Alternative Minimalist Visions of Language

Ray Jackendoff
Tufts University

1. Goals and constraints

The primary goal of modern linguistic theory (at least in the circles I inhabit) is an explanation of the human language capacity and how it enables the child to acquire adult competence in language.[1] Adult competence in turn is understood as the ability (or knowledge) to creatively map between sound and meaning, using a rich combinatorial system – the lexicon and grammar of the language. An adequate theory must satisfy at least three crucial constraints, which I will call the Descriptive Constraint, the Learnability Constraint, and the Evolutionary Constraint.

The Descriptive Constraint is that the class of possible languages permitted by the theory of the language capacity must account for the adult's competence in any particular language. The Learnability Constraint is that there must be a way for the child to attain adult command of lexicon and grammar, based on the input the child receives. To the extent that research reveals adult competence to be more and more complex, there is more for the theory of acquisition to explain.

One strategy to cope with the apparent complexity of language is to try to minimize what has to be learned. There are two prongs to this strategy. First, one tries to minimize the complexity of adult grammar through generalizations that explain apparent complexity in terms of the interaction of simpler principles. This part of the strategy ("Minimize Complexity") is nothing more than standard scientific practice. A classic case is Chomsky's (1957) account of the complexities of the English auxiliary in terms of simple underlying structures plus the three transformations inversion, affix hopping, and *do*-support.

The second prong of the strategy addresses the Learnability Constraint: it constitutes an important hypothesis about the nature of language. The idea is that, in addition to minimizing the complexity of the adult language per se, one minimizes what the child has to learn from the environmental input, by packing as much remaining complexity as possible into an innate Faculty of Language, or Universal Grammar. For instance, in order to implement a treatment of the English auxiliary that involves transformations, it is standard to posit that the

[1]Much of the material in this paper is drawn from Culicover and Jackendoff 2005, and therefore I am indebted to Peter Culicover for his part in developing it, as well as for comments on an earlier version of the paper. I have also benefited from comments and questions at numerous presentations of this paper, including CLS.

possibility of transformations is made explicit by the Language Faculty.

This strategy, "Minimize Learning by enriching UG", is a hallmark of Principles and Parameters theory (Chomsky 1981): the hypothesis is that the child only has to acquire the settings of some finite number of parameters, from which all the properties of adult grammar follow. This is also a fundamental hypothesis behind Optimality Theory (Prince and Smolensky1993; McCarthy 2004), where the grammar consists of a finite set of universal violable constraints, and all the child must acquire is their ranking.

The strategy of enriching UG is however in tension with the Evolutionary Constraint. Of course, good science demands that Universal Grammar, like any other theory, be minimized. But there is an empirical issue as well. If there is an innate Universal Grammar which the child brings to bear on language acquisition, it must be built into the child's brain by standard (if hardly understood) mechanisms of biological development, and therefore it must ultimately be coded in some fashion on the human genome. Insofar as linguistic competence is not attainable by apes, the human genome must in relevant respects differ from the chimpanzee genome, and the differences must be the product of biological evolution. The richer Universal Grammar is, the more the burden falls on evolution to account for the genetic differences that make it possible for humans but not apes to acquire language. The Evolutionary Constraint, then, puts a premium on minimizing the number and scope of genetic innovations that make the human language capacity possible – and therefore on minimizing the richness of Universal Grammar.[2] This constraint is what gives the notion of "minimalist inquiry" its empirical bite.

At the same time, the force of the Evolutionary Constraint must be tempered with an important ramification. The language capacity might in principle be altogether specialized in the brain, or it might be simply the result of the interaction of more general human cognitive capacities, or it might the result of a mixture of special and more general capacities. To the extent that it involves more general capacities (the "Broad Faculty of Language" in the terminology of Hauser et al. 2002), less work remains for specialized capacities (the "Narrow Faculty of Language"). The general capacities certainly include the general auditory capacity (shared with apes), Theory of Mind (only partially present in apes), vocal imitation (not present in apes), the ability to understand pointing gestures (not present in apes) (Povinelli 2000, Tomasello et al. 2005), and general

[2]I recognize that this is a plausibility argument only, in that the mechanisms by which the genome builds the brain and by which the brain encodes innate structure are far from well understood. It is conceivable (though I think unlikely) that a richly structured Universal Grammar could be the consequence of a single genetic alteration. Ultimately this will be an empirically resolvable question.

capacities for pattern recognition and categorization. Although a complete account of human evolution must account for the uniquely human aspects of these more general capacities, in a sense the theory of language gets them "for free." So for the purposes of the Evolutionary Constraint, the crucial issue is what aspects of the language capacity are sui generis – totally unique, or at least specialized versions of more general capacities. One goal of a theory that posits such specialized aspects might well be to find a plausible evolutionary path along which they could have arisen.

On the other hand, there are many researchers inside and outside linguistics who claim there *are* no specialized aspects of the language capacity. If they are right, the Evolutionary Constraint has no force at all within linguistics. At the end I will lay out what parts of the language capacity I think are special for language and which I think the Evolutionary Constraint must speak to.

2. Two kinds of minimalism

One approach to satisfying the Evolutionary Constraint appears in the Minimalist Program (Chomsky 1995, 2002). The hypothesis is that the Narrow Faculty of Language is "perfect", in the sense that it satisfies the Descriptive Constraint – the adult ability to map between sound and meaning – with an absolute minimum of specialized machinery. The complexity of language is posited to arise only by virtue of the interactions of the Narrow Faculty of Language with independent properties of sound and meaning (that is, the Broad Faculty of Language). The goal, then, is to eliminate the richness of Universal Grammar as conceived of in Principles and Parameters theory, while still deriving the same or better empirical results.[3]

Another approach might be to take a different tack on the strategy Minimize Learning. The idea is that, instead of the theorist always attempting to eliminate grammatical complexities that the child must learn, sometimes one must acknowledge them as uneliminable complexities, and seek a theory in which they are learnable. One way to pursue this approach is the Simpler Syntax Hypothesis (Culicover and Jackendoff 2005): to minimize the elements of linguistic structure that the child cannot infer from the overt form. These are the aspects of language that are most difficult to infer from the input, and which hence place the greatest burden on innateness and therefore on evolution.

Notice, of course, that the meaning (or semantic/conceptual structure) of words, phrases, and sentences is entirely covert and cannot be eliminated from the

[3]This goal has been interpreted in some quarters as a recantation of Chomsky's decades-long insistence on the richness of Universal Grammar. However, reducing the complexity of Universal Grammar is an important goal in terms of both scientific methodology and acknowledgment of the empirical importance of the Evolutionary Constraint.

theory, since it is necessary in order to drive inferences. So the question is what *other* covert elements *are* dispensible, in particular covert structure within the syntactic component.

The most obvious such elements are null categories such as PRO and *pro*. But far more basic elements of mainstream linguistic theory are also suspect. Covert syntactic structures such as Deep Structure and Logical Form cannot be directly inferred from observable forms, so the child must innately know that they exist and how to construct them. Moreover, covert syntactic structures are taken to be related to observable forms by principles of movement which are also not inferrable from the input alone. Thus a possible goal for minimalist inquiry could be to eliminate movement from linguistic theory as well.

Such an attempt might well provoke an astonished outcry from practitioners of mainstream theory: How is it possible to do syntax without null elements, covert syntax, and movement? To begin an answer, we must examine the motivations for these aspects of syntactic theory, motivations that have been taken for granted since the 1960s. Intuition, enshrined in "folk linguistics", tells us that language closely reflects meaning: there is a strong match between form and function. A great deal of analytic philosophy, especially in the 20th century, has been devoted to showing that this intuition is false. One of the important ideas behind generative grammar from its inception is that, despite appearances, there is indeed a strong form-to-function match – but not at the surface. Rather, the match exists at a level of covert syntax: Deep Structure in the so-called Standard Theory (Katz and Postal 1964, Chomsky 1965), Logical Form in later versions (Chomsky 1981). Thus it can be claimed that "Deep Structure/Logical Form determines/encodes meaning."[4]

According to this hypothesis, which Culicover and Jackendoff 2005 call "Interface Uniformity", all mismatches between semantics and surface syntax are consequences of the relation between covert syntactic form and the surface. Thus, following a heuristic advocated by Katz and Postal 1964, every time one finds a semantic distinction (e.g. distinctions in aspect, illocutionary force, or quantifier scope), one assumes a homomorphic syntactic distinction in covert syntax, and then the descriptive problem is to account for how this covert distinction is obscured in the course of deriving the surface form. Conversely, when one finds two different surface forms that mean (about) the same, one assumes a common (or near-common) covert syntactic form, and the descriptive problem is to account

[4]Chomsky (e.g. 1965, 1966) has claimed that the Cartesian linguists also had such a view. My reading of the quotes he adduces as evidence is that these linguists actually recognized that form is radically mismatched to meaning. Chomsky's reading arises because they had no independent characterization of meaning, so they were forced to couch their analyses in terms of ordinary sentences. (See Jackendoff 2002, 73 for somewhat more detailed discussion.)

for how the observable differences arise in the course of derivation.

An alternative is to assume that Interface Uniformity is a violable default option in language rather than a rigid requirement. This allows the possibility of encoding mismatches between form and meaning directly in the relation of meaning to surface form. For instance, instead of mapping active and passive sentences from a common underlying syntactic structure which in turn is mapped to their common semantics, one might view the active and passive as alternative ways of mapping directly from meaning to surface form.

One might justifiably ask if this is really an alternative. After all, the theory has to encode the same mismatches between meaning and surface form either way. However, consider the resulting architectures:

Transformational grammar/Principles and Parameters/Minimalist Program:
meaning ⇐ *simple mapping* ⇒ covert syntax ⇐ *complex mapping*
⇒ surface syntax

Direct mapping:
meaning ⇐ *complex mapping* ⇒ surface syntax

From the point of view of minimizing the number of components of grammar (and hence the number of components with which the learner has to be pre-equipped genetically), there is a difference between the two, and the direct mapping theory is a priori superior. Still, a practitioner of mainstream theory might ask: How can a direct mapping theory be feasible? In fact, feasibility is not an issue: HPSG (Pollard and Sag 1987, 1994) and similar "monostratal" theories have worked out direct mapping theories in detail.[5]

3. Ways in which covert syntax is not minimal

On the other hand, an argument from feasibility is not sufficient: it still has to be shown that a direct mapping theory is empirically superior to a theory involving covert syntax. In order to begin to make a case for direct mapping, let us look at two bread-and-butter syntactic phenomena, passives and long-distance dependencies, and see how the two approaches compare.

3.1. Passive. Under the covert syntax approach, standard in mainstream generative grammar since the 1950s, passive sentences are the result of deforming active or active-like underlying forms. In particular, in passive sentences, a postverbal NP has been moved to surface subject position.

[5]HPSG has not been especially concerned with satisfying the Learnability or Evolutionary Constraints. Nevertheless, the theory is readily reinterpreted in these terms.

This approach raises two sorts of problems. First, a movement theory of the passive is committed to the existence of a surface NP position to which the underlying postverbal NP moves. But there are various constructions in which passive VPs appear without such a subject position.

(1) a. The man followed by the FBI is my brother.
b. Followed day after day by the FBI, John went slowly nuts.

In (1a), *the man* is not the subject of *followed*: it is the head of an NP, and *followed by the FBI* is a participial relative clause. In (1b) there is no overt subject of *followed* at all. So what has undergone movement? The movement theory, bolstered by Interface Uniformity, standardly posits a null (or deleted) NP that has undergone movement, and in fact we are used to taking such null elements for granted. Often it is even taken to be a virtue of classical generative grammar that it reveals to us the existence of such null elements, which could not have been expected from an inspection of the surface.

But positing such null elements makes the theory less than minimal. The adult must know more about syntax, namely that there is a null element in (1). In turn this makes the problem of acquisition more difficult: the child, hearing the pattern in (1), has to *know* that there is a moved NP, even though there is none to be heard, perhaps by virtue of Interface Uniformity, and therefore must have a Universal Grammar equipped to posit it and its movement.

In a direct mapping theory, this problem need not arise. The "understood" argument of *follow* in (1) is indeed present in meaning – this is what it means to say it is "understood." But that does not mean that this argument has to be present in syntax. Rather, the syntax-semantics interface can provide the possibility of pieces of meaning that are simply not expressed in syntax. (Almost) everyone agrees that there are such aspects of meaning, for example illocutionary force, metaphor, and the logical connections among sentences in discourse – though these aspects of meaning are usually called "pragmatics." The innovation, then, is only to extend such options to more basic grammatical phenomena. Again, rules of this sort have been developed within the context of monostratal theories, so their feasibility is not in doubt (see also Culicover and Jackendoff 2005, chapter 6).

A second problem for a movement theory of passive is that it requires the movement to be semantically "blind", since syntactic movement is by definition insensitive to semantic considerations, in particular to the meanings of individual words. But consider the "prepositional passives" in (2).

(2) a. The bed was slept in/on/*under/*beside by John.

b. The telescope was looked through/*inside by the technician.

The difference between the good and bad cases seems to hinge on the semantics of the surface subject. It appears that the prepositional passives are acceptable only if they express the "proper function" of the subject (in the sense of Millikan 1984 or the "telic quale" of Pustejovsky 1995), i.e. what it is *for*. Beds are for sleeping in or sleeping on; they are not for sleeping under or beside. Telescopes are for looking through; they are not for looking inside (of). Thus a characterization of these cases of the prepositional passive must refer to the semantics of the surface subject and its interaction with the semantics of the verb and preposition. A semantically blind movement theory does not have access to this information.

One might try to save the movement theory by suggesting that the good cases undergo syntactic or lexical "reanalysis", so that at the point where passive takes place, *sleep in* and *look through* "count as" syntactic verbs but *sleep under* and *look inside* do not (as in Hornstein and Weinberg 1981). The difficulty is that this reanalysis constitutes extra complexity in the syntax. Moreover it does not solve the problem, since now although passive is semantically blind, reanalysis is not. And reanalysis cannot just be sensitive to the local semantics of the verb and preposition, since acceptability depends on the choice of noun as well. Thus it cannot be a "lexical rule":

(3) a. *The giant telescope was slept in by the technician.
b. *The 4-poster bed was looked through by John.

Furthermore, there are purely syntactic difficulties with reanalysis, in that it presumes that the underlined parts of (4) "count as" a verb. (However, these cases do not depend on proper function.)

(4) a. They were <u>taken considerable advantage of</u>.
b. She was <u>taken good care of</u>.
c. You have been <u>spoken clearly to</u>.

Such difficulties can be overcome, but again only at the expense of adding complexity to the syntax that the child must presume in advance in order to be able to learn the passive construction.

In a direct mapping theory, this difficulty does not arise. To be sure, it is still necessary to explain why prepositional passives like (3) are dependent on proper function. But since the passive construction is taken to be not a deformation of syntactic structure, but rather a noncanonical mapping between meaning and syntax, the semantic features necessary to check for proper function are at least accessible to the rule: nothing in the theory requires that passive be

semantically "blind", and therefore no further complications need be added to the description of passive to incorporate these cases. (It still remains, of course, to explain how children come to be sensitive to these restrictions. But that issue arises in any approach.)

3.2. Long-distance dependencies. The same difficulties accrue to the movement theory of long-distance dependencies such as *wh*-movement. First, a movement approach commits the theory to uniformly moving a *wh*-phrase to the front, either in surface structure or at LF. This is plausible and intuitively attractive in examples such as (5).

(5) Which man did Fred see (t)?

But consider cases like (6).

(6) a. Zero tensed relatives: the man [I love (t)]
b. *for-to* relatives: the man [for you to vote for (t)]
c. Comparatives: The book is longer [than I thought it was (t)]

Here there is nothing at the beginning of the bracketed clause that could have been moved from the gap position at the end. Moreover, (6b) and (6c) even preclude an overt phrase at the front:

(7) a. *the man for whom for you to vote
b. *the book is longer how long than I thought it was.
*the book is longer than how long I thought it was.

Since the 1960s, the movement theory has always been saved by positing that a phrase has indeed moved, but either it is null or else it is deleted subsequent to movement. Again, the adult language user must know that this null element is present in covert syntax, and the child, in the absence of observable evidence, must know to posit this element and its movement – a complication in Universal Grammar.

A direct mapping theory, in contrast, treats long-distance dependencies in terms of a noncanonical mapping between semantics and syntax. The semantic form involves a logical operator such as question or (for relative clauses and comparatives) lambda-abstraction, which takes scope over an open proposition containing a bound variable. The mapping to syntax, depending on the language and the construction in question, may map a *wh*-phrase into the syntactic position corresponding to the bound variable (in situ *wh*), or it may map a *wh*-phrase into the position corresponding to the operator, as in (5), or it may simply leave the bound variable unexpressed, as in (6). There is no movement, and there are no unnecessary syntactic elements. To be sure, the well-known constraints on long-

distance dependencies still must be accounted for. But, as shown by the example of HPSG, this need not be done in terms of movement.

Moreover, there is a long tradition outside mainstream generative grammar of accounting for at least some of the constraints in terms of semantics, in particular information structure and theory of referential dependencies (e.g. Kuno 1987, Erteschik-Shir 1979, Kluender 1992, Csuri 1996, Van Valin 1998). Which brings us to the second difficulty with the movement theory: like the movement theory of passive, it is committed to long-distance dependencies being semantically "blind." However, as observed as long ago as Erteschik 1973, there are long-distance dependencies where lexical semantics makes a difference.

(8) a. What did Bill say/??grumble that Harry would like (t) for lunch?
b. The man who Bill said/*grumbled that Harry met (t) ordered a bagel.
c. This book is longer than you said/*grumbled that it was (t).

When the verb is an ordinary verb of saying or telling, the long-distance dependency is acceptable; but when it is a verb of manner of speaking (e.g. *grumble*, *growl*) it is unacceptable, although on the face of it there is no syntactic difference. One may attempt to syntacticize the difference by saying that the two kinds of verbs take different syntactic complements that have different consequences for extraction (as suggested by Chomsky 1981, for instance). But then the child must *know* to assign the verbs different complements in the absence of evidence, again a nonminimal solution, and this difference in complementation is still a matter of stipulation.

By contrast, in a direct mapping theory, the principles of long-distance dependencies are part of the syntax-semantics interface, so they have natural access to the lexical semantics of verbs. The theory still needs an *explanation* of why the differences in (8) occur -- and how they are learned. But at least they are in a domain that is natural to the principles in question.

In short, a direct mapping theory does not make any of these empirical problems go away, but it makes a solution possible in terms of minimal covert structure. By contrast, the movement theory requires null elements, a covert level of syntax, and particular hitches in the syntax that correlate in theoretically dubious fashion with the semantic peculiarities of the constructions in question. Thus the direct mapping approach, insofar as it can meet the Descriptive Constraint, requires less machinery within Universal Grammar through which the child can infer all this hidden structure. And to the degree that Universal Grammar is leaner, the theory stands a better chance of satisfying the Evolutionary Constraint.

More generally, we might recall an early result in generative grammar:

learning the structural descriptions of ordered transformations was shown to be one of the most severe obstacles to language acquisition (Wexler and Culicover 1980). This was argued to be a strong justification for a sophisticated Universal Grammar that builds in many general constraints on movement. However, as noted here, this makes UG nonminimal and less amenable to satisfying the Evolutionary Constraint. Thus, to the degree that a direct mapping theory can avoid positing movement, it has a hope of eluding the difficulties that movement poses to acquisition and therefore the need for UG constraints on movement. In principle this offers an advantage with respect to the Evolutionary Constraint.

4. Basic mechanisms for building syntactic structure

I want next to look at some foundational assumptions of the Minimalist Program, in particular how it builds syntactic structure, with the goal of seeing (a) how empirically adequate and (b) how minimal the theory actually is.

The approach to structure in the Minimalist Program (Chomsky 1995, 2002; Lasnik 2002) starts with the necessity of combining units recursively to form utterances. It is suggested that the simplest possible way of doing so is through an operation called Merge, which takes two constituents and sticks them together, giving the composite the label of one of the constituents:

Merge: Take A and B and create either $[_A$ A B] or $[_B$ A B] or $[_A$ B A] or $[_B$ B A].

The process of building a sentence begins by selecting a "numeration", a set of elements chosen from the lexicon. The elements of the numeration, plus composites built from them by Merge, are the constituents that undergo Merge to form an utterance.

What is in the lexicon, from which a numeration can be selected? Minimally, it has to include words and/or morphemes, coded minimally, i.e. nonredundantly. In the Minimalist Program's vision of language (along with much of mainstream generative grammar dating back to the 1960s), all redundancy is squeezed out of the lexicon into rules. Chomsky 1965 quotes Bloomfield 1933 approvingly: "The lexicon is really an appendix of the grammar, a list of basic irregularities."[6]

This all sounds very reasonable and simple, but I want to point out a number of assumptions that lie behind it.

- The organization of syntactic structure is to be characterized in terms of

[6]Whatever Chomsky's interpretation, it is not clear that Bloomfield considered the lexicon nonredundant, or even whether he even cared. The deeper point, as will be seen presently, is that Chomsky follows Bloomfield in regarding the lexicon as altogether separate from the grammar.

ordered derivations which put pieces together one after another. That is, the grammar is conceived as *derivational* or *proof-theoretic*.

- The optimal and minimal way to put syntactic structures together is in terms of strictly binary branching.
- The lexicon is nonredundant.
- There is a strict division between the lexicon and the grammar (the rules/regularities): these are entirely different sorts of entities in linguistic theory.
- Semantics is strictly locally compositional (or Fregean): the meanings of sentences are built up word by word, and the combination of word and phrase meanings is dictated by syntactic configuration.

An alternative vision of language, most aspects of which have appeared in the literature, involves replacing all of these assumptions. This section and the next will explore this alternative and compare it to the Minimalist Program.

4.1. Licensing by constraints rather than derivation by rules. First, the organization of syntactic structure can be formalized in terms of *licensing* rather than derivation. A structure is licensed by checking whether each piece of the structure and each relation among pieces of the structure is licensed by a relevant principle. This is the conception behind *constraint-based* (or *representational*, or *model-theoretic*) formalizations of grammar, the characteristic formalization of all the monostratal theories.

Is this any different from a derivational formalization? At first they might be thought to be notational variants, translatable one-to-one into each other. However, this proves true only to a first approximation, as will be seen.

A first case of difference between the two, pointed out by Postal 2004 (chapter 6), involves constructions such as (9) that admit non-English terminal elements:

(9) a. The space alien said 'klaatu barrada nikto' to Gort.
b. [Teenspeak:] And then, I'm all like, [gesture of exasperation].
c. The sign @ was invented in 1451.
d. *Sklerf* does not rhyme with *nikto*.
e. *Jean est mangé le pain* is ungrammatical in French.

These cannot be described using a derivation that starts with a numeration of lexical items, since they include elements that are not part of the lexicon of English. One would not want to counter by saying that *klaatu*, @, gestures of exasperation, and so on *are* all lexical items of English, for that swells the lexicon beyond all recognition. In fact, since arbitrary sentences and non-sentences of any other language can be inserted in (9e), the lexicon would have to be infinite.

Nor would one want to simply say that the sentences in (9) are ungrammatical, underivable by the grammar. That would imply, among other things, that linguists, who use metalinguistic utterances like (9d,e) all the time, are speaking and writing ungrammatically, making use of some unspecified auxiliary system of communication. It is important also that the content of the non-English elements in (9) is involved in the truth-conditions of the sentences, for example (9e) becomes false if *a* is substituted for *est*. This means that these elements cannot be treated as undigested lumps in the structure of the sentence, inserted under a wild-card category such as *Name* which somehow evades the numeration.

For a constraint-based model there is a more convenient escape hatch. The fact that sentences must be composed of words of the language is, like many other constraints, a violable default. Making it a default allows for particular constructions, such as those in (9), to license constituents that override this constraint and thereby "leak out of the language". Nevertheless, such constituents are not totally unconstrained: they still have selectional restrictions. For instance, the elements in quotative contexts such as (9a) have to be linguistic utterances in *some* (purported) language, the element in (9b) has to be an expressive linguistic or nonlinguistic gesture, and so on. And there is nothing to exclude these elements having content that contributes to the meaning of the sentence – it depends on the semantics of the construction in which they are embedded.[7]

Other differences between derivational and constraint-based theories will emerge as we go along.

4.2. Unify rather than Merge. We now turn to another assumption behind the MP, the notion that Merge is the simplest possible basis for building structure. In constraint-based approaches such as HPSG, LFG, and Construction Grammar, the fundamental combinatorial device is taken to be not Merge, but instead *unification* (Shieber 1986). Unification is sort of a Boolean union on feature structures: unifying A and B results in a composite that shares all common features of A and B and preserves all distinct features of A and B. (10) gives two examples:

(10) a. Unification of [V, +past] and [V, 3 sing] = [V, +past, 3 sing]
(not *[[V, +past] [V, 3 sing]]*, as with Merge)

b. Unification of [$_{VP}$ V NP] and [V, +past] = [$_{VP}$ [V, +past] NP]
(not *[[V, +past]][$_{VP}$ V NP]]*, as with Merge)

[7]One can imagine extending this argument to utterances incorporating code-switching. But this takes us beyond the scope of the present article.

Merge can be stated as a special case of unification, namely the unification of two elements with a piece of structure (or a grouping) containing unspecified terminal elements:

(11) Merge A & B = [A B]
Step 1: Unification of A and [*x y*] = [A *y*]
Step 2: Unification of B and [A *y*] = [A B][8]

Thus Merge can be reduced to Unify, but Unify cannot be reduced to Merge. Moreover, as seen in (10), Unify permits operations that cannot be simulated with Merge.

This does not prove that unification is *simpler*, conceptually or empirically, only that the two are not equivalent. However, unification appears to be a ubiquitous relationship in cognition. For instance, given that the visual system appears to process shape in one brain area, color in another, and motion in another (Koch 2004), an appropriate formal characterization would seem to be that a visual percept has a structure involving the unification of all these features. A sequential derivation in which the features are Merged into a tree structure seems considerably less likely (though of course not out of the question, pending a formal analysis of vision). Thus the Evolutionary Constraint, which seeks to reduce the evolutionary distance between humans and other apes, by deriving aspects of language from more general aspects of cognition, favors unification over Merge for the fundamental combinatorial operation of language. In short, like Merge, unification makes possible the representation of hierarchically embedded combinatorial structure, and moreover it comes "for free" from more general properties of cognition.

4.3. Redundant rather than nonredundant lexicon. Next consider the assumption that the lexicon is nonredundant. Although Chomsky has often asserted that the lexicon is nonredundant, no one to my knowledge has ever formulated a way to squeeze all the redundancy out of the lexicon into rules. We will shortly see many empirical problems for this view (cf. also Jackendoff 1975). Moreover, it seems a reasonable guess that redundancy is characteristic of the brain – that redundant representations help stabilize brain processing and make it more reliable. In particular, there is experimental evidence that high-frequency regular plural nouns are stored in the mental lexicon, even though they are totally redundant (Baayen et al. 2002). Thus, both on grounds internal to language and on grounds of the Evolutionary Constraint, there seems little reason to maintain

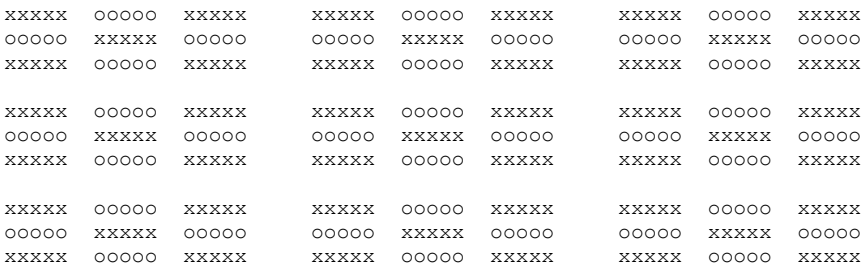
[8]Since unification is actually checking rather than derivation, the steps of relating A, B, and [*x y*] to [A B] are in arbitrary order.

this assumption, as long as the theory offers an account of regularity and productivity in the lexicon.

4.4. Multiply branching rather than binary branching trees. Now let us turn to the assertion that strictly binary branching trees are minimal. There is actually a tension here in what counts as minimal. Binary branching trees indeed have the minimal number of branches per node and the most restricted principles for constructing trees. But on the other hand they require a tree involving *n* terminal elements to have at least *n-1* nodes (one from Merging the first two elements, plus one more corresponding to the Merging of each remaining terminal element with the tree). By contrast, multiply branching trees aloow more complex possibilities for any single node, but they are simpler in the sense that they require fewer nodes in the tree. In the extreme case of a totally flat tree, there is only one node for any number of terminal elements. A priori there is no way to decide which desideratum is more important: fewer branches per node, or fewer nodes per tree.

However, if we again look at more general mechanisms of cognition, we find that multiply branching recursion is present elsewhere. Consider the visual array in (12).

(12)



This array is seen as little rows of five *x*s and *o*s, collected into squares made up of three rows. The squares are further combined into arrays of nine squares forming a bigger square, and three of these squares are collected into a row. Clearly the process could go on further, collecting a number of such rows into a bigger square, and so on. Thus the grouping of visual arrays represents a type of "discrete infinity" in Chomsky's sense. This shows that recursion is not unique to language, as Hauser et al. (2002) speculate (though recursion is likely unique to human language among natural communication systems).

The principles that govern visual grouping do not apply just to artificial stimuli such as (12): they are classical gestalt laws (Wertheimer 1923) that enable us to segment the visual field into regions and objects, and objects into parts and parts of parts. As Wertheimer observes, they apply not only to static visual grouping but also to temporal grouping in vision and audition, including music (Lerdahl and Jackendoff 1983). To the extent that other animals segment the

visual field more or less as humans do, recursive principles are not unique to humans (contra another speculation of Hauser et al.).

Notice next that there is no motivation for treating the groups in (12) as organized by binary branching, say [*x* [*x* [*x* [*xx*]]]]. Rather, they involve simple 5-ary and ternary branching. Thus, from the point of view of the Evolutionary Constraint, recursive n-ary branching comes "for free" as part of the human cognitive repertoire, and Universal Grammar gains nothing by constraining linguistic trees to binary branching. In fact, constraining linguistic trees to binary branching adds complexity rather than reducing it.

There are, of course, other respects in which linguistic trees are unique among cognitive capacities. Most obvious is the repertoire of elements that can enter into the trees, namely the set of syntactic categories. Syntactic trees also differ from recursive visual groupings in that each constituent has a head; they share headedness with, for example, syllabic structure (which is not recursive in the strong sense of unlimited depth of embedding) (see Culicover and Jackendoff 2005, Pinker and Jackendoff 2005, Jackendoff and Pinker 2005 for discussion) and with certain structures in music (Lerdahl and Jackendoff 1983).

In short, the argument for binary branching based on its alleged simplicity is spurious.

In fact in certain respects, binary branching forces extra complexity. For instance in the English verb-particle construction (*look the answer up*), it is difficult to find any non-theory-internal reason for bracketing like either [*look* [*the answer up*]] or [[*look the answer*] *up*]. Similarly, as argued by Chomsky 1965, there is no non-theory-internal reason to posit binary branching in the case of multiple prenominal adjectives such as *the handsome tall intelligent man*.

Two other arguments for exclusively binary branching have appeared in the literature. First, Kayne 1983 and Larson 1988 have claimed that strictly binary branching simplifies binding theory: if branching is binary, linear order can be eliminated from the conditions on binding, leaving only c-command, a structural condition. But is it such a simplification to eliminate linear order from binding conditions? After all, linear order is something that the learner gets "for free" in the signal, whereas c-command requires the learner to impute structure that is not visible. A priori, then, a minimal theory should use linear order for binding wherever possible, adding structural conditions such as c-command only where necessary.

Second, Haegeman 1992 has claimed that binary branching simplifies acquisition. Consider a learner encountering a string *A-B-C*. In principle, this might have any of the structures in (13).

(13) a. [A B C]
b. [[A B] C]
c. [A [B C]]

If UG stipulates that the branching is always binary and in a uniform direction, say (13c), then the learner does not have to decide which is correct. Therefore, says Haegeman, binary branching is simpler for the learner. However, another simple possibility is that UG stipulates minimal embedding, i.e. the flat structure (13a), as the default in the absence of evidence, and that the learner only imputes structure where the evidence pushes in that direction. Neither stipulation requires more of the learner than the other.

I conclude that uniform binary branching is not a minimal assumption in terms of either the Descriptive or Evolutionary Constraints, and there is good reason to permit multiply branching trees, as permitted by Simpler Syntax.

5. Addressing acquisition: What does the child have to learn?

So far we have questioned the first three of the assumptions behind the Minimalist Program's technique for building structure. To address the other two, we return to the primary goal of accounting for language acquisition and ask the question: What does the child have to acquire?

5.1. Words. Clearly the child has to acquire tens of thousands of words. This requires a learning procedure that cannot be reduced to the setting of a finite number of parameters.[9] For the Minimalist Program, this is not an important issue, since the lexicon is assumed to be separate from the grammar, and certainly can require its own independent learning theory. However, many linguistic phenomena call this assumption into question, as we will see in the course of this section and the next.

5.2. Fixed expressions. The child must learn a large number of fixed expressions, clichés, song titles, and so on, of the sort illustrated in (14).

(14) a. all hands on deck
b. Any friend of yours is a friend of mine.

[9]One might think that if Fodor 1975 is right that all word meanings are innate, word learning could amount to setting a finite (though *very* large) set of parameters. But Fodor's claim makes a mockery of the Evolutionary Constraint: it requires that all the word meanings be coded on the genome. And even if he were right, the learning procedure would be nontrivial: all the usual problems arise for the child trying to identify what meaning the speaker has in mind. Moreover, the meaning of a word does not completely determine its syntactic properties, and these are well known to present vexing problems for acquisition as well.

c. faster than a speeding bullet
d. weapons of mass destruction
e. Can't Buy Me Love
f. The Grapes of Wrath

My own interest in such expressions (Jackendoff 1997a) was piqued by the television game show *Wheel of Fortune*, in which contestants guess expressions letter by letter. A large proportion of the multi-word puzzles are fixed expressions like (14). Considering that the show uses four or five puzzles a day and has been on the air for decades with no sign of running out of puzzles, we conclude that such expressions number in the tens of thousands, an order of magnitude similar to the number of words.

One might object that these items do not belong in the lexicon. Still, they are familiar to speakers, and distinguishable from novel expressions with the same grammar, such as (15). This is why contestants can guess them.

(15) a. all feet on deck
b. Any sandwich of yours is a sandwich of John's.
c. slower than a lumbering tortoise
d. weapons of minimal destruction
e. can't buy me popcorn
f. the grapefruit of scorn

So at the very least the fixed expressions belong in an "expression-icon." Is this part of knowledge of language? What else could it be, given that these items are made up of phonological, syntactic, and semantic structures? The cognitive capacity responsible for learning them has to be (at least partly) the language capacity.

Consider that a fixed expression is made up of standard words, combined in standard syntactic structures, and with meanings composed (more or less) canonically from the meanings of their parts. This means that, aside from the fact of its existence, a fixed expression is completely redundant. Yet there is no way to squeeze its redundancy out into rules, without expunging the item entirely from the knowledge of language. Hence there is no way to strictly segregate rules from an idiosyncratically stored material, as assumed by mainstream generative grammar. I know of no discussions of fixed expressions that address this issue within mainstream thinking, in particular within the Minimalist Program.

5.3. Idioms. The child must also acquire idioms of the language such as (16). As is well known, these run into at least the thousands.

(16) a. kick the bucket
b. take NP to task [note: discontinuous]
c. son of a gun
d. down in the dumps

Idioms too are largely redundant, being (mostly) made up of existing words (mostly) arranged in standard syntactic structures (with exceptions such as *be that as it may* and *for the most part*). The only part that is not redundant is the meaning, which cannot be predicted from the meanings of the words.

The meaning of idioms presents a difficulty for building structure by means of Merge (and for classical lexical insertion as well). Recall that Merge introduces words individually into trees, and the meaning of a phrase is to be built up from the meanings of its constituent words. In the case of an idiom this is impossible: the meaning cannot be determined unless all the words are present together.

One can imagine a couple of ways to deal with this. One would be to say that idioms are actually introduced under a lexical category – for instance, *kick the bucket* is actually not a VP, as it looks, but simply a V (this is proposed in Culicover 1967 and obliquely in Chomsky 1981). However, an idiom such as *take NP to task* has an open argument position that looks suspiciously like a direct object. Is one to say that a V can have an open argument position within it? Or does *to task* obligatorily extrapose from the V, coincidentally landing just where a PP argument would be expected? Both these solutions (a) introduce further complexity that the child must figure out by dint of Universal Grammar and (b) are patently attempts to force the theory into line with uncomfortable data.

Another solution would be to say that *kick* has a second meaning, 'die', that only can be used in the context of *the bucket*. Coincidentally, in just this context, *the* and *bucket* must also have second meanings that happen to be null. Then the meaning of the idiom can in fact be introduced with a single word. The difficulty with this solution is its arbitrariness. There is no non-theory-internal reason to concentrate the meaning in just one of the morphemes. Moreover, it requires *take* to have the reading 'criticize' in the context of *to task*, 'rob' in the context of *to the cleaners*, 'exit' in the context of *a powder*, and so on – and each of the words in these phrases has to be meaningless in the context of *take* and the rest of the idiom. In other words, a lexicon populated exclusively by words has to be furnished with voluminous cross-reference in order to account for the existence of idioms, essentially building the description of an idiom into each of its words – a massive redundancy.

The issue of which idiom chunks can undergo movement has received some discussion in mainstream generative grammar, but (to my knowledge) the

more fundamental issue of how idioms get into syntactic structure in the first place has not been seriously addressed.[10] In view of the fact that there are about as many idioms as adjectives, this is a yawning gap in the empirical coverage of mainstream theory.

These problems do not arise in a constraint-based unification formalization of structure. *Kick the bucket* is listed as a lexical VP meaning 'die', and it unifies straightforwardly with an ordinary VP structure; *take NP to task* unifies with an ordinary VP, leaving an open NP slot to be filled by the direct object. Moreover, the partial redundancy of idioms is not a problem: it is just what one might expect. Thus we have uncovered another difference between derivational formalizations employing Merge and constraint-based formalizations employing Unify.

5.4. Noncanonical utterance types. English (and I presume other languages) contains perhaps a couple dozen utterance types whose syntax does not fall under standard X-bar theory. Here are some examples:

(17)	a. *PP with NP*:	Off with his head! Into the trunk with you!
	b. *How about X?*	How about a cup of coffee? How about we have a little talk?
	c. *NP+acc Pred?*	What, me worry? Him in an accident? John drunk? (Akmajian 1984)
	d. *NP and S*	One more beer and I'm leaving. One more step and I shoot. (Culicover 1972)
	e. *Scores*	The Red Sox 4, the Yankees 3.
	f. *The more ... the more*	The more I read, the less I understand. (McCawley 1988, Culicover and Jackendoff 1999)

These utterance types are part of the knowledge of English, and the learner must acquire them. There are two possibilities for their structure: either they are more or less brute force exceptions, listed in the grammar in terms of their surface structure, or they are derived by massive and exceptional deformation from more canonical forms. In at least two cases, (16d) and (16f), the literature cited argues that the latter solution is not viable. And consider its consequence for acquisition: the child not only must learn the surface form, but must infer a canonical covert syntactic form and the complex details of the derivation.

I conclude that the best solution is more or less brute force: these

[10]One possibility would be to extend Merge, so that it could apply idiomatic phrases as a whole. However, this would still present difficulties with discontinuous idioms such as *take NP to task*.

utterance types are learned and stored as exceptional pieces of syntax, complete with a special interpretation. That is, they are like idioms except that they have exceptional syntactic structure as well. Such sentence types are among the original motivations for Construction Grammar (e.g. Fillmore, Kay, and O'Connor 1988), in which the grammar contains explicit *constructions*, stored pairings of syntactic structure and meaning.[11] For the moment we leave open whether these are to be part of the rules, part of the lexicon, or a separate component that some call the "constructicon."

In a constraint-based unification formalism, it is easy simply to admit these structures as another sort of idiom, integrated into sentences by the usual processes. It is less clear how they can be dealt with in terms of Merge, where sentences are built from the bottom up, and their meanings are determined solely by the meanings of the words. In particular, such a treatment does not provide an obvious account of (a) how the top of the tree comes to have one of the acceptable noncanonical forms but not one of the unacceptable ones, and (b) where the meaning comes from. To say, for example, that *with* has some special meaning just in case it occurs in the context *PP__NP* in effect conceals the construction inside the lexical entry of *with* – a hidden notational variant of the constructional solution. Moreover, in the case of (17c,e) there is no characteristic morpheme to which the meaning can be attached and in which the construction can be hidden. I hesitate to work through straw men beyond this point, because such utterance types are never (to my knowledge) addressed in P&P and MP literature. In any event, this is a further case in which derivational and constraint-based formalism differ in their consequences.

5.5. Syntactic nuts. In addition to the noncanonical utterance types, English has many noncanonical phrase types ("syntactic nuts", to use Culicover's (1999) term) about which speakers have perfectly clear intuitions – and which of course children must learn. Here is a sample of four.

• *Numbers*. Every language has a peculiar little part of the grammar for productively characterizing numbers.

(18) three hundred fifty-five billion, fourteen million, one hundred twenty-five thousand, six hundred thirteen and five sixteenths

• *Names of geographical features*. English has a collection of nouns such as *ocean, river,* and *lake* that denote geographical features. These can be combined with names to create names of geographical features. Each noun combines with

[11] In some formulations of Construction Grammar (e.g. Goldberg 1995), the grammar is taken to consist *entirely* of meaningful constructions.

names in a characteristic fashion; there are altogether four frames, of which each noun partakes of one or two:

(19) a. the Atlantic Ocean, the Hudson River, the Mediterranean Sea
b. the Bay of Biscay, the Gulf of Aqaba, the Sea of Azov
c. Arrowhead Lake, Wissahickon Creek, Laurel Hill, Loon Mountain, Biscayne Bay
d. Lake Michigan, Mount Washington

These frames are productive. For instance, if one wanted to name a new ocean (say on Titan or in a science fiction epic), one could call it *the Bojoric Ocean* but not **the Ocean of Bojoric*, **Ocean Bojoric*, or **Bojoric Ocean*. A new lake could be *Gloggle Lake* or *Lake Gloggle* but not **the Gloggle Lake* or **the Lake of Gloggle*; a new mountain could be *Mount Snoopy* or *Snoopy Mountain* but none of the other six possibilities. Thus these combinations are rule-governed.

What is the appropriate way to characterize these frames? Frames (19a,b) look like noun phrases, especially since adjectives can be inserted (*the majestic Hudson River*), while frames (19c,d) look like compound proper nouns (e.g. *my beloved Loon Mountain*, parallel to *my beloved Dr. Bartolo*). So it is unclear whether the rules in question belong "in the grammar" or "in the lexicon." And if they are "in the grammar", each word must still explicitly list which frame(s) it fits into, perhaps through some variant of subcategorization – in effect coding the rules within the lexicon as well as within the grammar. Thus the strict division between phrasal rules and lexicon begins to blur.

In terms of learning, the simplest solution is that the child learns each word along with the frame(s) in which it can occur, whether compound or phrasal. There seems little reason to suppose that the child infers a uniform covert underlying structure for all of them, in accord with their uniform semantics, then develops movement rules (and sometimes insertion of *the*) in a lexically specific fashion. That would just conceal the same distinctions in a layer or two of derivational complexity.

• *Focus reduplication*. Reduplication has typically been studied in the context of "exotic" languages. But English too contains a reduplicative construction, quite common in colloquial speech (though rare in writing), called *cloning* by Horn 1993 and *focus reduplication* by Ghomeshi et al. 2004.

(20) a. You make the tuna salad, and I'll make the SALAD-salad.
b. Would you like some wine? Would you like a DRINK-drink?
c. Do you LIKE-her-like her?
d. Are you guys, um, LIVING-together-living together?

The construction has two distinct meanings, distinguished by pragmatics. The first reading denotes 'standard/stereotypical subtype of reduplicant', as in (20a), where *SALAD-salad* means 'green salad', i.e. the stereotypical salad. The second reading denotes 'special or extreme subtype of reduplicant'. For instance, in (20b), *DRINK-drink* denotes 'drink of hard alcohol'; in (20c), *LIKE-her-like-her* denotes 'like her a whole lot'; and in (20d), *LIVING-together-living together* denotes 'living together as sexual partners' as opposed to being mere housemates.

Unlike most reduplication rules discussed in the literature, this one is not confined to single words (see Lidz 2001 for another such case, in Kannada). For instance, it can reduplicate a verb plus a cliticized object pronoun (20c) or a verb and complement that form an idiom (20d). Some speakers can even reduplicate a verb and its object: I recently heard *I haven't really SEEN-Boston-seen-Boston.*

It is not clear what the structure of such phrasal reduplications is. Ghomeshi et al. 2004 explore in some detail how the structure might be produced, both using a Minimalist approach and using a constraint-based approach. They encounter considerably less artificiality in the latter but still leave unsolved puzzles. In both approaches, the best solution seems to be a syntactic affix which is associated with the appropriate meanings, and whose phonological structure is an operator that says roughly 'reduplicate the stem I am attached to'. The latter seems necessary in any event for any reduplicative phenomenon.

For present purposes, the point is that it is hard to decide whether focus reduplication is to be regarded as a rule in the grammar or "in the lexicon". 'Reduplicate stem' certainly looks like a productive rule, but it is triggered by a particular morpheme, which is not expected of a rule "in the grammar." However, since the reduplication applies to syntactic constituents, not just to words, the rule cannot take place "in the lexicon" "before" lexical insertion. Again the strict distinction between lexicon and grammar blurs, in a different way than in the previous case.

• *N-P-N construction*. English has a minor construction with the structure N-P-N (Williams 1994):

(21) a. dollar for dollar
b. face to face
c. house by house
d. month after month
e. book (up)on book

The choice of preposition is highly restricted: those in (21) are the only productive ones, although there are a few idioms with other prepositions such as *limb from limb* and *hand in hand*. The two nouns must always be identical,

except again in a few idioms such as *hand over fist* and *tongue in cheek*.

There is little evidence for further internal structure. In particular, there is no evidence for *P-N* forming a PP, excluding the first noun: there is no way to separate *P-N* as a constituent or interpose something before the *P*. Determiners are impossible (22); adjectives can appear before the nouns only in very strictly constrained fashion (23).

(22) a. *a/the face to a/the face
b. *a day after the day
(23) a. month after miserable month
b. miserable month after miserable month
c. *miserable month after month
d. *miserable month after lovely month

The construction can appear in adverbial or adjectival positions (24-25); N-P-N expressions with *after* and *(up)on* can also appear in NP contexts, although the others cannot (26).

(24) a. We matched their contributions dollar for dollar.
b. Face to face, Bill was frightening.
c. We inspected the chimneys house by house.
(25) a. a dollar-for-dollar exchange
b. a house-by-house inspection
c. month-after-month boredom
(26) a. Student after/*by student failed the oral exam.
b. They flunked student after/*by student.
c. We accumulated volume upon volume of phonology texts.
d. Student after student who flunked the exam complained about it.

(26c,d) show also that when the construction appears in NP contexts, it can take appropriate postnominal complements and modifiers.

Detailed examination of the construction (Jackendoff to appear) reveals a thorough interpenetration of regular, semiregular, and idiosyncratic characteristics. Each preposition has a special range of meanings in the construction, and each meaning shows a different degree of regularity. There is nothing in standard X-bar theory or other treatments of canonical phrase structure that predicts the characteristics of the construction: the repetition of the noun, the defective phrase structure associated with it, and its apparent syntactic category (AP/AdvP/NP). If a new phrase structure rule were added to the usual armamentarium (parallel to, say, the conjunction schema), it would not predict that the regularity and productivity of the structure would be so dependent on the particular choice of preposition. A theory based on bottom-up Merge faces the

further problem of guaranteeing the necessary identity of the two nouns and, where present, adjectives. Whatever the analysis, the learner has to learn all the details of this construction, and there seems little likelihood that they are predictable through some finite set of parameters or lexical features stipulated by Universal Grammar.

The upshot of this section is that the language is riddled with little patterns that (a) must be learned, (b) do not follow naturally from anything in UG (at least standard conceptions thereof), (c) present difficulties for a theory that builds structures in terms of Merge, (d) in many cases fall uncomfortably between rules of grammar and lexical rules. Presumably every language has lots of these sorts of things. (See Culicover 1999, Culicover and Jackendoff 2005, Goldberg 2005 for more examples and much more discussion.)

6. Can these examples be disregarded as "peripheral"?

An advocate of the Minimalist Program might be tempted to respond to these phenomena by consigning them to the status of "merely" "peripheral". This term of course adverts to Chomsky's (1981) distinction between "core" and "periphery" in language: the problem of language acquisition and the goal of "perfection" are supposed to apply only to "core" grammar, which includes such well-studied phenomena as argument structure, passive, raising, long-distance dependencies, and basic cases of binding. Non-core phenomena such as idioms, syntactic nuts, and so forth are irrelevant to the problem of characterizing core grammar, and so it is a reasonable scientific idealization to set them aside for the present.

There are at least four reasons for rejecting such a response. The first is that it explicitly abandons the Descriptive Constraint: the theory of core grammar is not responsible for the structure of the language any more – only for part of it. The rest is postponed indefinitely – including even the structure of the lexicon and the treatment of the tens of thousands of idioms.

The second reason for rejecting this response is that one cannot know in advance what belongs to core and what to periphery. Without exploring the boundary, one cannot make a judgment except on question-begging grounds ("It doesn't have an account in the theory, so let's make the hypothesis that it is peripheral.") Many of the phenomena mentioned in the previous section have such a borderline status.

The third reason for rejecting this response (due to Culicover 1999) concerns acquisition. It is not just the core phenomena that present a problem for the learner: everything has to be learned somehow or another. At the very least, we know that a learning procedure for the core is not going to be able to account for lexical acquisition. So suppose we discover a learning procedure that can

acquire the words of the language and all the "peripheral" grammatical phenomena. What is to say that this procedure cannot learn the "core" as well, making a separate "core UG" unnecessary? If we adopt a research strategy of idealizing away from the "periphery" and its acquisition, we will never investigate this question, which threatens the notion of "core" with irrelevance.

These three reasons for rejecting the appeal to the core-periphery distinction are methodological. The fourth reason is empirical: there are syntactic nuts that interact closely with indisputably "core" areas of grammar. Here are two.

6.1. VP constructions. Goldberg 1995, Jackendoff 1990, 1997b and Goldberg and Jackendoff 2004 discuss a number of VP constructions in English, some of which are illustrated in (27).

(27) a. He sang/drank/slept/laughed his head off.
(V his head off = 'V excessively')
b. Bill belched/lurched/joked/laughed his way out of the restaurant.
(V X's way PP = 'go PP while/by V-ing)
c. Sara slept/drank/sang/laughed the whole afternoon away.
(V NP away = 'spend NP amount of time V-ing')
d. Bill drank the pub dry. (V NP AP = 'make NP AP by V-ing')
e. The trolley squealed/rumbled around the corner.
(V PP = 'go PP, motion inducing V-ing sound)

In these constructions, the verb does not license the constituents of the VP (underlined in (27)). Singing and sleeping have nothing to do with one's head, belching and squealing are not verbs of motion, and so on. Moreover, the postverbal NP in (27a-d) is not some sort of syntactic adjunct. Rather, it appears to occupy and thereby "use up" the object position, since the verb is prohibited from taking an object more appropriate to its meaning:

(28) a. *He drank scotch his head off.
b. *Bill told jokes his way out of the restaurant.
c. *Sara sang folk songs the whole afternoon away.
d. *Bill drank bourbon the pub dry.

Goldberg and Jackendoff argue that these constructions are to be thought of as a sort of idiom: a VP structure in which the verb, although still the syntactic head of the VP, functions as an argument. That is, the syntactic and semantic glosses in (27) are essentially the way the constructions are listed in the grammar. Such an account is readily statable within a constraint-based unification formalism that admits constructions of the sort discussed in the previous section. Aside from the resultative (27d), no account of constructions of this sort is known to me in the

literature of P&P/MP; Goldberg and Jackendoff point out many obstacles that such an account would have to surmount.

If there is a core-periphery distinction in the grammar, these constructions surely belong in the periphery. Yet they intimately involve the determination of argument structure, which unquestionably belongs to the core.

6.2. "Signatures" of long-distance dependency constructions. One of the major advances of classical generative grammar was the recognition that all long-distance dependency constructions share similar constraints (Ross 1967). These similarities came to be captured in terms of a single derivational process, *Move wh* (Chomsky 1977), which in turn was later subsumed by a more general process, *Move α* (Chomsky 1981). This further has come to be thought of a special case of Merge (Chomsky 2001).

What has been lost in this development is the differences among the various long-distance dependency constructions, particularly the configurations permissible at their left edge. (29) summarizes some of these; they involve not just the fronted phrase, but also the status of the complementizer, the subject, and the tense, and whether inversion takes place. We might call this configuration the "signature" of the construction.

(29)
- a. *Wh*-questions: *wh*-phrase at front, allows preceding preposition (*From whom...?*), requires inversion in main clauses *except* with *how come*, which prohibits it.
- b. Infinitival indirect questions: *wh*-phrase at front (*I wonder who to see*), allows preceding preposition (*I wonder to whom to speak*), disallows *for*-subject (**I wonder who for Bill to see*)
- c. *Wh*-exclamatives: *what (a)* or *how* at front (**Who we saw!)*, does not allow preceding preposition (**From what a gentleman we received a letter!*), no inversion
- d. Tensed relative clauses
  - i. *wh*-phrase at the front, allows preceding preposition (*the man [to whom you spoke]*) – or
  - ii. *that* at the front, no preceding preposition (*the man [that you spoke to]*) – or
  - iii. zero at the front (*the man [you spoke to]*); normally subject cannot be the gap (**the man [t spoke to you]*)
- e. Infinitival relative clauses
  - i. *wh*-phrase at the front, must be preceded by preposition (*the man [to whom to speak]*; **the man [who to speak to]*) – or
  - ii. *for*-NP subject at the front, no *wh*-phrase (*the man for you to hire;* **the man with whom for you to talk*) – or
  - iii. bare infinitive at the front (*the man to see*)

f. Free relatives
wh-phrase or *wh-ever* phrase at the front, no preceding preposition (*what you saw; whatever you saw; *to whomever you speak*)

g. *So* topicalization
so-phrase at front, inversion in main and subordinate clauses (*so many men did he see that ...*; *I can assure you that so many men did he see that ...*)

h. *The more* constituents
the more/the A-er phrase at front, must be in parataxis with another such clause (or a comparative clause) (*the more you read, the less you understand; you understand less, the more you read*)

Let's notice just one detail here, as representative of the problems posed by (29). In infinitival relatives (29e), a fronted *wh*-phrase must be preceded by a preposition: *the man to whom to speak* but **the man whom to speak to*. By contrast, infinitival indirect questions (29b) permit a *wh*-phrase to appear with or without a preceding preposition: *I wonder to whom to speak*; *I wonder whom to speak to*. Moreover, both tensed relatives and tensed indirect questions allow both possible forms. Thus there appears to be nothing principled in the syntax or semantics of these constructions that predicts the particular constraint on infinitival relatives. Certainly it is not predictable from a general rule that just says to move a phrase to the front. Something special has to be learned to account for speakers' command of each of these constructions.

In the early days of generative grammar, prior to Move α, each of these constructions was thought to be the product of a different rule. I recall how as graduate students in the late 1960s we tried to account for them through movement rules with very precise structural descriptions, supplemented by very particular deletions that took out just the right pieces after they were moved. It was not easy. As far as I can tell, the attempt to account for these idiosyncrasies was largely abandoned with the onset of Principles and Parameters theory, as the search for plausible crosslinguistic parameters came to be a more pressing concern. To be sure, one must choose one's battles. But if anything, an account of the signatures of long-distance dependencies has receded further in the face of the increasing concentration on generalization and "perfection."

In a constraint-based (i.e. non-movement) theory, the signatures can be learned as idiosyncratic phrase-structure configurations associated with surface forms, that is, as syntactic idioms with particular constructional meanings. Such an approach has been explored in some detail by HPSG, especially Sag 1997 and Ginzburg and Sag 2000. The generalizations concerning long-distance dependencies are not a consequence of movement, but rather a consequence of relating the signature to a gap within the clause, along lines discussed earlier in section 3.2.

Again, the idiosyncrasies of signatures might be considered issues of the "periphery" and hence not a major concern for now. Still, they are intimately intertwined with the core phenomena of long-distance dependencies, and therefore an adequate account of the core cannot disregard them.

The conclusion illustrated by these two cases and argued in more detail by Culicover and Jackendoff 2005 is that it is a mistake to disregard the "periphery" of language and concentrate on the "core." Rather, summing up this section and the last,

There is an unbreakable continuity between "core" and "peripheral" phenomena and between the "core" generalizations of language and complete lexical idiosyncrasy.

This conclusion has been reached by several independent lines of research, in particular HPSG, Construction Grammar, Cognitive Grammar, and item-based acquisition (Tomasello 2003) as well as Simpler Syntax. If correct, it is a deep and important insight that forces a major rethinking of our vision of language. Such a rethinking is impossible within the assumptions of mainstream generative grammar.

This disparity between constraint-based theories and mainstream generative grammar leads to the further conclusion:

Derivational and constraint-based formalizations are not notational variants, and constraint-based theories are more adequate for expressing the continuity in the texture of linguistic structure.

Moreover, to be *very* blunt, if we review the basic properties of implementation of phrase structure in the Minimalist Program – binary Merge beginning with a numeration from the lexicon, a nonredundant lexicon, the distinction between the core and periphery, the distinction between the lexicon and rules, and semantically blind movement – we have found that all of them are either formally non-minimal, empirically inadequate, or methodologically unsound. The vision of language in mainstream generative grammar has led to huge advances in our understanding over the past fifty years, but it is time to move on.

7. Learning and innateness – satisfying the Evolutionary Constraint

Practitioners of the Minimalist Program might still appeal to the Learnability Constraint: In order to meet the Descriptive Constraint, the constraint-based theory requires a proliferation of rules. How are all these rules acquired, particularly if another goal is to keep the innate components of the Narrow Language Faculty at a minimum?

The first response to such critics, of course, would be to firmly remind them that contemporary derivational theories have sorely neglected the Descriptive Constraint, so there is no way of knowing what problems they raise for acquisition. But there is a more positive response as well.

Within a constraint-based theory, what is the difference between a word and a rule? Both are pieces of structure stored in the lexicon/constructicon. What makes something specifically a rule is that it has variables as part of its structure, and these variables must be satisfied by unification with something else. (30) illustrates the smooth transition from an idiosyncratic structure to very general principles of language (variables within these structures are notated in italics).

(30) a. VP idiom – no variables: [$_{VP}$ [$_{V}$ kick] [$_{NP}$ [$_{Det}$ the] [$_{N}$ bucket]]]
b. VP idioms with variable: [$_{VP}$ [$_{V}$ take] *NP* [$_{PP}$ [$_{P}$ to] [$_{NP}$ task]]]
[$_{VP}$ *V* [$_{NP}$ *pro*'s head] [$_{Prt}$ off]]
c. VP structure with more variables: [$_{VP}$ *V (NP) (PP)*]
d. Head parameter for VP: [$_{VP}$ *V* ....]
e. X-bar theory: [$_{XP}$ ... *X* ...]

(30a) is a stereotypical idiom: a VP with all the phonological material filled in and a stipulated meaning. The examples in (30b) introduce a variable. *Take NP to task* is an idiom with a direct object to be filled in both in syntax and interpretation; *V pro's head off* is one of the aforementioned constructional idioms, in which the verb is a variable and fits into the interpretation of the idiom. (30c) is composed entirely of variables; it is a notational variant of a standard phrase structure rule for VP. (30d) bleeds more structure out, leaving only the stipulation that the verb is initial in the VP – in effect the setting of the head parameter for the English VP. Finally, (30e) says that an XP has an X somewhere within it; this is a way of stating X-bar theory, the hypothesis that a phrase has a head of the appropriate category.

(30) illustrates the larger point that the "core" principles of phrase structure are general schemata along the lines of (30d,e), whereas more idiosyncratic rules and fully specified items are usually specializations of these schemata. That is, these items fall into an *inheritance hierarchy* (to use a term common in constraint-based frameworks): (30a,b) are special cases of (30c), (30c) is a special case of (30d), and (30d) is a special case of (30e). On the other hand, there can also be idiosyncratic rules that are *not* specializations of more general principles, for instance the N-P-N schema (e.g. *day after day*), which is not an instance of X-bar theory.

Inheritance hierarchies are not specific to language: they are more broadly useful for characterizing knowledge of nonlinguistic categories (e.g. birds and

mammals are special cases of animals; cats are special cases of mammals; my late cat Peanut is a special case of cats). Thus this fashion of arranging items in memory comes "for free."

The Minimalist Program has speculated that there is only one grammar – that all language-specific differences are coded in the lexicon. In a way, the treatment of words and rules in terms of inheritance hierarchies realizes this speculation trivially. The only "procedural" part of the grammar is Unify, the constraint-based counterpart of Merge, and the lexicon contains not only all the words but all the rules as well!

Turning back to acquisition, this formulation of rules lends itself to an "item-based" approach to acquisition (Clark 2003, Tomasello 2003, Culicover and Nowak 2003, but also Braine 1971). The idea is that particular constructions are learned holistically – even particular VPs like *eat a cookie*. The collection of such items is constantly subjected to a "sifting" operation that seeks structure in the collection. When multiple items have a part in common, this operation creates a new item (i.e. a rule) that consists of the common part plus a variable that corresponds to the parts that differ from item to item. By applying this operation recursively to its output, gradually more and more general schemata are created.[12]

Notice that such a procedure is feasible only in a formalization of language in which words and rules are stated as data structures in a common format. A theory that contains rules like Move, which are qualitatively different from words, does not lend itself to this sort of learning, and thus is more likely to require a transcendental leap in acquisition, supported by UG.

How is this procedure different from plain analogical learning? There are two major differences. First, the learning procedure does not *just* make analogical connections between existing items. Rather, it creates *new* items that explicitly capture the similarities among existing items. These new items contain variables (or slots), which is what makes them rule-like and potentially open to productive generalization. As Marcus (1998, 2001) has shown, the notion of a formal variable is crucial to characterizing the combinatoriality of language. The inability of connectionist models of analogical learning to encode the notion of a variable is fatal to their aspiration of accounting for language acquisition.

A second difference from analogical learning is proposed by Jackendoff 2002 (chapter 6). Just creating more general structures in memory is probably not

[12]One can imagine possible refinements that make this account somewhat more "brain-friendly", for instance decreasing the threshold of activation of a rule-like item as it is invoked more frequently. Perhaps a rule "goes productive" when its activation threshold is lower than the particular items from which it is abstracted.

enough to guarantee the emergence of the language in the learner. A counterpart of the Poverty of the Stimulus argument applies here: how is the learner (or the learner's brain) to know which of the many possible generalizations are the right ones? Traditionally, the role of Universal Grammar is to guide the learner's generalizations.

The present formalization offers an attractive way to capture this intuition: UG can be stated as a set of very general schemata – within the same data structure format as the words and rules. These schemata serve as endpoints on the inheritance hierarchy, such that an intermediate-level generalization (i.e. a language-specific rule) that falls under one of them is easier to learn, or is more highly valued, or is easier to activate, than a generalization that does not. The result would be a tendency for rules to fall under these schemata and hence to conform to a more or less universal prototype, while competing generalizations would be extinguished. On the other hand, if the input data of the language force generalizations that fall outside the prototype, the result is a syntactic nut like N-P-N. In a sense, then, UG is a set of "attractors" on rules, toward which the learner will be drawn in creating a grammar.

Returning to the strategies for satisfying the Learnability Constraint of section 1: this learning procedure strikes a compromise between two ways to Minimize Learning. On the one hand, since words and rules are stated in a common format, there is only one kind of thing to learn rather than two. On the other hand, the common format makes it possible to learn a large class of rules, of all degrees of regularity, so there is no need to minimize the structure of particular grammars and the differences among languages (unlike P&P, where the full range of differences is built into UG). The learner still needs UG to guide learning, however, so it is not the case that language acquisition comes entirely "for free."

Here is a best guess for what might fall into UG, on this conception (following Jackendoff 2002 and Culicover and Jackendoff 2005).

- The basic organization of conceptual structure (meaning). Although there are undoubtedly aspects of meaning that depend on having language, many fundamental elements grow directly out of primate cognition, for instance individuation, the type-token distinction, the organization of space, and important aspects of social interaction. They therefore belong in the Broad Faculty of Language and come "for free."

- The notion of words being used symbolically to communicate intentionally about the perceived world. The jury is still out on the extent to which this notion is present in language-trained apes, but it is certainly present robustly in babies. I think Deacon (1997) is essentially correct in seeing this as *the* evolutionary breakthrough in the emergence of language

in our species. The rest is refinement.

- The use of variables in stored structures and the use of unification to combine them, permitting productivity and recursion. As this occurs also in conceptual structure (and likely elsewhere, e.g. the formulation of complex actions (Jackendoff 2006)), it constitutes part of the Broad Faculty of Language.

- Basic principles of phrase structure, including
  a. Distinctions between lexical and phrasal categories, and possibly a distinction between verbs and all other lexical categories
  b. X-bar theory (30e), through which each lexical category projects a phrasal category with the same label
  c. Other common alternatives such as the conjunction schema

  The existence of categories per se is not specific to language, but this particular set of categories looks like part of the Narrow Faculty of Language. In particular, as is well known, it does not align completely with anything in semantics.

- Basic default principles of the syntax-semantics interface
  a. Semantic heads defeasibly map to syntactic heads, semantic arguments to syntactic arguments, and semantic modifiers to syntactic adjuncts. This is the present theory's counterpart Interface Uniformity in mainstream theory; however it is violable, for instance in constructional idioms such as those in (27).
  b. Agent First (more generally, principles for relating thematic roles to linear order)
  c. Topic First, Focus Last (more generally, principles for relating information structure to linear order)

  These principles, although they are grounded in more general cognitive principles of alignment, pertain to specifics of language structure, so they are by necessity aspects of the Narrow Faculty of Language.

- Basic principles of morphological agreement and case-marking. To the extent that such principles depend on distinguishing a syntactic category of verbs, which undergo agreement and determine case-marking, and a category of nouns, which impose agreement and undergo case-marking, they are specific to language.

- Basic principles of long-distance dependencies. Probably largely specific to language.

- Culicover and Jackendoff 2005 show that the principles above are not sufficient to characterize the mapping between syntax and semantics in

English. We propose an additional layer of structure, the *Grammatical Function Tier*, which deals with the linearization of just the NP arguments of verbs (subjects, objects, and indirect objects), disregarding the verb's PP arguments and adjuncts as well as the arguments of nouns, adjectives, and prepositions. The Grammatical Function Tier is involved in the formulation of passive, raising, structural case, agreement, pro-drop, and intra-clausal binding. Lest this tier seem an excessive addition to the theory, we observe that it has clear counterparts in every sufficiently worked out theory of syntax: functional structure in LFG, essentially all the structure studied in Relational Grammar, the complement hierarchy in LFG, and Abstract Case in P&P/MP. Given that this tier lies deep in the internal workings of language, it is hard to see what could constitute a parallel in any other cognitive capacity. We therefore feel confident is assigning it to the Narrow Faculty of Language.

- In addition, there are UG aspects of phonology and morphology. At the very least, the child has to come equipped to create a large repertoire of symbols by *digitizing* the sound in the input, through decomposing it into syllables and segments, the latter forming a structured space of feature oppositions. Such organization is not known in any other animal communication system, and thus it constitutes part of the Narrow Faculty of Language. And surely there is more to UG for phonology and morphology than this.

This heterogeneous class of elements is not "perfect" by any means. But it appears *relatively* minimal, given the need to satisfy the Descriptive Constraint. Unlike the Minimalist Program, this conception of grammar allows for a proliferation of learned rules, under a potentially realistic learning regimen. In particular, learning rules is mostly an extension of learning words. The Narrow Language Faculty is to be seen as a "toolkit" that results in tendencies toward language universals.

Unlike P&P and the Minimalist Program, this vision of language is nicely compatible with an incremental evolution of the language faculty (see Jackendoff 2002). In particular, it is possible to imagine a "protolanguage" in Bickerton's (1990) sense, which has words, organized phonologically and used symbolically, and in which linearization is a product not of phrase structure, but only of simple direct principles such as Agent First and Focus Last, statable over linear phonological structure. Such a communication system would consist of an interface between semantics and phonology, without syntax in between. It would be eminently useful for communication, if not as flexible and precise as modern language. In fact, it is not far off from a characterization of pidgins (Bickerton 1990, Givon 1995) and the speech of poor second language learners (Klein and Perdue 1997). The evolution of syntax could then be seen as the emergence of

new sorts of constraints between semantics and phonology that yield adaptive advantages in precision, efficiency, and flexibility of communication. The additions are somewhat heterogeneous (in particular dividing into morphological and phrasal flavors). However, some are logically dependent on others – for instance verb agreement and case depend on there being a noun-verb distinction – which offers some clues as to order of evolution.

This is not to say that the constraint-based frameworks have solved all the problems of language description, language acquisition, or language evolution. Indeed, the foregoing constitutes at best a prolegomena to a large-scale research program. What I hope to have shown, though, is that there is a robust alternative to mainstream generative grammar when it comes to attacking the goal of explaining language at the deepest possible level. And my prediction is that if any approach to language is eventually going to achieve this goal, at the same time making meaningful contact with cognitive neuroscience and evolutionary biology/psychology, it will be an approach growing out of constraint- and construction-based minimalism, not out of the Minimalist Program.

References

Akmajian, Adrian. 1984. Sentence types and the form-function fit. *Natural Language and Linguistic Theory* 2: 1-23.

Baayen, Harald, Robert Schreuder, Nivja de Jong, and Andrea Krott. 2002. Dutch inflection: The rules that prove the exception. In S. Nooteboom, F. Weerman, and F. Wijnen (eds.), *Storage and Computation in the Language Faculty*, 61-92. Dordrecht: Kluwer.

Bickerton, Derek. 1990. *Language and Species*. Chicago: University of Chicago Press.

Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart & Winston.

Braine, Martin. 1971. On two types of models of the internalization of grammars. In D. I. Slobin (ed.), *The Ontogenesis of Grammar: A Theoretical Symposium*. New York: Academic Press.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge, MA: MIT Press.

Chomsky, Noam. 1966. *Cartesian Linguistics*. New York: Harper & Row.

Chomsky, Noam. 1977. On wh-movement. In P. W. Culicover, T. Wasow, and A. Akmajian (eds.), *Formal Syntax*, 71-132. New York: Academic Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

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

Chomsky, Noam. 2001. Beyond explanatory adequacy. *MIT Working Papers in Linguistics* 20.

Chomsky, Noam. 2002. *On Nature and Language*. Cambridge: Cambridge University Press.

Clark, Eve. 2002. *First Language Acquisition*. Cambridge: Cambridge University Press.

Csuri, Piroska. 1996. Generalized dependencies: Description, reference, and anaphora. Doctoral dissertation, Brandeis University.

Culicover, Peter W. 1967. The treatment of idioms within a transformational framework, IBM Boston Programming Center Technical Report, Cambridge, MA.

Culicover, Peter W. 1972. OM-sentences. *Foundations of Language* 8: 199-236.

Culicover, Peter W. 1999. *Syntactic Nuts: Hard Cases in Syntax*. Oxford: Oxford University Press.

Culicover, Peter W. and Ray Jackendoff. 1999. The view from the periphery: The English comparative correlative. *Linguistic Inquiry* 30: 543-571.

Culicover, Peter W. and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.

Culicover, Peter W. and Andrzej Nowak. 2003. *Dynamical Grammar*. Oxford: Oxford University Press.

Erteschik, Nomi. 1973. On the nature of island constraints. Doctoral dissertation, MIT.

Erteschik-Shir, Nomi. 1979. Discourse constraints on dative movement. In T. Givon (ed.), *Discourse and Syntax, 441-467*. New York: Academic Press.

Fillmore, Charles, Paul Kay, and Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64: 501-539.

Fodor, Jerry A. 1975. *The Language of Thought*. Cambridge, MA: Harvard University Press.

Ghomeshi, Jila, Ray Jackendoff, Nicole Rosen, and Kevin Russell. 2004. English focus reduplication: The salad-salad paper. *Natural Language and Linguistic Theory* 22: 307-357.

Ginzburg, Jonathan, and Ivan A. Sag. 2000. *Interrogative Investigations*. Stanford, CA: CSLI.

Givon, Talmy. 1995. *Functionalism and Grammar*. Philadelphia: John Benjamins.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele. 2005. *Constructions at Work: Constructionist Approaches in Context*. New York: Oxford University Press.

Goldberg, Adele and Ray Jackendoff. 2004. The English resultative as a family of constructions. *Language* 80: 532-568.

Haegeman, Liliane. 1992. *Introduction to the Theory of Government and Binding*. Oxford: Blackwell.

Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how does it evolve? *Science* 298: 1569-1579.

Horn, Lawrence. 1993. Economy and redundancy in a dualistic model of natural language. In S. Shore and M. Vilkuna (eds.) SKY 1993, *Yearbook of the Linguistic Association of Finland*, 31-72.

Hornstein, Norbert and Amy Weinberg. 1981. On preposition stranding. *Linguistic Inquiry* 11: 55-91.

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51: 639-671.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1997a. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1997b. Twistin' the night away. *Language* 73: 534-559.

Jackendoff, Ray. 2002. *Foundations of Language*. Oxford: Oxford University Press.

Jackendoff, Ray. to appear. NPN: Another family of constructions in English.

Jackendoff, Ray and Steven Pinker. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition* 97: 211-225.

Katz, Jerrold and Paul M. Postal. 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, MA: MIT Press.

Kayne, Richard S. 1983. Connectedness. *Linguistic Inquiry* 24: 223-249.

Klein, Wolfgang and Clive Perdue. 1997. The Basic Variety, or: Couldn't natural language be much simpler? *Second Language Research* 13: 301-347.

Kluender, Robert. 1992. Deriving island constraints from principles of predication. In H. Goodluck and M. Rochemont (eds.), *Island Constraints: Theory, Acquisition and Processing*, 223-258. Dordrecht: Kluwer.

Koch, Christof. 2004. *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts and Company.

Kuno, Susumu. 1987. *Functional Syntax: Anaphora, Discourse and Empathy*. Chicago: University of Chicago Press.

Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19: 335-391.

Lasnik, Howard. 2002. The Minimalist Program in syntax. *Trends in Cognitive Sciences* 6: 432-437.

Lerdahl, Fred and Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Lidz, Jeffrey. 2001. Echo reduplication in Kannada and the theory of word-formation. *The Linguistic Review* 18: 375-394.

Marcus, Gary. 1998. Rethinking eliminative connectionism. *Cognitive Psychology* 37: 243-282.

Marcus, Gary. 2001. *The Algebraic Mind*. Cambridge, MA: MIT Press.

McCarthy, John (ed.). 1994. *Optimality Theory in Phonology: A Reader*. Oxford: Blackwell.

McCawley, James D. 1988. The comparative conditional construction in English, German, and Chinese. *Proceedings of the 14th Annual Meeting of the Berkeley Linguistics Society*, 176-187. Berkeley, CA: Berkeley Linguistic Society.

Millikan, Ruth. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Pinker, Steven and Ray Jackendoff. 2005. The faculty of language: What's special about it? *Cognition* 95: 201-236.

Pollard, Carl and Ivan A. Sag. 1987. *Information-Based Syntax and Semantics*. Stanford, CA: CSLI.

Pollard, Carl and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Postal, Paul M. 2004. *Skeptical Linguistic Essays*. Oxford: Oxford University Press.

Povinelli, Daniel. 2000. *Folk Physics for Apes*. Oxford: Oxford University Press.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Piscataway, NJ: Rutgers University Center for Cognitive Science.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33: 431-484.

Shieber, Stuart. 1986. *An Introduction to Unification-Based Approaches to Grammar*. Stanford, CA: CSLI.

Tomasello, Michael. 2003. *Constructing a Language*. Cambridge, MA: Harvard University Press.

Tomasello, Michael, Malinda Carenter, Jesp Call, Tanya Behne, and Henrike Moll. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28: 675-691.

Van Valin, Robert D. 1998. The acquisition of wh-questions and the mechanisms of language acquisition. In M. Tomasello (ed.), *The New Psychology of Language*, 221-249. Hillsdale, NJ: Erlbaum.

Wertheimer, Max. 1923. Untersuchungen zur Lehre von der Gestalt (Laws of organization in perceptual forms). *Psychologische Forschung* 4: 301-350. Translation in W. Ellis (ed.), *A Source Book of Gestalt Psychology*. London: Routledge & Kegan Paul (1938), 71-88.

Wexler, Kenneth and Peter W. Culicover. 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.

Williams, Edwin. 1994. Remarks on lexical knowledge. In L. Gleitman and B. Landau (eds.), *The Acquisition of the Lexicon*, 7-34. Cambridge, MA: MIT Press.