J. Linguistics **37** (2001), 367–392. © 2001 Cambridge University Press
DOI: 10.1017/S0022226701008817 Printed in the United Kingdom

REVIEW ARTICLE

Parameters and the periphery: reflections on *Syntactic nuts*[1]

JANET DEAN FODOR

Graduate Center, CUNY

(Received 3 March 2000; revised 26 August 2000)

Peter W. Culicover, *Syntactic nuts: hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press, 1999. Pp. viii + 244.

1. Deducing learners from languages

In this oddly named book, Peter Culicover's project is 'to investigate the properties that the learning mechanism for language must have by investigating the properties of language itself as bounding conditions on such a mechanism' (1). In practice this amounts to an extensively illustrated reminder of the fact that natural languages abound in irregularities and exceptions, and that these have to be learned just as much as the big broad language specific properties such as head-direction or pro-drop. This rules out parameter setting as the sole mechanism for syntax acquisition, and calls for a learning mechanism that seeks out and evaluates generalizations in the input language sample.

A short first chapter sets the scene. Culicover (henceforth, C) reviews the classic statement in *Aspects* (Chomsky 1965) of the relation between linguistic theory and language acquisition, and expresses skepticism about how linguistics has been approaching the explanatory goals that Chomsky outlined. For C, abstract analysis is not explanatory; parameters are not explanatory; the number of grammars is not finite; the major generalizations are not syntactic but relate syntactic and conceptual structure; core and periphery are not distinct; less is innate than has recently been assumed. Thus C takes his stance. There are no definitive arguments here for these assumptions, but as a statement of intent to explore a new explanatory terrain they are clear enough.

The centerpiece of the book is the detailed discussion in chapter 2 ('Categories') and chapter 3 ('Constructions') of the unruly behavior of

[1] I am grateful to Mark Baltin, Judy Bernstein, Marcel den Dikken and Fritz Newmeyer for their advice and information, and to Maggie Tallerman and Bob Borsley for very helpful feedback on an early draft.

many linguistic phenomena. These are the syntactic nuts, the hard cases for linguistic theory to account for. C has put together a flourishing collection of them over the years, and devotes much care to running them through all the standard distributional tests in order to determine what special conditions apply to them. This is exemplary descriptive linguistics, and there's lots of it, presented clearly and very compactly. It is also entertaining, fun to follow along and scribble in the margins, maybe note some even more bizarre properties of the construction that is under the microscope.

Matters take a more serious turn with the claim that these exceptional phenomena cannot be accounted for by syntactic theory, and that the general patterns they violate could not be due to syntactic principles of Universal Grammar (UG). A similar conclusion is drawn in the final chapter (chapter 4, 'Constraints'), where the linguistic examples reveal exceptions to otherwise general locality constraints on movement and are taken to show that no such constraints exist in grammars or in UG: the existence of irregularities entails that even regularities cannot be innate. A functional explanation of locality restrictions ascribes them instead to a least-effort tendency on the part of learners.

At the most general level, the outcome for linguistic theory is that its responsibilities are diminished. The consequence for language acquisition is that learners must do more work with less innate assistance. They cannot rely even on a predefined set of syntactic categories (see, for example, page 85) or on structural guidance from X-bar principles (for example, page 95). These are too simplistic to cover all the facts, and so it is implausible to suppose that they are psychologically real. If this is correct it is cause for concern. Any erosion of the presumed innate basis for language acquisition increases the challenge for acquisition theory, and things are bad enough already. Despite years of effort, it has proven extremely difficult to show how the syntax of a natural language could be acquired even if it were completely innate except for 20 or 30 binary choices to be determined by the input sample. Though setting parameters sounds easy, attempts to implement the process have run into serious snags (Clark 1994). Gibson & Wexler (1994) showed that under some not too implausible assumptions, successful learning is not guaranteed even in a miniature domain with only 8 grammars to choose from. With less innate guidance, learning can reasonably be expected to be more difficult still. Though it may seem natural to regard the disorderliness of language as evidence that UG is not very restrictive, from a learnability point of view the opposite conclusion might be more appropriate. As Howard Lasnik has observed, the wilder the facts, the more help children would seem to need in acquiring them.

A sufficiently powerful learning device might compensate for lack of innate knowledge. Some success has been reported for UG-free syntax learning in a connectionist framework (Elman 1993), but to what extent the grammar attained resembles human grammars is unknown. C claims to have a

'dynamical system' learning model (vi) that can extract just the right generalizations from the input sample, but it is outlined in only sketchy terms in this book. A full presentation is promised in a companion volume to be co-authored with Andrzej Nowak. In the meantime we are told what the learning mechanism must be able to do, but very little about how it does it – so little, that it's fair to wonder whether it really can.[2]

One property of C's learner is clearly stated: though purely formal syntactic constraints are scarce, the learner can exploit syntax-semantics correspondences. A major theme of the book is that natural language grammars are framed in terms of correspondences between syntactic structure and conceptual structure, as proposed by Jackendoff (1997). For example, on X-bar theory C writes:

> The picture that appears most natural … is that the X′ Schema, to the extent that it captures true generalizations, reflects the structure of conceptual representations. To the extent that there is some uniformity in the organization of conceptual structure, there will be a uniformity in phrase structure. But the characterization of phrase structure is in principle unbounded in its variability (96).

I think it's true that formal work on language learnability would benefit from greater attention to the contribution of meaning (inferred from the context) in establishing sentence structure for learners. This has been neglected in much of the recent work on mechanisms of syntactic parameter setting, though it was a major premise of earlier investigations of learnability such as Wexler & Culicover (1980). For some readers, C's adoption of the idea that grammars explicitly state correspondences between form and meaning may be the most interesting aspect of this work. But in this discussion I will set aside the role of conceptual structure in order to focus on other aspects of C's picture of how languages and learners relate to one another.

2. Learners need markedness

Linguistically oriented research on language learnability has had different preoccupations at different periods: the complexity of input sentences in the 1970s; the lack of negative evidence and the need for the Subset Principle in the 1980s; ambiguous input in the 1990s. C's concern with the existence of very narrow as well as very broad grammatical generalizations falls squarely under the negative data problem: without negative instances to disconfirm overly broad hypotheses, incurable overgeneration may result. There appear to be two ways in which this could be avoided: either learners are highly

[2] An example: '… the learner … has to compare all of the words of the language with one another, form hypotheses about which of them function in a similar way, and on the basis of these similarities determine what the categories are' (37).

conservative, or else UG specifies how far an observed fact can be generalized. C opts for the former, though the latter is linguistically much more interesting if we can pull it off; I will return to it below.

One reason why parameter theory (Chomsky 1981) was embraced by acquisition theorists was that it undercuts this whole question of how far learners should generalize the specific facts they are exposed to. It does so by eliminating entirely the process of formulating generalizations expressed in terms of rules or constraints. Apart from composing a lexicon, all a learner must do is select between syntactic alternatives that are fully prefigured innately. Because they are innate they can also be innately priority-ordered where this is necessary to satisfy the Subset Principle and avoid overgeneration. But C's mission in this book is to wean linguists away from complacently invoking parameter setting to escape the negative data problem.

The main line of argument relating language facts to learning mechanisms can be summarized as follows:

(i) Natural languages exhibit many idiosyncrasies as well as broad structural regularities; the 'periphery' of language is vast and strange. (Note: I use the terms CORE and PERIPHERY for convenience in what follows, though C urges that there is no sharp divide but 'a continuum along which a full spectrum of possibilities can be found, from the very idiosyncratic to the very general' (vi).)
(ii) These odd phenomena fall beyond the scope of parametric description, since it is not plausible that every possible structural peculiarity is innately prefigured in the value of some parameter.
(iii) Therefore, the periphery cannot be acquired by parameter setting. So not all of syntax can be acquired by parameter setting (plus lexical acquisition).
(iv) Native speakers are consistent and confident in their judgments of exceptional constructions, and know things about them to which their exposure must have been minimal. Evidently, peripheral learning is not inferior to the learning of regularities.
(v) Any learning mechanism capable of acquiring the imperfect generalizations of the periphery is presumably capable of acquiring exceptionless generalizations also.
(vi) The preferred hypothesis is that there is a single learning mechanism for syntax. This is not only parsimonious but respects the fact that idiosyncratic properties may be intermingled with core properties in the same construction.
(vii) Therefore, a language is acquired not by parameter setting, which plucks a few facts from the input to trigger generalizations that go far beyond them. Instead, all language knowledge must be acquired by induction over the available input sample, by a CONSERVATIVE ATTENTIVE LEARNER (CAL).

In short: the periphery cannot be assimilated to the core, but it is also not disjoint from the core; hence the core must be assimilated to the periphery, and both must be learned in an 'old-fashioned' way as set out in Chomsky (1965) rather than Chomsky (1981, 1986, 1995); learning theory must give up triggering, and turn back to hypothesis formation. To make this work, C recognizes that he will need an evaluation measure, in other words a theory of markedness, just as Chomsky did in the pre-parametric learning theory of *Aspects*. The reasoning in points (i)–(vii) is extended as follows:

(viii) Evidence that learners make conservative grammar choices, and the uniformity of acquisition outcomes, indicate that learners' choices are guided by a universal evaluation metric of some kind.

(ix) For acquisition of all of language including the periphery, the descriptive options are too numerous to be individually listed and ranked; markedness scales must therefore be projected on a general basis.

This is all in sharp contrast with recent work in the Minimalist Program, whose primary task Chomsky (1995: 8) says is 'to show that the apparent richness and diversity of linguistic phenomena is illusory and epiphenomenal'. But whether or not one finds its conclusions palatable, (i)–(ix) must be recognized as a rational and responsible line of thought. C is pretty convincing on the scale of the periphery, and on how deeply it infiltrates the core (48): natural language is not a rigorously tidy system with just a faint nimbus of irregularity at its furthest fringe. C's goal of developing an acquisition algorithm that is comfortable with both the regular and the peculiar therefore has much to commend it. 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. (I am assuming here, though perhaps this is wrong, that distinct learning mechanisms for core and periphery in syntax cannot be identified with the two learning styles proposed for inflectional morphology; see Clahsen 1999, Pinker 1999, and many references there.) So it is certainly not impossible that the prominent patterns which are commonly cast as parameters are just one end of the continuum of eccentricity that C envisages. Whether or not it ultimately succeeds, treating core and periphery as continuous thus seems well worth the try.

However, there is a baby-with-the-bathwater tendency in C's thinking which is likely to discourage linguists of various persuasions. Universal syntactic categories like Noun, Verb and Adjective don't predict all distributional properties of a word, so universal syntactic categories are banished.[3] Structural constraints like Subjacency don't apply absolutely, so

[3] C continues to use standard structural categories for descriptive purposes in the chapters that follow. He does not illustrate how grammar rules are formulated when categories are represented as regions of semantic space (42ff.).

out with structural constraints. Here, I think, C is overreacting. For anyone who has reasons for wanting to hold on to standard syntactic categories and constraints, I don't believe it will be difficult to characterize them in a way that accommodates variability. This, after all, is what a theory of syntactic markedness would do; see discussion below.

Fortunately, the various strands of C's thought can be teased apart and assessed separately. This means one can follow C's lead in trying to develop a theory of syntactic oddity even if one disagrees with him on how to go about it.

3. The empirical basis for a theory of markedness

Markedness certainly seems to be the right place to look for a way to allow for many degrees of generality without inviting total anarchy. But markedness theory for syntax has met with little interest over the years, or even strong passive resistance (though finding a natural home these days in Optimality Theory). C goes all the way back to *Aspects* for his inspiration, though he develops the theory somewhat differently than Chomsky. For C, markedness is part of the learning mechanism, not part of linguistic theory; this is the 'Antigram' hypothesis (224ff.). For instance, since island constraints on extraction are not absolute they must be understood 'not as components of grammars, but as guidelines for CAL' (222).

The grammatical or extra-grammatical status of markedness can be important to how acquisition proceeds, but it won't be an easy matter to settle. C's kind of close descriptive work could contribute some well-worked-through case studies. The book does include some. By adopting and extending the complexity measure proposed by Hawkins (1994), C predicts markedness scales for some clusters of extraction constraints (while rejecting Hawkins' claim that such preferences are often grammaticalized). Parasitic gaps are ranked (180ff.); they are claimed to be least marked in *without*-clauses (e.g. *These are the reports that I filed without reading*), next best in temporal clauses with PRO subject, then in tensed clauses, and so on. Control and other such relations are briefly discussed (200ff.). Based on directness of the correspondence between syntactic and conceptual structure, C argues that control of an embedded subject is less marked than control of an embedded object (as in *This table needs painting*), but only a little less marked than empty operator constructions (as in *I bought this book for you to read*). But surprisingly, C does not draw on his descriptive analyses of the 'syntactic nuts' in chapters 2 and 3, which make up the heart of his book, as a basis for developing a general theory of how constructions can deviate from the norm.

It comes as a bit of a shock to discover, after working one's way through them, that this is not what the syntactic nuts are there for. It seems that their role is only to support points (i)–(iv) above; that is, to demonstrate that

learners do acquire a great many highly particular facts. But surely C's examples are good for more than this? Consider a simple case. C notes (93ff.) some contrasts among degree expressions in adjective phrases, as in (1).

(1)	a	very	tall	woman	*very	tall	a	woman	very	tall	women
	*a	so	tall	woman	so	tall	a	woman	*so	tall	women
	*a	that	tall	woman	that	tall	a	woman	*that	tall	women

These are just the tip of an iceberg. The degree words that compel the AP to precede the determiner include also *how* (e.g. *How tall a woman is she?*, **A how tall woman is she?*). Like *so* and *that*, this occurs only in the singular, or perhaps only if not string-vacuous (**How tall women are they?*). In this respect, interrogative *how* differs from exclamatory *what*, which precedes both the determiner and the rest of the AP, and can do so even when the determiner is not overt (*What a tall woman she is! What tall women they are!*). For a *how*-phrase the pre-determiner position is obligatory even in an echo question (*She is **how** tall a woman? *She is a **how** tall woman?*). For some speakers there is alternation with an *of*-form (*How tall of a woman is she?*; see Borroff 2000), which may be a colloquial attempt (a language change in progress?) to make sense of the word order of *how tall a* by relating it to constructions like *how tall for a woman*, *how much of an idiot*.

What can this package of facts tell us? It doesn't seem to be completely off the wall, but exemplifies some type of movement (or its equivalent in nonmovement theories) within DP, possibly into specifier position, controlled by a few specific lexical items. This kind of movement is not as common as, for instance, movement of a *wh*-object to [Spec,CP], but it is by no means unique (see Valois 1991, Bennis et al. 1998). So this is just the sort of phenomenon we need to feed into a theory of markedness: not the norm, but not an out-and-out idiom either.[4] It's clear what sorts of questions need to be asked: What properties of this movement within DP make it more marked than fronting an object in its clause? How does the grammar of English (as opposed to the grammar of Dutch) license it? Does it employ richer phrase structure principles, to base-generate the surface order as is? or to provide a marked landing site for AP movement? Does it use a special strong feature to attract the AP upward? Does it shift the point of spell-out, so that LF-scope-related movement occurs in overt syntax, getting the operator high enough in the DP to be able to scope out of it? Even within one linguistic theory there may be several ways in which the data might be handled in principle; which of them is correct needs to be established. If that can be done then we can go on to ask: What other phenomena could be characterized using the same kinds of grammatical machinery? Do those

[4] For present purposes I will assume that infrequency across languages is a sign of greater markedness. The question of how markedness values ought to be established is too deep to be discussed here.

other phenomena also occur – if not in English then in other languages? Do they occur more often or resist language change more stoutly than constructions which deviate further (as far as we can tell) from familiar forms?

I have framed these questions in syntactic terms, where C would look to syntactic-conceptual structure correspondences. But that's not important here; the general project clearly transcends any particular theory of grammars. Ideally, each of the oddball examples analyzed in the book would contribute a fact or two to a substantial database that needs to be assembled so that we can discern the trends and the limits of exceptional behavior: when a construction deviates from the norm, in what ways can it deviate, and by how much? But curiously, and disappointingly, C doesn't raise these questions.

4. Can exceptions be explained?

In the typical case, C first argues that some properties of a given construction are *sui generis*, i.e., don't follow from general principles (on any reasonable linguistic theory). Then he summarizes these properties in a formula expressed in a fairly standard lexical/structural/featural notation.[5] An example is his (39), from chapter 3, for the *no matter* small clause construction (e.g. *You should continue with the lecture no matter how confused the students*).

(2) (= C's (39), ch.3)
no matter [Predicate $[+\mathrm{WH}]_i$ NP_i]
where reference of NP is a definite generic.

It is implied that this is the form in which the facts are mentally recorded in the grammar, and that a grammar would contain many such formulae to describe the periphery. Some may constitute lexical entries, but not all have a particular lexical head to pin the deviant behavior on, for example the phrase structure schemata for infinitival relatives in C's (251) of chapter 3.

(3) (= C's (251), ch.3)
N [[$_{\mathrm{PP}}$ P NP [+REL]] to VP]
N [to VP]
N [[$_{\mathrm{PP}}$ for NP] to VP]

These formulae are to be taken, I think, as representative of C's theory of the ways in which marked constructions can stretch the limits of the descriptive

[5] Some of these descriptive formulae use terms such as 'sentence-initial position' or 'to the left of' rather than structural specifications (e.g. on pages 53, 104, 164). C is drawn to a theory in which linear precedence relations are defined in the absence of tree structure: 'it is worth exploring whether the learner should be assumed to seek a structure-dependent account of every phenomenon it encounters' (103).

apparatus of the language faculty.[6] He doesn't strive for any deeper representation of the facts than this because he is convinced it is futile to do so.

A repeated theme, capping the discussion of many examples, is that once the properties of a construction have been identified and shown to be *sui generis*, there is nothing more to be said about it.[7] On the *no matter* construction, for instance, he comments:

> [T]he description [(39)] simply expresses the facts, while a theory explains why (39) is possible and why some equally complex construction is impossible. But notice that in our various attempts to formulate a reduction, we were only able to translate the idiosyncrasy of (39) into a different vocabulary; at some level, the fact that (39) is possible must be stipulated. The notion that a more abstract theory 'explains' the possibility of (39) is an erroneous one, in fact, unless the reduction can be accomplished in terms of strictly general and independently motivated principles. (120)

I will return to the *no matter* construction below. For present purposes, let's accept C's claim that it has 'an irreducible ad hoc component' (119). Even so, this argument against the possibility of an explanatory account of (39) does not go through because it mixes up two quite different notions of explanation. A theory might explain an apparently idiosyncratic construction

---

[6] Given INDEPENDENT evidence for C's claims that grammars directly specify syntactic/conceptual correspondences, and that grammatical descriptions are neither abstract nor deep, then the assumption that core and periphery are similar in kind (e.g. page 35) might entail that peripheral facts MUST be represented in something like the style of (2) and (3) 'as part of the syntactic side of the correspondence rule' (21).

[7] Related points are made under the head of 'the Encoding Argument'. No general definition is given, but at first mention it takes the following form. I quote it because it seems less plausible when paraphrased.

> In the case of subcategorization it is possible to assign traditional category labels and subcategory features to classes of elements; in the case of syntactic subregularities and idiosyncrasies it is possible to derive the surface forms from more abstract representations … The question is, are these accounts descriptions of the language faculty in the human mind/brain, or are they simply formal encodings [i.e., not psychologically real, JDF] of the behavior of the real language mechanism in the mind/brain? I will suggest that in general the grammatical accounts are encodings of the linguistic knowledge. The Encoding Argument is based on the simple observation that in order to determine what the grammatical category or grammatical description of a phenomenon is, the learner must first determine precisely what the superficial generalization is and what the superficial idiosyncratic facts are … Once the learner has identified the special properties and made the generalizations, the learner knows the relevant facts about the language in this domain, and we may say that the learner has 'acquired' this part of the language in some concrete sense. There is nothing more about the language that the learner acquires in virtue of assigning various elements to the linguist's categories. Hence this assignment is an encoding, not an explanation. (67–68)

AWAY, as not really idiosyncratic at all; or it might explain how it is that the construction is possible and yet (genuinely) idiosyncratic.

C's reference to 'translation' reflects debates about the proper level of abstractness of linguistic analyses. Linguistic theories differ in this respect, and C favors concrete analyses (see page 87). In this passage he is fending off the claim that greater abstractness brings greater explanatoriness IN THE SENSE THAT more properties of the construction fall under the general principles of an 'abstract' theory than under those of a 'concrete' theory. In other words, he is defying any other theory to squeeze out more normality from his exceptional examples than he has done. But even if he wins that challenge, another baby has just been tossed out with the bathwater. The other sense of explanation, which has got lost here, concerns the issue raised above: how phenomena that are admittedly idiosyncratic are licensed. How is the descriptive apparatus (whatever it is, within a given linguistic theory) to be extended to accommodate them? Do the observed cases constitute a natural class of extensions of that apparatus? Are any parts of a grammar inviolable even in extreme cases? The kind of theory that could answer these questions could indeed tell us why (39) is possible but some other construction is not.[8]

THIS is the kind of theory of exceptionality that could contribute to acquisition theory, by allowing learners to watch out for the potential special cases while simultaneously sailing ahead on generalizations that are guaranteed by UG to be exceptionless. This bears on point (iv) above: learners' knowledge of properties of peripheral constructions despite the rarity of relevant evidence (see Culicover & Jackendoff 1999 for instances of this). An approach to this puzzle is to suppose that what little input the learner encounters is analyzed by the language faculty in an innately determined way (just one of the many ways in which a generalization seeker like CAL might record the facts). How it is assimilated will determine which general principles it will interact with, and that will establish properties of the construction that are (so far) unattested in the input. This, of course, is just the 'poverty of the stimulus' argument for UG; it is a traditional point but still valid.

The descriptive apparatus of one linguistic theory might well be more adaptable than that of a competitor for purposes of explaining why learners encode peripheral constructions as they do. And if other things were equal, that would be the better theory; explanatory success in this sense could give one linguistic theory an edge over another. Perhaps in the end C might (or

[8] Since the periphery is a major locus of cross-linguistic variation, this investigation should be largely coextensive with the investigation of possible language differences, and a theory of one can supplement a theory of the other. For instance, if the idea can be sustained that cross-language variation is restricted to the lexicon and formal features of functional categories (Borer 1984, Chomsky 1995), then obviously the same should be true of all idiosyncratic phenomena within a language.

might not) win his campaign against abstractness via this path – but only by seeing first that a detailed description of each oddball construction is not where inquiry ends but where theory construction begins.

5. What can be *sui generis*?

There is no lack of hypotheses to be explored. For example, many of C's examples center on a lexical item, and these items are mostly not nouns and verbs but 'closed class' items or functional categories. Chapter 2 discusses *either*, *or*, *nor* and related words, quantifiers and determiners, and various prepositions (*during*, *since*, *notwithstanding*, etc.). Chapter 3 covers *no matter*, *-ever* (as in *Whatever the reason* (*is*), *I will not condone your behavior*), sluice-stranding (*John went to NY with someone*, *but I couldn't find out who with*) which is restricted to a few *wh*-words and prepositions, and *not*-topics (*Not in my car you won't*), as well as *wh*-movement and other phenomena not tied to particular lexical items. (The non-colloquial character of some of C's examples may suggest they are not learned spontaneously in childhood. But other peripheral constructions are acquired early; see Fodor 1994.) The peculiarities of many of these constructions look as if they might reflect extensions of normal selection relations. The preponderance of such cases may be an accident of C's choice of examples to present; but if not, it is an important (though not unexpected) contribution to a theory of idiosyncrasy.

The point can be illustrated with *no matter*. In one variant, *no matter* takes a finite embedded *wh*-question, as in *No matter how much you complain*, *you are going to eat your spinach*. The *no matter* part is an odd collocation, perhaps only historically related to the verb or noun *matter*, but it functions fairly normally as a subordinating conjunction (like *after* or *although*), except that (according to C) it selects a *wh*-complement (like *despite*). However, in the variant summarized by (39), *no matter* takes a small clause complement, which somewhat surprisingly seems to have a [Spec,CP] landing site for *wh*-movement, and which must meet some additional and unusual conditions. As (39) records, the *wh*-phrase must be the predicate, as in *no matter how reasonable his complaint* but **no matter how many of his complaints reasonable* (compare the finite form *no matter how many of his complaints are reasonable*). And the predicate must be what is *wh*-fronted: **no matter which actor how intelligent* is impossible, distinctly worse than ?**no matter how intelligent which actor*, even though the latter is a superiority violation. C states that the subject must be a definite generic NP, as in *no matter how tall the building* but **no matter how tall it/a building/John/every student*; some speakers allow an indefinite generic, but not a pronoun or proper name. And this must be all there is: no embedding (*no matter how tall Mary told Sally the building* *(*is*), *no matter how tall the building* *(*is*) *going to be*), and no adverbs between the subject and where the copula would be in a full clause

(*no matter how tall the building probably* *(*is*), though note *no matter how tall the building* (*is*) *eventually*).

Some but not all of these properties are shared by other 'free range' constructions that C discusses (with *despite*, *-ever*, *notwithstanding*, and so forth). There also seems to be some echo of these restrictions, though I can't think why, in verbless exclamations of an archaic flavor such as *How bright the stars!* but not **How bright Betelgeuse!* So, though unusual, the constraints on *no matter* may not be unique. It is not unusual that the lexical head *no matter* selects the types of complement it takes, but it is more unexpected that its small clause complement is +wh. It is not unusual that the +wh element must be fronted, presumably to [Spec,CP], as for the complement of *wonder*. But it is strange that *no matter* appears to be controlling properties of the *wh*-phrase such as its predicate status or its underlying (or trace) position, and also properties of the small clause's subject (unless the restrictions on the subject prove to be semantic).

It seems that the selectional dependencies are reaching down deeper into the structure here than is usual.[9] A traditional view is that a head selects only its sisters (complements). Baltin (1989) argues for selection of the head of the complement. By Spec-head agreement this could control some properties of the complement's specifier, such as obligatoriness of a +wh feature (unless satisfied in C by *whether*); this is essentially the *wh*-criterion imposed via selection. It would be more of a curiosity, though, for Spec-head agreement to demand a predicate as specifier. If any dependencies hold between the head of the complement and ITS complement, the domain of influence could stretch even further down (though presumably not indiscriminately, e.g. not to an adjunct in the complement). Baltin suggests, for instance, that *declare* selects *that* as head of its complement, and *that* selects a finite I as head of its complement. For C's *no matter* small clauses, this machinery might suffice. *No matter* would select an empty +wh C as head of its complement, and that would require a +wh [Spec,CP]. The ban on embedding, expressed by the subscripts in (39), could be handled if the empty C head selected by *no matter* could select a small clause with an empty predicate (perhaps with generic meaning, thereby causing the subject to be generic). If this empty predicate requirement were satisfiable only by a trace, for recoverability reasons, it would entail the predicate status of the phrase in [Spec,CP], so that this would not need to be independently stipulated. (Acceptable adverbs, as in *no matter how tall the building eventually*, would have to be adjoined at a higher level, or extraposed from the fronted predicate: *no matter how tall eventually the building*.)

[9] What I am loosely calling 'selection' here may be effected in different ways in different theories: selection of underlying form or of surface form, via features on functional heads or SYNSEM specifications in lexical entries, and so on. Though the mechanisms are of interest, I will keep this discussion neutral between them as far as possible.

If something like this approach, employing extended selection requirements, should turn out to be correct for *no matter*, it would raise the question of how, and how far, the usual scope of selection can be expanded in exceptional cases. Is there post-movement selection? Is there selection across intervening heads? One conjecture would be that there is never action at a distance – that however else they deviate, each selection relation is strictly local, as in the unmarked case, though they may create more elaborate patterns by chaining together, as suggested above. This and other potential limits remain to be established, by reference to other kinds of exceptional selectional phenomena.[10] If there are limits, and if learners know what they are, they won't need to waste effort anxiously scanning input sentences for selectional dependencies which could never occur even as exceptions. Note that the work of testing the input for possible exceptions is equally great whether there are any in the target language or not. As long as, for all the learner knows, any construction MIGHT be subject to some odd restriction, all constructions must be scrutinized for all possible restrictions (Fodor 1985, Pinker 1989). For this reason, an innate limit which excludes some otherwise imaginable phenomenon is of value even if the range of natural language phenomena remains strictly infinite.

A different approach to *no matter* would relate its special properties not to selection but to constraints on the remnants in ellipsis constructions.[11] The result would be not unlike C's (39), but motivated now in terms of

[10] An interesting possibility is that an exceptional selection relation is at the heart of the conditional comparative construction (e.g. *The more it rains*, *the angrier I get*) discussed by C and more fully by Culicover & Jackendoff (1999), though at first sight this may look to be a 'pure' syntactic anomaly. *Either-or* coordination in English, and similar phenomena in other languages, give evidence that a selectional dependency can hold between the lexical markers of coordinated constituents: *either* needs *or*, no other word of its category will do. The conditional comparative construction also has matched elements introducing its two clauses (e.g. *the more … the angrier*), though these are not base-generated lexical items but are inside phrases which have been moved so that they are in mutually accessible positions. They are at the very top of their respective clauses because of two facts noted by C as unprecedented: movement of the comparative phrase is obligatory, and it does not allow pied piping of a preposition. Conceivably, then, the primary eccentricity of the conditional comparative is a selectional dependency of a type that is characteristic of coordination but which uncharacteristically holds here between the heads of phrasal constituents (or possibly between features that have percolated to their dominating nodes). This dependency must hold at a post-movement stage, or else before movement between features that will subsequently force movement, though the latter may be more difficult to engineer here than in the case of *no matter*.

[11] Though it may not be entirely possible, I would like the discussion here to be read as neutral with respect to particular derivational mechanisms for ellipsis (e.g., LF copying versus PF deletion; surface remnants in situ or moved out first; see Lasnik 1999 for discussion), in order to keep the focus on the larger difference between lexical selection mechanisms and ellipsis-related mechanisms. Any approach which standardly characterizes the possible surface remnants in ellipsis phenomena, however it does so, will serve for present purposes. (I lack space to discuss additional empirical requirements that C observes, such as the amnestying of island constraints and facts of stress retraction.)

grammatical mechanisms. For Gapping, constraints on the prosodic contour and on how many constituents remain overt are familiar. For sluice-stranding, C's investigation (section 3.2.3) uncovers what looks like a simple surface filter permitting only certain lexical combinations of *wh*-element + preposition to remain: *who to* is good, and so is *what with*, but *who near* is not, nor is *what under*, or any phrasal *wh*-NP such as **how much money with*. Following Lobeck (1995), C gives distributional evidence that these are the remnants of a CP, which rules out a simple base-generation analysis (C's (105)). I think it is clear that they are two separate remnants, the *wh*-element having moved out of the PP as in C's (107), not (106); this is needed to account for the acceptability of examples like *I know he did that but I don't remember what for* with purposive *what for*, for which stranding is obligatory (*What did he do that for?* but **For what did he do that?*, though there is a rhetorical *for what* as in: *He did it but for what!*). If grammars standardly specify the permitted surface remnants of ellipsis constructions, then there is no problem about ensuring that everything else is null in a clause with sluice-stranding (**I know she gave the books away but I forget who Moby Dick to*). Remnant specification also offers a clue to the superficial lexical filter on sluice-stranding, so it may not be necessary to conclude that grammars can impose such filters just anywhere (see also Merchant 1999). We might speculate that wherever remnant categories are specified in the normal case, particular remnant lexical items may be specified as a special case. Then sluice-stranding could select *who to*, much as a verb (e.g. *make*) may in the marked case select a particular noun (e.g. *headway*) as its object; see Baltin 1989.

Circling back now to *no matter*: it could be that the language faculty assimilates it to ellipsis phenomena, representing the small clause in terms of its surface remnants (*wh*-predicate and subject), much as C does in (39). The impossibility of additional structure such as embedding and intervening adverbs would follow. This approach to *no matter* could relate C's examples to similar examples with only one remnant (such as *No matter how fit*, *Sam hasn't a chance of finishing the marathon*, with a *wh*-predicate but no overt subject), and to other constructions which permit a similar ellipsis of the copula with fronted predicate and generic subject, such as the conditional comparative (e.g. *The higher the stakes*, *the lower his expectations*; see Culicover & Jackendoff 1999).

It would take a great deal more work to decide which if any of these ideas is correct. But unlike C, I am claiming that the analysis matters. There may be a significant amount of 'irreducible idiosyncrasy' on ANY imaginable analysis of a construction, but it is still of interest where in the grammar that idiosyncrasy resides. A very simple illustration of this is the unexpected locally bound pronoun in English 'ethical datives', e.g. *I think I'll have me a nice hot bath*. This might be a violation of Principle B, or it might stem from analyzing the ethical dative NP as an adjunct rather than an argument

(Fodor 1994). Which is correct clearly bears on the theory of markedness; for example, on whether there is a scale of possible deviations from core binding principles, and on how much leeway there is in the conceptual/syntactic mapping for oblique complements such as recipients and beneficiaries.

To summarize the main points: any or all of the conjectures I have outlined here may be false, but they are at least conjectures of the right kind, addressing the question of which grammatical principles or mechanisms have the flexibility to permit departures from the norm. These proposals do not eliminate the need to stipulate some very particular facts, but they do seek to relate the stipulations to the general ecology of natural language grammars. This, I believe, is the route we have to take toward a true theory of the periphery. I'm sure others can make a better job of it than I can. But C's examples are a goldmine for such a theory, and somebody should be making good use of them.

6. Acquiring generalizations

How does CAL, the conservative attentive learner, go about its work? There are a few passages in this book which address this (pages 28–30, 196, 213–216). Concerning much of the mechanism I can only guess, and Culicover & Nowak (in preparation) may show that I have guessed wrong. But there are basic points worth making whether they apply precisely to C's model or not.

CAL can do two things which are known to be difficult for a learning model without systematic negative evidence. It can discover partial regularities and it can recognize obligatoriness. CAL 'generalizes where possible, and is attentive to the evidence about the generality of the correspondence [between conceptual structure and form]. The attentive learner knows that certain regularities hold for a particular subset of items, while others hold for all or virtually all members of a category. The attentive learner learns that certain conceptual structure properties must be expressed, while others need not be' (30). The mechanism for this has something in common with statistical learning algorithms: a learner like CAL 'is capable of generalizing less than globally when the evidence warrants' because it 'is not only sensitive to the structure of input but in some way takes into account the relative frequency of instances and patterns'; it can thereby acquire 'a range of generalizations, from those that cover large-scale categories such as NP and VP, to those that concern a small number of lexical items' (197). As noted above, markedness as well as statistics helps the learner to find the right level of generalization (196).

These are bold claims in view of the more or less total collapse in the past of 'little linguist' models of hypothesis formation and testing. Perhaps reliance on conceptual structure makes all the difference, or perhaps a dynamical system model can do what others cannot; we must wait and see.

But a nagging doubt remains: perhaps CAL only appears to be able to do these things because C has not grasped the full extent of the implementation problems. Confidence is not increased by the complete lack of references to any of the works in the literature that have grappled with this; Pinker, for example, is nowhere in the bibliography. I will address three points here: how CAL compares with parameter setting; the need for mental bookkeeping to keep track of the bodies of data over which generalizations are formulated; and the way in which markedness interacts with simplicity in guiding the choice of hypotheses.

First, an out-and-out mistake about the relation between parameter setting and the discovery of generalizations. This would be almost too silly to mention except that I have heard it quoted approvingly, although all it does is muddle the issues that matter. C gives an excellent characterization of parameter triggering (30): 'In syntax, the triggered learner sets parameter values on the basis of exposure to key exemplars; the full range of evidence is … unnecessary for learning'. This contrasts with CAL which 'must actually have access to all of the data and does not go significantly beyond its experience'. But later we are told that

> the [parametric] learner must identify the correct generalization as a prior condition for setting the parameters. So the learner must observe that the language in question has or lacks overt *wh*-movement, has or lacks partial *wh*-movement, and so on, as a step on the way to setting the parameters. Arriving at these generalizations is close if not equivalent to learning the language (146)

so nothing is gained by actually setting the parameters. This is just wrong.[12]

There are two respects in which parameter setting is based on incomplete evidence. If several surface facts are associated with the same parameter, observing one suffices to trigger knowledge of the others. The classic example (whether or not it remains valid on current theory) is the triggering of the negative value of the null subject parameter by an overt expletive (Hyams 1986). C rejects 'deep parameters' such as this (on pages 7, 16) and I won't discuss them further here. Another defining characteristic of parameter setting is that it is incremental. A grammar change is made in response to an individual input sentence; there is no memory for prior sentences. Hence

[12] This is a form of the Encoding Argument (note 7 above), now misapplied to parameter setting. It should be noted that there IS an apparent paradox of parameter setting but it is not this one. The learner must identify the correct STRUCTURE of sentences (i.e., of input word strings) as a basis for setting the parameters; but since sentence structure is determined by the grammar, establishing the correct structure seems to demand prior setting of the parameters (Valian 1990). This holds for any method of syntax acquisition unless it can somehow relate grammars directly to word strings. Several different solutions to this problem (such as the learning by parsing method outlined in section 7 below) have been proposed and debated in recent research.

there is no question of the learner having access to a simultaneous array of input sentences so that they can be compared, contrasted and cross-checked to see how far a grammatical generalization extends. So there is no possibility of the learner establishing the generalization prior to setting the parameter. Rather, the generalization is established BY setting the parameter. I believe this is the key to solving the extent-of-generalization problem for language acquisition, and I will argue below that it can be adopted by non-parametric theories too. So it is worth getting this point straight.

In parameter setting what sets the bounds on generalizations is UG, not induction over the input. I noted earlier that this UG solution to the extent-of-generalization problem is the alternative to strict conservatism for a learner without negative data. And it is a welcome alternative. Seriously conservative learning is not a practical option once the periphery is taken into account (Fodor 1994). It requires the learner to creep toward a target generalization, passing through along the way EVERY smaller generalization (compatible with the data) that any natural language might exhibit. There can be an enormous number of such intermediate steps, as C's own descriptive work makes clear – arguably too many for this type of learning to be compatible with the speed at which children master basic facts about their language (see Wexler 1998). Thus it seems that learners MUST make some (highly constrained) leaps ahead of their data. Parameter setting is a means of choreographing these leaps. The learner observes a sentence with a *wh*-object in initial position, not adjacent to its verb. This sentence must have other properties as well; we may imagine that the object is feminine plural and animate, the verb is stative and nonfinite, with a deontic auxiliary, the subject is a proper noun, and so forth. The parameter value for overt *wh*-movement is triggered, and it instantly generalizes over all these particularities to a grammar which permits movement of singular *wh*-objects as well as plural ones, objects of non-stative verbs as well as stative verbs, and so on. Furthermore, UG guarantees (setting aside known complications here) that this is the RIGHT generalization, or at least UG guarantees that it is not wrong but can be extended into the full generalization by subsequent input. The contribution of UG can be very particular. Because the parametric generalizations are innately encoded, they don't have to meet any uniform criterion but can differ in both character and extent from one parameter to another. For example, unlike *wh*-movement, a parameter concerned with agreement would NOT extend a co-occurrence generalization from plural to singular NPs.

Incremental learning (without memory for past inputs) is characteristic of parameter setting, which C rejects, but it is also a central feature of the powerful work on the learning of Standard Theory transformational rules to which Culicover himself contributed (Wexler & Culicover 1980). To revert to a labor-intensive paradigm-based pattern-finding mechanism (e.g. on page 37) would seem to be a move in the wrong direction. But C gives an argument

to establish the feasibility of the inductive approach. The lexicon is clearly learned in this fashion, not by triggering. And syntax learning is not significantly different from lexical learning; 'the difference appears to be one of scale, not of type' (33). Therefore 'I suggest … that the CAL that is capable of dealing with lexical correspondences of varying degrees of generality should be capable of dealing with syntactic correspondences of varying degrees of generality' (35). But there is a flaw here. Syntax learning CANNOT be assimilated to lexical learning if the learning method is the discovery of generalizations by amassing instances. That requires past inputs to be recorded in memory so they can be consulted again later. This is feasible for the lexicon. Each lexical item may be mentally recorded, in each of its forms, and with each of its argument structures, etc., as it is encountered. Later perhaps, some properties of these items may be erased from memory because they fall under patterns that the learner has now spotted. But this cannot be true for syntax. There is surely no mental place in which a child stores all the sentences he or she hears, in order to work over them later.

I turn now to markedness as a determinant of the extent of learners' generalizations. Markedness theories standardly hold (perhaps even as a point of definition; see Battistella 1990) that learners adopt the least marked hypotheses compatible with their evidence. A natural psychological assumption is that learners are least-effort devices that do no more than is necessary to accommodate the evidence. Hence it is plausible to identify least marked with simplest, as in Chomsky 1965. For C this is especially advantageous, since simplicity offers an open-ended projectible markedness scale to satisfy point (ix) in section 2 above. All that's required to make this work is some notation or format for representing the contents of grammars (rules, principles, constraints, filters, etc.) such that less marked grammars are simpler than more marked grammars. For C the basis for the simplicity ranking reflects (at least in part) the transparency of syntactic/conceptual structure relations:

> On the present markedness approach to linguistic universals, in formulating a correspondence between form and meaning the learner seeks to reduce as much as possible deviance from the conceptual structure representation. (198)

In practice, it is extremely difficult to devise a satisfactory representation system for the simplicity metric to operate over. Typically in linguistics a shorter rule (e.g. fewer features, fewer context specifications) is more general than a longer rule. So a simplicity metric implies that learners prefer more general rules to more specific rules; but this is the opposite of conservative learning and would be utterly disastrous. On the other hand, it would be uncomfortable to have to claim that conservative learners are 'most-effort'

learners that adopt the most complicated hypotheses compatible with the evidence.[13]

Beating the simplicity/generality correlation is perhaps the most difficult aspect of constructing a nonparametric learning model. C becomes mired in this problem in chapter 4. I have been stuck in the same mud myself, having tramped over much the same ground in a series of papers, some with Stephen Crain, a few years ago (e.g. Fodor & Crain 1987; Fodor 1989, 1992a, b). I have never felt such a sense of struggle as in trying to get the simplicity metric to work, to reconcile the need for learners to generalize and the need for learners to be conservative enough to detect exceptions. The pull of the opposing considerations (learners clearly do/clearly don't generalize) is so strong that there's a danger of just seesawing uselessly between them. Culicover succumbs to this at more than one point. For example, he writes:

> Linguistic phenomena vary in their complexity, and the complexity of phenomena correlates in part with how naturally they are learned. On the markedness perspective, the learner will formulate the most general (i.e. least marked) grammar consistent with his or her experience. (196)

But later on the same page he notes:

> The approach taken here is a natural descendant of Chomsky's early proposals on markedness, in which he hypothesized that the complexity of a rule of grammar correlates with the degree of specificity of its structural description. Very specific rules are possible, on this view, but are more marked than those that are relatively general. CAL formulates the **most specific hypothesis** consistent with the available evidence and its prior experience. (196; C's emphasis)

The learner favors the most general hypothesis; the learner favors the most specific hypothesis. Simplicity favors the first; conservatism demands the second. Which does C intend?

In Fodor (1992a) I showed that the simplicity/generality correlation can be broken by adopting a SPECIFIC DEFAULTS principle: every feature (or other

[13] In his treatment of extraction in chapter 4, C adopts from Hawkins (1994) a notational format for representing long-distance dependencies. He proposes that every node initially blocks extraction, and a learner discovers one by one the nodes that bridge extraction in the target language, adding these into the representation of permitted movement (or binding) dependencies. Fodor (1992a, b) presented a similar approach defined over SLASH feature paths in a GPSG framework. The purpose is to encourage conservative learning in a least-effort system by making the grammar more complex the more extensive the extraction patterns it permits. This contrasts with the more usual assumption that every node type is transparent to extraction unless specifically designated as a blocking node. However, C does not extend this general technique for managing the simplicity/generality relation to other constructions in his book which do not involve the Hawkins metric. Also, he does not comment on how to avoid the other horn of the dilemma: that postulating more costly rules for more general phenomena may result in highly complex adult grammars which fail to capture valid generalizations; see discussion below.

element of syntactic descriptions) has an innately specified default value. This ensures that when some property or feature F is not specified in a grammar rule, the rule does not license a free choice of values of F (a superset rule) but only the default value (a subset rule). In consequence, a simple rule which leaves much underspecified does not license more sentences than a fully specified rule does; it licenses less-marked sentences than the specified rule. If a learner adopts a rule on the basis of an input sentence with the default value of F, the learner can leave feature F unspecified in the rule. If the input exhibits the marked value of F, the rule must specify the marked value of F, or else the learner's grammar won't license the input (a violation of the Greediness principle; see Gibson & Wexler 1994). If a language permits both values of F, both must be explicitly represented in the grammar (since underspecification licenses only the default); thus this superset situation is the most complex to represent, as conservative learning requires.

The problem with this approach is that – unless learners radically restructure their grammars at some point – the final grammar attained is not as svelte as would normally be assumed by linguists unconcerned with learnability problems. As linguists, we wouldn't expect a grammar (an adult grammar) to have to state explicitly that both singular and plural NPs can be *wh*-fronted, as can both masculine and feminine NPs, animate and inanimate NPs, and so forth; it should say just that NPs (or XPs?) can be fronted. Though learners must be conservative, we want adults to have compact grammars that capture the broad generalizations. Rule collapsing notations can reduce the formal cost of stating post hoc generalizations, but the cost must not be eliminated entirely or else conservatism would not be fostered. For this reason, I believe the Specific Defaults principle is best combined with a very richly specified UG which tells the learning system what properties it needs to bother with in which contexts, and what it can safely ignore. A lot of learning work can be avoided, and the complexity of the final grammar greatly reduced, if UG narrows down the facts that learners need to record. For agreement relationships the learner must note number and gender, but for *wh*-movement it need not: UG vouches for their irrelevance, so Specific Defaults does not insist that they be specified. Specific Defaults was originally worked out for rule learning, but note how close this approach is to the idea above that a parameter for *wh*-movement, by not differentiating genders or numbers of moved items, is UG's way of telling the learner that it is safe to generalize across these features.

7. Parametric and non-parametric learning

This convergence suggests that a non-parametric theory might try co-opting what parameter theory does well: incremental learning, and UG-guidance concerning what can and cannot be generalized. I have found in recent work that when parameter setting is implemented, it is more like rule acquisition

than has standardly been supposed. This is because the major problem for parameter setting has been to 'decode' the parametric information carried by an input sentence. That is: how can a learner tell, on encountering a novel sentence, which combinations of parameter values COULD license it? Those are the only values worth hypothesizing. But there are 2^n possibilities for n binary parameters – a vast number. Decoding on this scale is made possible by taking parameter values to be tree fragments (or equivalently, by taking them to be rules, schemas or lexical entries that define tree fragments). Parameter setting is the process of fitting these tree fragments, or 'treelets', to an input sentence, to find out which of them are needed to parse it. Those which are essential to a successful parse should be adopted into the learner's grammar. How the tree fitting is accomplished by the parsing routines is described elsewhere (Fodor 1998, Sakas & Fodor 2001). That it works (except for weakly equivalent languages) is documented by Bertolo et al. (1997).[14]

This process of learning by parsing is unlike the familiar metaphor of tripping parametric switches, but the latter demonstrably does not work for natural languages because their significant structural properties are not superficially evident in word strings (Clark 1994).[15] But importantly, the process of learning by tree fitting is not exclusive to parameters, as switch-setting was; it can just as well be described as rule learning, or HPSG rule schema activation, or discovering the elementary trees of a TAG grammar for the target language. Hence this method could be employed by a linguistic theory without explicit parameters, to provide an incremental learning algorithm that is closely guided by UG.

To meet C's standards, however, we cannot stop there. We must see whether this system, designed for the tidy world of binary parameters, can

[14] How does taking parameter values to be treelets help? Efficient and accurate parametric decoding requires that a parameter value can be added to a grammar and immediately contribute to the parsing of sentences, without any need to recompile the grammar for that purpose. Treelets satisfy this condition, permitting the following learning procedure. The learning system attempts to parse an input string with the currently hypothesized grammar. Wherever in a sentence that parse attempt fails, the learner temporarily recruits into the current grammar all additional parameter values that are consistent with UG, and continues parsing the string. Any parameter value(s) that are needed for a successful parse are adopted into the learner's grammar; other parameter values that were temporarily added but contributed nothing are not adopted. (Like standard parameter setting, the treelet model at present deals only in structural properties, not correspondences with conceptual structure. I think it could be adapted to the latter, and it might then work even better.)

[15] Other implementations of parameter setting, such as variants of random walk learning as in the Triggering Learning Algorithm of Gibson & Wexler (1994), or a genetic algorithm as proposed by Clark (1992), consume more time, input sentences, processing resources or memory than is believed to be appropriate for modelling human language acquisition; for discussion see Sakas & Fodor (2001), Fodor (2001). Cue-based learning, as proposed by Dresher & Kaye (1990) for phonology and Lightfoot (1997) for syntax, and assumed by many linguists, fails as switch-setting does if cues are superficial, and faces resource problems if cues are deep; for discussion see Gibson & Wexler (1994) and Fodor (2001).

cope with the full range of phenomena that C has drawn attention to. I don't know whether it can, but some parts of the apparatus are already in place. Somehow UG must project an intricate web of innate information, specifying all the properties of sentences that are predictable, and what the options are for filling in the rest. Neither default properties nor UG-guaranteed properties need to be registered in the learner's grammar; the Specific Defaults principle will prevent overgeneration by rules/treelets with these features left unspecified. The degree of markedness of any particular grammatical construction will be a function of how many non-default features it takes to fit it. A sentence that can be parsed by highly underspecified treelets, with the rest supplied by UG, will be closer to the core than one which needs very particular treelets with many marked features. When the learner's parsing routines analyze an input sentence, they will give priority (as always) to the least costly analysis, using as few marked properties as possible. Therefore the learning procedure will find in the sentence as much regularity as exists there. These predictable aspects need no action by the learning mechanism; only what is truly exceptional about a construction will be stored in the grammar. Syntactic nuts are just the extreme case where almost every property is exceptional.[16]

Note that on this approach, unlike C's, all the interesting and difficult work is done by linguistics. The treelet-parsing learning mechanism is simply an efficient means of delivering this linguistic information to learners at the rockface. A learner confronted with a novel target language sentence can know what changes in the grammar it demands, without having to compare it with other sentences, or run through all potential derivations for it one by one.

Expanding the capability of this system from twenty or thirty binary parameters to the full-scale variability of a natural language will entail, I assume, expanding out each parametric treelet into a family of related treelets, with less and more marked members. The former will be less specified, and favored by learners where possible; the latter will be more complex, and resorted to only where the input cannot be accommodated more simply. The details of these clusters of related treelets are for each

[16] Possibly what C intends by a syntactic nut is not just a highly marked construction but one that is isolated from others in the language. This would come about if the unmarked counterparts of the distinctive marked features of the construction are not licensed in that context. A tenet of markedness theory is that acceptability of a marked feature may entail acceptability of its unmarked counterpart in the same contexts, though not vice versa. But this entailment may not hold for all features. Where it does, a deviant form will lie at the outer edge of a block of related forms; e.g. for the conditional comparative (see note 10 above), everything in between normal coordination and *the more … the merrier* would be acceptable in English. Clearly this is not so. The conditional comparative stands out on its own, showing that UG does not encourage learners to fill in that region of linguistic space. Which marked feature values entail their unmarked values is another aspect of UG that could be mapped out on the basis of exceptional examples like C's.

linguistic theory to decide. In HPSG, for example, it might be that each family corresponds to one general rule schema (e.g. the schema for coordination, or relative clause modification, or simple declarative clauses) instantiated with less and more marked feature values (e.g. +finite versus −finite for a declarative clause). The theory would need to take a clear stand on what is innate and invariant about each schema, which of its features never need to be specified, and what the defaults are for those of its features that can vary across languages.

Perhaps this is all just an exercise in optimism. Until we try it we won't know whether it really can meet C's tough criterion.[17] Certainly it will be important to see how this linguistically-based approach to the extent-of-generalization problem compares with an approach that puts extra power into the learning mechanism. Note that in the treelet system the learner arrives at a generalization on the basis of a single input sentence, just by mentally recording it in less than full detail, having stripped away all its UG-predictable properties. Once in the grammar, the underspecified description interacts with general UG principles to generate additional sentences. It seems to me that this could preserve the incremental baby even for those like C who prefer to throw out the parametric bathwater. In fact, if I'm right that even parameter setting must be implemented in this way, then we have a rapprochement between parametric and non-parametric approaches to learning which would have seemed impossible a few years ago. Treelets are the lingua franca of linguistics and they make it possible to explore in a relatively theory-neutral way how new licensing power can be added to a grammar in response to a novel input string encountered by a learner.

8. Conclusion

In this book, C has made a case for the claim that the existence of peripheral phenomena in natural languages (which, after all, might not have been the case) makes a difference to how the core should be described, altering the whole shape of linguistic theory and the conduct of linguistic research as well

[17] The projectability of an unlimited number of parameter value treelets is implicit in the model. A treelet can contribute to the generation of sentences if and only if it interacts successfully with the derivational principles of UG, which fill in the features it leaves unspecified and connect it up to other treelets. The principles of UG thus implicitly define the set of possible (usable) parameter values. Though the details will vary from one linguistic theory to another, any tree fragment (from a single feature to a complete tree) which can participate in successful derivations is a potential parameter value. This great wealth of alternatives should not overload processing as long as the parser gives priority to least marked analyses. But other damaging complications could well arise as the system is scaled up to include the periphery. Because grammatical phenomena differ with respect to what properties are predictable and what must be noted and stored, the specifications of default values for features must be context sensitive and possibly complex (Fodor 1992a). Also, the function which sums markedness values as treelets combine to fit sentences may not be simply additive.

as the modelling of language acquisition. Not everyone, I think, will want to accept these conclusions. This is a topic on which linguists have strong convictions, sincerely held but quite contrary. Those who want linguistics to reveal what the human mind is capable of may regard it as absurd to ignore the wealth of facts the periphery provides. Others may find it equally absurd to purport to study exceptions without first establishing the general principles they are supposedly exceptions to. And perhaps there are still others who are confident that the exceptions are only apparent.

Not being alarmist about odd cases is almost built into the practice of language description. Working linguists are trained, and rightly so, to resist as long as possible the thought that some particular fact might be irreducibly itself. The goal is to discover the design of human language, and design cannot be found unless it is looked for. The looking may have to be sustained and aggressive, and it is open-ended: there is no point at which it can safely be declared of any grammatical phenomenon that it will not yield to any further understanding. Could it ever be justified for someone to blow the whistle and assert that such efforts will never completely succeed?

I think C is right that one can make a rational estimate, which is neither doomsdayism nor faintheartedness but is totally in keeping with the goal of constructing an explanatorily adequate linguistic theory. If natural languages have ragged edges we need to know, because it is relevant to the evaluation of distinct but roughly co-extensive theories of core grammar. If one of them does better than the others at predicting the existence and character of natural language oddities, then (other things being equal) that is the theory we should prefer. Of course C could be wrong that we are NOW at a stage at which the true extent of deviations is discernible. His present judgment is that when all analysis is eventually completed there will be no smooth bedrock visible but just an awkward collection of odd-shaped sharp pebbles. This assessment may well be premature. There is interesting current work that sees the most general principles of the Minimalist Program at play in some of the constructions branded as exceptional by C (e.g. Bennis et al. 1998). Even so, it makes good sense to consider now what the consequences for acquisition would be if the irregular residue does prove to be substantial, as C conjectures. We can't just wait and let linguistic analysis reveal whatever rocks it may in its own good time, because that would put acquisition theory on hold indefinitely.

C suspects that research effort has been being wasted on a learning model (parameter setting) that is inherently at odds with the character of natural language. But he doesn't make a strong case for the direction he wants to go in instead. The disappointment of *Syntactic nuts* is that neither the linguistic nor the psycholinguistic part of the project arrives at a satisfactory destination. The account of learning is only a trailer for the more substantial presentation to come. And the linguistic analysis which is the main concern of the present volume stops short just where it could have broken new

ground. It stresses the EXISTENCE of marked phenomena, whereas it is their CHARACTER that can be most informative about the nature of grammars. Perhaps the book works best as a call to arms, to legitimize the study of linguistic peculiarities and draw in other linguists and learning theorists to help undertake it. There has been a research vacuum on this topic and C has stepped in to fill it on behalf of us all. If we would rather it were filled in some other way, it is up to us to set to work on it.

REFERENCES

Baltin, M. R. (1989). Heads and projections. In Baltin, M. R. & Kroch, A. S. (eds.), *Alternative conceptions of phrase structure*. Chicago: University of Chicago Press.

Battistella, E. L. (1990). *Markedness: the evaluative superstructure of language*. Albany, NY: State University of New York Press.

Bennis, H., Corver, C. & den Dikken, M. (1998). Predication in nominal phrases. *The Journal of Comparative Germanic Linguistics* **1**. 85–117.

Bertolo, S., Broihier, K., Gibson, E. & Wexler, K. (1997). Cue-based learners in parametric language systems: application of general results to a recently proposed learning algorithm based on unambiguous 'superparsing'. Paper presented at the 19th Annual Conference of the Cognitive Science Society, Stanford, CA, August 1997.

Borer, H. (1984). *Parametric syntax*. Dordrecht: Foris Press.

Borroff, M. L. (2000). Degree phrase inversion in the scope of negation. Ms., State University of New York at Stony Brook.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris Publications.

Chomsky, N. (1986). *Knowledge of language: its nature, origin and use*. New York: Praeger.

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

Clahsen, H. (1999). Lexical entries and rules of language: a multidisciplinary study of German inflection. *Behavioral and Brain Sciences* **22**. 991–1060.

Clark, R. (1992). The selection of syntactic knowledge. *Language Acquisition* **2(2)**. 83–149.

Clark, R. (1994). Finitude, boundedness and complexity. In Lust et al.

Culicover, P. W. & Jackendoff, R. (1999). The view from the periphery: the English comparative correlative. *Linguistic Inquiry* **30.4**. 543–571.

Culicover, P. W. & Nowak, A. (in preparation). *Dynamical syntax*. Oxford: Oxford University Press.

Dresher, E. & Kaye, J. (1990). A computational learning model for metrical phonology. *Cognition* **34**. 137–195.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition* **48**. 71–99.

Fodor, J. D. (1985). Why learn lexical rules? Paper presented at the 10th Annual Boston University Conference on Child Language Development, Boston, MA. [Circulated as 'The procedural solution to the projection problem'.]

Fodor, J. D. (1989). Learning the periphery. In Matthews, R. J. & Demopoulos, W. (eds.), *Learnability and linguistic theory*. Dordrecht: Kluwer Academic Publishers.

Fodor, J. D. (1992a). Learnability of phrase structure grammars. In Levine, R. (ed.), *Formal grammar: theory and implementation*. Oxford: Oxford University Press.

Fodor, J. D. (1992b). Islands, learnability and the lexicon. In Goodluck, H. & Rochemont, M. (eds.), *Island constraints*. Dordrecht: Kluwer Academic Publishers.

Fodor, J. D. (1994). How to obey the Subset Principle: binding and locality. In Lust et al.

Fodor, J. D. (1998). Unambiguous triggers. *Linguistic Inquiry* **29.1**. 1–36.

Fodor, J. D. (2001). Setting syntactic parameters. In Baltin, M. R. & Collins, C. (eds.), *Handbook of contemporary syntactic theory*. Oxford: Blackwell Publishers.

Fodor, J. D. & Crain, S. (1987). Simplicity and generality of rules in language acquisition. In MacWhinney, B. (ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

Gibson, E. & Wexler, K. (1994). Triggers. *Linguistic Inquiry* **25**. 407–454.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
Hyams, N. (1986). *Language acquisition and the theory of Parameters*. Dordrecht: Reidel.
Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
Lasnik, H. (1999). On feature strength: three Minimalist approaches to overt movement. *Linguistic Inquiry* **30.2**. 197–217.
Lightfoot, D. (1997). Catastrophic change and learning theory. *Lingua* **100**. 171–192.
Lobeck, A. (1995). *Ellipsis*. Oxford: Oxford University Press.
Lust, B., Hermon, G. & Kornfilt, J. (eds.) (1994). *Syntactic theory and first language acquisition: cross-linguistic perspectives*. Vol. 2: *Binding, dependencies, and learnability*. Hillsdale, NJ: Lawrence Erlbaum.
Merchant, J. (1999). *The syntax of silence: sluicing, islands, and identity in ellipsis*. Ph.D. dissertation, University of California, Santa Cruz.
Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press.
Pinker, S. (1999). *Words and rules: the ingredients of language*. New York: Basic Books.
Sakas, W. G. & Fodor, J. D. (2001). The structural triggers learner. In Bertolo, S. (ed.), *Language acquisition and learnability*. Cambridge: Cambridge University Press.
Valian, V. V. (1990). Logical and psychological constraints on the acquisition of syntax. In Frazier, L. & de Villiers, J. (eds.), *Language processing and language acquisition*. Dordrecht: Kluwer Academic Publishers.
Valois, D. (1991). *The internal syntax of DP*. Ph.D. dissertation, UCLA.
Wexler, K. (1998). Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua* **106**. 23–79.
Wexler, K. & Culicover, P. W. (1980). *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Author's address: The Graduate Center,
City University of New York,
365 Fifth Avenue,
New York, NY 10016,
U.S.A.
E-mail: jfodor@gc.cuny.edu