur Thráinsson and C. Jan-Wouter Zwart, 253–281. Amsterdam: John Benjamins.
- Travis, Lisa. 1989. "Parameters of phrase structure". *Alternative conceptions of phrase structure*, ed. by Baltin, Mark R. and Anthony S. Kroch, 263–279. Chicago: University of Chicago Press.
- Vainikka, Anne. 1989. *Deriving syntactic representations in Finnish*: University of Massachusetts Ph. D. dissertation.
- Vainikka, Anne and Yonata Levy. 1999. "Empty subjects in Finnish and Hebrew". *Natural Language and Linguistic Theory*, 17.613–671.
- Valian, Virginia V. 1990. "Logical and psychological constraints on the acquisition of syntax". *Language processing and language acquisition*, ed. by Frazier, Lyn and Jill de Villiers, 119–145. Dordrecht: Kluwer.

- Valian, Virginia V. 1991. "Syntactic subjects in the early speech of Italian and American children". *Cognition*, 40.21–81.
- Verrips, M. and Jürgen Weissenborn. 1992. The acquisition of functional categories reconsidered. Paper presented to the Workshop on Crossing Boundaries, Tübingen
- Williams, Edwin. 1981. "Language acquisition, markedness, and phrase structure". *Language acquisition and linguistic theory*, ed. by Tavakolian, Susan L., 8–34. Cambridge, MA: MIT Press.
- Zanuttini, Raffaella. 2001. "Sentential negation". *The handbook of contemporary syntactic theory*, ed. by Baltin, Mark and Chris Collins, 511–535. Oxford: Blackwell.

Author's address:

Frederick J. Newmeyer
University of Washington,
Department of Linguistics
P.O. Box 354340
Seattle, WA 98195–4340
USA
fjn@u.washington.edu