

Linguistics in Cognitive Science: The state of the art

RAY JACKENDOFF

Abstract

The special issue of The Linguistic Review on “The Role of Linguistics in Cognitive Science” presents a variety of viewpoints that complement or contrast with the perspective offered in Foundations of Language (Jackendoff 2002a). The present article is a response to the special issue. It discusses what it would mean to integrate linguistics into cognitive science, then shows how the parallel architecture proposed in Foundations seeks to accomplish this goal by altering certain fundamental assumptions of generative grammar. It defends this approach against criticisms both from mainstream generative grammar and from a variety of broader attacks on the generative enterprise, and it reflects on the nature of Universal Grammar. It then shows how the parallel architecture applies directly to processing and defends this construal against various critiques. Finally, it contrasts views in the special issue with that of Foundations with respect to what is unique about language among cognitive capacities, and it conjectures about the course of the evolution of the language faculty.

I am honored that *The Linguistic Review* has suggested *Foundations of Language* (Jackendoff 2002a, henceforth FL) as a “unifying starting point” for discussion of the role of linguistics in cognitive science. The fruits of this discussion, a special issue of the journal (Ritter 2005b, henceforth TLR), represent a fascinating microcosm of the state of the art, offering a profusion of fields and points of view.

This response to the special issue has two goals. First, I wish to reiterate the positions taken in FL, defend them against the critiques in some of the papers, and show how they bear on or even resolve issues raised by some of the others.¹

1. Several authors issue blasts against generative grammar, characterizing it purely in terms of

Second, I want to imagine what some authors' responses to each other might be. Many of the papers provide perfect illustrations of problems with other papers, or emphasize an important aspect of language that other papers neglect. In other cases – the more interesting – one paper can build on others to suggest a larger synthesis. My overall goal is to promote further conversation among disciplines, bearing in mind that “everyone will have to endure some discomfort and give a little” (FL: xiii).

Section 1 deals with general theoretical and methodological issues: what it means to integrate linguistics into cognitive science. Section 2 recapitulates the parallel architecture, the fundamental theoretical innovation in FL; sections 3 and 4 answer critiques from mainstream generative grammar and from opponents of mainstream generative grammar respectively. Sections 5 and 6 discuss issues of language acquisition and the continuing need for Universal Grammar, including the “layered” character of UG. Section 7 discusses the relation of linguistic theory to theories of processing and memory. This lays the groundwork for a more thorough discussion in section 8 of what is special about the language capacity and what follows from more general properties of human cognition. Finally, Section 9 turns to the question of the evolutionary roots of the language capacity and the traces they have left in the structure of the modern brain and modern language; to my astonishment, it is here above all that the papers in this special issue may lead to new insight.

1. What does it mean to unify linguistics with cognitive science?

1.1. *Four levels of inquiry about language*

Linguistics perforce became a cognitive science when linguists took it as a central premise that the nature of language depends on its instantiation in the minds of speakers. This is one of the founding perspectives of generative grammar (FL: Chapter 2), and it is rightly identified above all with Noam Chomsky. To unify linguistics with cognitive science is to take seriously all the implications of this premise for the nature of language. Conversely, to unify cognitive science with linguistics is to take seriously all the insights about the mind/brain afforded by inquiry into the nature of language.

On this view, language is a system of knowledge² stored in the memory of

Chomsky's position (or a caricature thereof); they overlook the fact that FL anticipates many aspects of their critiques, acknowledging some and answering others. I really think that a major obstacle to unification is this insistence on viewing Chomsky's linguistics as the only linguistics worth discussing, which one would think we could get over by now.

2. FL introduces the term “functional knowledge” or “f-knowledge” here, to distinguish f-knowledge of language from conscious knowledge of facts and from the philosophers' notion

a speaker. Any particular language, e.g., English, is an abstraction rather like a species: it is a convenient idealization over a community of speakers for whom the languages in their heads are similar enough to be mutually fluently intelligible. Linguists committed to cognitive science claim that the formal structures they propose have a “psychological reality,” that is, they correspond to something actually present in the minds of speakers. Much dispute in the literature concerns how linguists as well as those outside of linguistics proper should understand such a claim.

The strongest interpretation of the claim would be that those aspects of language that are part of a speaker’s knowledge are present in memory, and those aspects of language that are the result of combinatorial rule application (e.g., complex novel sentences) are constructed online in speakers’ minds in the course of production and perception. Ideally, these pieces of knowledge and processes of combination can be studied experimentally, in particular probing the differentiation among aspects of speakers’ knowledge (and whether this conforms to theoretical predictions), as well as the time-course and sources of complexity in combinatorial processing.

Moreover, since the mind is instantiated by the nervous system, a full unification would show how knowledge of language and the combinatorial processing of language are achieved by the brain. This would have to proceed at two levels: the broad localization of various aspects of language and language processing – a prospect currently within reach (e.g., Baynes and Gazzaniga 2005; Poeppel and Embick 2005) – and a detailed account of how neural assemblies actually accomplish the functions they fulfill – a considerably more distant prospect.

These three levels of inquiry – the formal, the functional/psychological,³ and the neuroscientific – are replicated in the study of language acquisition. At the formal level, one can ask what lexical items and rules of grammar the child has acquired at any particular stage, that is, the state of the child’s knowledge of language and how it changes over childhood. One can also ask what knowledge the child must have in order to be able to acquire language – the well-known but frequently misunderstood Universal Grammar of generative theory.⁴ At the functional/psychological level, one can frame the inquiry in terms of the gradual building of linguistic aspects of the child’s memory and the gradual refinement of the child’s language processing system – we must remember above

of “justified true belief,” neither of which is appropriate for characterizing “knowing English” or “knowing Navajo.”

3. The distinction I make here between formal and functional is different from the usage of some subcommunities in the field, represented in the special issue of TLR by Tomasello, Bybee/McClelland, and Goldberg/Del Giudice. They use “formal” for what I mean by “syntactic”; they use “functional” for what I mean by “semantic/conceptual” (which includes pragmatics), as well as for the effects of relative complexity in processing and acquisition.

4. See FL: Section 4.2, for a host of persistent misinterpretations.

all that the child acquires language through using it (Culicover (2005: 228)). At the neural level, one can ask what changes take place in the child's brain in the course of language acquisition, and what it is about the substrate of the human brain that potentiates these changes – as it were, the neural correlate of Universal Grammar. A similar three-level story can be constructed for second (or late) language acquisition and for language impairments.

The process of language change (not discussed in FL but raised in TLR by Tomasello and Bybee/McClelland) is somewhat more complex: it is necessary to treat the range of variation in a speech community, the potential sources of innovation, the pressures within a grammar that may lead to change, and how such pressures are reflected in acquisition. At the formal level, this amounts to the standard study of language variation and language change. At the psychological/functional level, the pressures for grammar change might be reflected in processes that reduce processing complexity or optimize the organization of memory; at the neural level one could ask how measures of functional complexity and optimality result from the way neurons instantiate linguistic function.

There is also a fourth level of inquiry. The structure of human brains arises from the interaction of the genetic code with processes of biological development. These processes require input in the form of nutrition and in the form of information (or differential stimulation); development is also constrained by physical principles. The cognitive differences between humans and our primate relatives are a consequence of differences in brain structure that have developed over the past 6 million years or so. The language capacity, being one of these differences, thus belongs to the subject matter of biology and evolutionary psychology.

1.2. The treatment of the competence-performance distinction

As stressed by Wilkins (2005: 272) and Walenski and Ullman (2005: 328), a full unification of linguistics with cognitive science would demonstrate the rich interpenetration of all these levels of inquiry, from formal through psychological/functional to neural and biological/evolutionary. What I am calling the formal level corresponds roughly to what Marr 1982 calls the “computational” level; it is a characterization of what function the brain computes. My functional/psychological level corresponds to Marr’s “algorithmic” level, a characterization of the actual processes the brain uses to compute this function. My neural level corresponds to Marr’s “implementational” level, a characterization of how the neurons carry out these processes.⁵ The linguist’s distinction

5. Ritter (2005a, 119) sees a different alignment of Marr’s levels, identifying the computational

between *competence* and *performance* corresponds, I think, to the formal level versus everything else (FL: Section 2.4).⁶

Marr discusses his three levels of inquiry in the context of admonishing psychologists and neuroscientists for unduly neglecting the computational level – for undertaking experimentation without first asking in detail what the system has to accomplish and how the formal/mathematical constraints on the task constrain possible processing solutions. Marr’s complaint applies equally to many nonlinguists’ characterization of language. For this reason FL opens (Chapter 1) with a sketch of all the formal structure – phonological, syntactic, and semantic – of a very simple sentence of English, setting this as a lower bound on the formal complexity that any unified theory of language has to account for.

Linguists, on the other hand, have emphasized the formal level at the expense of the rest. Chomsky (1965) originally introduced the competence/performance distinction as a principled justification of doing what linguists normally do: investigating word order, case marking, and so on in a variety of languages, without having to be tied to actual behavioral data such as speech errors, coughs, and speakers losing track of what they are saying (FL: 29–30). FL (33) notes that the distinction has gradually hardened over the years: as experimental research led to increasing doubts about the rules, representations, and derivations proposed by formal linguists, linguists often responded that issues of processing are irrelevant to competence theory, while simultaneously insisting on the relevance of competence theory to processing. As FL, Walenski and Ullman (2005: 328–329), Ferreira (2005: 367–370) and Marantz (2005: 431) observe, such an attitude is an obstacle to unification, in that it accords formal theory an unwarranted supremacy over experimental evidence and renders formal theory impervious to experimental disconfirmation. One consequence has been alienation of psycholinguists from formal linguistics (Myachykov, Tomlin, and Posner 2005: 350; Ferreira 2005: 368–370; FL: 74). Poeppel and Embick 2005 go so far as to suggest that attempts to integrate linguistic theory with brain science have led to “interdisciplinary cross-sterilization”.

Several contributors (MacNeilage and Davis; Culicover; Lieberman; Ferreira; and Myachykov, Tomlin, and Posner) observe and bemoan this disconnection. Marantz (2005: 432–438) attributes it to a “public relations problem”, as did I in Jackendoff 1988: if only linguists would make clearer what they

level with semantics (conceptual structure in FL), the algorithmic level with the symbolic formal level, i.e., the “computational system” of syntax that Chomsky takes as the foundation of the language capacity, and the implementational level with both processing and neural instantiation. I question this characterization, because Marr carefully investigates the “computational” character of very “unsemantic” stages of vision such as the primal sketch.

6. It does not parallel Saussure’s *langue/parole* distinction, as asserted by MacNeilage and Davis (2005: 163); see FL (29, N. 6).

really mean, everyone would recognize that we all have the same interests at heart. I now think I was unduly optimistic. Many prominent linguists (who shall remain nameless) don't want to be bothered about the psychological, neural, and biological ramifications of language, and many others give it only lip service – and they pass these attitudes on to their students.

An interesting case of disconnection that goes beyond “public relations” is the TLR paper by Boeckx and Piattelli-Palmerini, which, echoing Chomsky's recent writings (e.g., Chomsky 2000), speaks of unifying linguistics with biology and physics, deriving properties of language from deep and simple principles, and thereby going “beyond explanatory adequacy.” In their eagerness to unify the field, however, Boeckx and Piattelli-Palmerini entirely bypass the psychological/functional and neural levels of inquiry, which are the necessary steps between formal linguistics and biology. This is justified in terms of “a recurrent and deep-seated urge characteristic of the mature sciences” (454) to pursue maximal generality at the highest level of idealization, characterized as “Galilean method.” How one follows this urge, of course, is a matter of where to place one's bets – as they acknowledge. My bet, unlike theirs, is that unification of the science of language with everything else must include unification with the rest of psychology and neuroscience, and that only in the context of understanding how behavioral repertoires in general are constrained by genetics can we make connections between language and biology. I believe that the other contributors agree on this point, at least implicitly.

1.3. *Problems with levels of representation.*

Another interesting outlier among the contributors is Bybee and McClelland, who emphatically assert (397–398) that ultimately the formal level of inquiry will wither away (their italics):

[T]here is no analysis into units on any level or set of levels [e.g., phonological, morphological, or syntactic – RJ] that will ever successfully and completely capture the realities of synchronic structure or provide a framework in which to capture language change.

Bybee and McClelland acknowledge that units such as *adjective*, *subject*, *accusative case*, and *syllable* are unavoidable descriptive conveniences for linguists, but they suggest that these will eventually be discarded in favor of emergent properties of more adequate functionally-based models of neural behavior growing out of the connectionist tradition. Emphasizing that much less of language is freely combinatorial than generative grammar claims, they reject the necessity for combinatorial mechanisms, ignoring one of the original great lessons of formal linguistics (Chomsky 1957; Miller and Chomsky 1963). In

doing so, they miss Marr's point that a functional or neural model must be able to compute the function that the formal model requires – which in the case of language includes free combinatoriality as well as many larger fixed phrases. I will return to Bybee and McClelland's approach in Section 4.

A third and quite different outlier is Gross's paper on the proper approach to semantics. Coming out of the philosophical tradition, Gross is above all concerned that mental representations be connected properly to the real world, for example that "the presence of a cow can activate my cow-concept" (2005: 252, N. 6). But he never raises the issue of the mechanism by which this can happen. My cow-concept is not activated simply by the *presence* of a cow: I have to have *perceived* the cow. Cow-concepts, if they're in the brain, are instantiated in neurons that have no direct connection to the world outside the brain. Rather, the relation of my cow-concepts to cows has to be mediated by the perceptual system, for which the explanatory burden falls on psychology and neuroscience (FL: Chapter 10). Gross, by contrast, takes the causal relation of the cow to the cow-concept for granted, as though it's simply self-evident.⁷ The same goes for other philosophers who have discussed FL's conceptualist approach to semantics, such as Rey 2006 and Higginbotham 2003; similar problems accrue to Jerry Fodor's work appealing to intentionality (e.g., 1987, 1990, 2000; see FL: Section 9.4; Jackendoff 1992: Chapter 8; Jackendoff 2002b).

Gross opposes conceptualist semantics with a position he calls "methodological naturalism." A methodological naturalist can accept intentionality (a direct relation between concepts and the world), as long as one "insists . . . that one reject theories of, or adverting to, intentionality that are explanatorily vacuous, unintelligible, insufficiently fruitful, etc" (2005: 259). Yet he quotes (260, n. 25) a passage in FL (279) that advocates this very position: even if there is such a thing as intentionality, we still have to explain how the mind grasps meaning. In fact, the four chapters of discussion in FL about ontology, lexical decomposition, and phrasal composition demonstrate how an insistence on intentionality not only is "explanatorily vacuous" and "insufficiently fruitful", it actually gets in the way of insightful semantic/conceptual analysis. Thus I see Gross's position as actually warding off unification of semantics with the rest of cognitive science.

7. Though he adds in a note (2005: 260), "I see no reason to preclude an expression or concept's being about something in virtue of its standing in a complex mediated causal relation to that thing." If he accepts the perceptual system as instantiating such a relation, then we're on the same page. But he certainly doesn't sound to me like he is thinking along such lines.

1.4. Methodological concerns

One of the divides between linguistics and the other cognitive sciences is methodological. It has always been standard in linguistics to use grammaticality judgments on made-up examples, as well as attestations from actual speech and text, but this has been suspect in other quarters. (Some colleagues at my former university refused to consider linguistics a part of psychology or even of science, because linguists don't do experiments.) In TLR, Myachykov, Tomlin, and Posner (2005: 349) and Ferreira (2005: 371) criticize this methodology,⁸ and Marantz (2005: 432–438) offers a defense, paralleling (without citation) the discussion in Jackendoff 1994 (46–50). Briefly, my sense (and Marantz's, I think) is that grammaticality judgments are akin to well-known visual demonstrations such as the Necker cube, the duck-rabbit, the Kanizsa triangle, Escher's 3-dimensionally anomalous drawings, and Julesz's random-dot stereograms – all of them made-up examples. They are quick and dirty experiments for which the results are (purportedly) robust enough as to require no statistical analysis.

However, as Myachykov, Tomlin, and Posner and Ferreira observe, the methodology is easily abused. Linguists frequently do not offer enough control examples to sort out all the factors involved in an ambiguity or ungrammaticality. Major theoretical points often rest on assertions of delicate judgments that prove not to be uniform among speakers and that may well be biased by the writer's theoretical predispositions or by the writer's overexposure to too many examples. (In 1973, I gave up working on reflexives because after five years of playing with picture-noun reflexives my judgments were shot.) The solution is not to abandon grammaticality judgments altogether. It would cripple linguistic investigation to require that all judgments of ambiguity and grammaticality be subject to statistically rigorous experiments on naive subjects – especially when working on languages where speakers are hard to come by. Rather, as in all scientific inquiry, the solution is to be as careful as possible with grammaticality judgments, to control for all possible relevant factors, and by all means not to consider such judgments privileged over other sorts of data other than by virtue of their convenience.

8. Oddly, Myachykov, Tomlin, and Posner speak favorably of work in Cognitive Linguistics by Givón, Chafe, Langacker, and Talmy, all of whose work is (to my knowledge) based on intuitive judgments and not on experimental evidence. One might therefore suspect that the animus is really more toward mainstream generative grammar per se than toward its methodology.

1.5. *Connecting language to the rest of cognition*

The unification of linguistics with cognitive science does not pertain just to linguistics and psycho/neuro-linguistics. There must be unification with the study of all faculties of the mind/brain, at all levels of inquiry. FL (79) quotes Chomsky (1965: 207):

Notice that we do not, of course, imply that the functions of language acquisition are carried out by entirely separate components of the abstract mind or the physical brain.... In fact, it is an important problem for psychology to determine to what extent other aspects of cognition share properties of language acquisition and language use, and to attempt, in this way, to develop a richer and more comprehensive theory of mind.

At the formal level, this calls for an account of the mental structures (or representations) involved in other cognitive capacities such as vision (as in Marr 1982), music (as in Lerdahl and Jackendoff 1983⁹), the formulation of action (Jackendoff 2007b: Chapter 4 and references therein), and social cognition (Seyfarth and Cheney 2003, Jackendoff 2007b: Chapters 5–11). The important unifying issue is what aspects of linguistic representations are shared by other capacities: hierarchical structure, unbounded recursion, instantiation of variables, constituents with heads, and so on (Jackendoff 1987: Chapter 12; FL: Section 4.5). What seems altogether well established is that, contrary to the speculation of Hauser, Chomsky, and Fitch 2002, hierarchical structure and recursion are ubiquitous in higher human cognitive capacities and perhaps not uncommon in other animals as well (Gallistel 1990). Language is unique not for its recursion per se, but through being the only natural communication system whose signals have hierarchical structure that to some degree mirrors hierarchical structure in the messages being transmitted. This is still nothing to sneeze at, of course (Pinker and Jackendoff 2005; Culicover and Jackendoff 2005).

At the psychological/functional level, the relevant questions are to what extent the character of linguistic memory and language processing parallels that of other capacities of comparable complexity, such as understanding rich visual fields, planning and carrying out multistep actions, and, as Culicover (2005) points out, improvising jazz. To the extent that language can be shown to make

9. Tomasello (2005: 188) argues against Universal Grammar on the grounds that “no one to date has proposed anything like Universal Music”, which is exactly what Lerdahl and Jackendoff do (see especially chapter 11 and now Jackendoff and Lerdahl 2006). To be sure, it is not stated in terms of parameters like the versions of Universal Grammar Tomasello is criticizing, but then neither is the UG of FL (see Section 5).

use of domain-general mechanisms of memory access, learning, attention, and assembly of complex structures at multiple levels of representation, it becomes possible to understand how much of language is cut from the same cloth as the rest of the mind. Several contributors to TLR raise such commonalities as evidence against the hypothesis of UG; I will take up these issues in more detail in Section 8. For the moment, let me only say that showing there is *some* commonality between language and other capacities hardly demonstrates that the commonalities *completely* account for the language capacity. (This problem of substituting a universal for an existential quantifier sadly turns out to be a frequent logical slip in the critiques.)

At the neural level, familiar questions arise: is there anything special about the language areas of the brain? Are they devoted exclusively to language, or are they shared with other capacities such as control of the hands (Wilkins 2005) or general-purpose temporal hierarchical structuring (Lieberman 2005)? Is language “riding on top of” more primitive capacities such as the motor program responsible for chewing (MacNeilage and Davis 2005)? And so on.

Another important unifying issue is how language interfaces with other capacities, so that we can use our auditory systems to hear language, our motor systems to speak, and above all use language to express our perceptions and thoughts. The formal problem is alluded to by Boeckx and Piattelli-Palmerini, in the context of the Minimalist Program’s desire to derive as much of language as possible from “the specifications of the interfaces in our species” (2005: 461). However, as in everything Chomsky has written about the Minimalist Program, there is no exploration of what language interfaces *with*. In particular, what the Minimalist Program calls the “sensorimotor” interface is actually *two* independent interfaces: one with audition and one with motor control, both of which are rather well understood by now but which play absolutely no role in constraining the Minimalist Program. By contrast, FL (following a more detailed exposition in Jackendoff 1997), takes pains to explore the character of the language faculty’s interfaces with other systems – particularly including the interface between linguistic meaning and the visual system, a crucial unifying step in the formal theory of cognition (in my own work, see also Landau and Jackendoff 1993; Jackendoff 1996a; plus such works on language and space as Bloom et al. 1996 and Levinson 2003). A related issue is the extent to which language (the grammar and lexicon of particular languages, or the presence of language in general) can affect the content and course of reasoning (Slobin 1996; Levinson 2003; Li and Gleitman 2002; Boroditsky 2000; Jackendoff 1996b).

At the psychological/functional level, interface issues include the timing of access to the visual field and to world knowledge in the course of sentence parsing (e.g., Tanenhaus et al. 1995; Hagoort 2005). At the neural level, a corresponding question is how the parts of the brain responsible for language are

connected to other parts of the brain involved in perception, the formulation of action, and world knowledge, a central issue for Wilkins (2005).

In short, there are many dimensions in which linguistics can be unified with cognitive science. Some of these have been well explored, but some still present major challenges.

2. The parallel architecture

We now turn to the central innovation of FL, the parallel architecture. This architecture emerges from a reconsideration of deeply held founding assumptions of mainstream generative grammar that have persisted for nearly fifty years. FL (Chapter 5) observes that these assumptions have never been argued for; they have simply become an unspoken background to theoretical practice. As shown in FL and Culicover and Jackendoff 2005 (especially Chapters 2 and 3), many of the features of contemporary mainstream generative theory that contribute to alienation from the rest of cognitive science result directly from maintaining these founding assumptions in the face of massive counterevidence. FL and Culicover and Jackendoff 2005 further show that an alternative set of assumptions is better in tune with the basic facts of language and also lends itself to more direct integration with the rest of cognitive science. There is no space to go into the arguments here; suffice it to say that they offer sufficient reason to consider an alternative conception of generative grammar.

Four basic assumptions of mainstream generative theory are replaced:

- The basic “generative engine” of language, which provides it with its unlimited combinatorial richness, is the syntactic component (now called the “computational system”). The relation between complexes of sound and complexes of meaning is created by “interpreting” syntax into phonological and semantic structure. FL calls this the assumption of *syntactocentricity*.
- Generativity is formalized in terms of algorithm-like derivations, which build and alter trees step by step in a certain logical order.
- Because semantics is “interpretive”, all semantic combinatoriality must be derived through homomorphism with combinatoriality in syntactic structure. When such syntactic combinatoriality is not apparent in the surface form of a sentence, it is taken to be explicit in a covert syntactic form (Deep Structure in early work, later Logical Form) that is related to the surface form by a series of deformations (originally transformations). Culicover and Jackendoff 2005 call this assumption *Interface Uniformity*. It is closely related to Frege’s notion of compositionality in logic: “The meaning of a compound expression is a function of the meaning of its parts and of the syntactic rules by which they are combined” (Partee, ter Meulen, and Wall 1993). Marantz (2005: 440) states it as “the meanings of linguistic expressions are

built up recursively from the meanings of the constituents of the expressions, where a recursive definition of ‘constituent’ decomposes phrases eventually to atomic sound/meaning connections.”

- An even older assumption, tracing back to traditional grammar (and also present in formal logic), is that the rules of grammar and the lexicon are entirely distinct formal entities, and that the productivity of language is specifically a function of the rules. Mainstream generative grammar incorporates this assumption by ensconcing lexical items (complete with their phonology and semantics) as the terminal elements of syntactic trees, where they are carried around passively in the course of a grammatical derivation.

The replacements for these four assumptions are:

- The grammar contains multiple independent sources of generativity: phonology (and within it, tiers such as segmental/syllabic structure, stress grid, and intonation), syntax, and semantics (which also subdivides into tiers such as propositional structure and information structure). The correlations among these structures are established not by derivations but by *interface rules*, which relate substructures generated in different components. A phrase or sentence is well-formed if it has well-formed structures in all components, related in well-formed fashion by all relevant interface rules. The theory is named the *parallel architecture* on the basis of this characteristic. (A similar assumption, though not carried through to phonology, appears in LFG, Role and Reference Grammar, and Autolexical Syntax.)
- Well-formedness of structures is established through a constraint-based formalism (as in HPSG and LFG, for instance), so there is no inherent logical order of derivation. Furthermore, constraints may be violable, so that structural complexity (and less than perfect grammaticality) can arise through constraint conflict.¹⁰
- Because semantics is formally independent of syntax, it can be richer in structure than syntax: sentence meaning need not be entirely encoded in the words plus the syntactic structure. The default relation of syntax to semantics is indeed a homomorphism, as in Interface Uniformity, but mismatches are common and to be expected. Where mainstream theory resolves mismatches in the relation between covert and surface syntax, the parallel architecture resolves them in the syntax-semantics interface component.¹¹ These mis-

10. Optimality Theory is one instantiation of a theory with violable constraints, but there are others such as Harmonic Grammar (Smolensky and Legendre 2006), the preference rule systems of Lerdahl and Jackendoff 1983 for tonal music and of Jackendoff 1983 for semantics, and various forms of default logic.

11. Marantz (2005: 440) curiously inverts the intent of this assumption: “The very existence of these [interface] rules, one supposes, is meant to explain the appearance of isomorphism among the levels of structure...” Quite the opposite: they are meant to explain the frequent *deviations* from isomorphism.

matches can be investigated in detail because the parallel architecture, unlike mainstream theory, is linked to a well-developed and cognitively plausible theory of semantics, Conceptual Semantics (FL: Chapters 9–12).

- Words and rules of grammar are formalized in a common format, as pieces of structure stored in memory. Words and productive morphemes are small-scale interface rules, licensing correspondences between idiosyncratic pieces of phonological, syntactic, and semantic structure. Rules of grammar such as phrase structure rules and syllable structure rules are constraints that license structure within a single component of grammar. Other rules license regular relationships between two components, for instance stress rules (syllabic structure to stress grid) and argument structure linking rules (syntax to semantics). In between the words and the regular rules lies a continuum of other pieces of stored structure, including idioms and meaningful syntactic constructions. This feature of the parallel architecture is shared with HPSG, Construction Grammar, and some versions of Cognitive Grammar.

The parallel architecture requires defense against two lines of criticism: from those who wish to maintain syntactocentrism, and from those who think I have not gone far enough in rejecting it. The former line includes the TLR responses by Marantz and by Boeckx and Piattelli-Palmerini. Marantz more or less dismisses the parallel architecture, endorsing the critique by Phillips and Lau (2004), to which I therefore must respond as well. The latter includes the responses by Goldberg and Del Giudice and by Bybee and McClelland (Tomasello and Lieberman would probably be counted among these if their articles had referred to FL).

3. The parallel architecture is too radical: Critiques from the right

Marantz and Phillips/Lau voice what I gather is a common reaction among linguists of the mainstream persuasion: the parallel architecture requires three or more “generative engines” plus the interface components, and so it is a priori less economical and elegant than the syntactocentric architecture, which requires only syntax as its “generative engine.” A first reply to this reaction is that mainstream theory has for the most part neglected formal characterization of semantics and even of phonology (except where morphosyntax is involved), so it is only natural that syntax should appear to be the only formal “engine” needed.¹²

12. For example, as far as I know, the theory of Spell-Out (after a dozen years of the Minimalist Program) has not made clear how the derivation converts syntactic categories such as VP into prosodic categories such as Intonational Phrase, which do not correspond one to one with syntactic constituency (FL: 118–121). Phillips and Lau propose to deal with the mismatch between intonation and syntax in processing terms, evading the formal question of how the

Phillips and Lau admit, moreover, that the well-formedness of sentences may well depend on *constraints* specific to semantics or phonology. This turns the tables, for now the syntactocentric architecture requires two sorts of principles, algorithmic rules and constraints, whereas in the parallel architecture *all* components consist of constraints – including the syntax. Furthermore, the syntactocentric architecture admits precisely the sorts of constraints that the parallel architecture takes to be generative. Thus criteria of economy and elegance in fact prefer the parallel architecture.

A variant of this argument has come up several times in conversation: the parallel architecture seems to presume that one generates structures separately, then checks to see whether they can be matched up. This is a less elegant method than generating everything at once syntactocentrically. But this confuses competence models with performance models. The parallel architecture on its own is not a theory of sentence perception or sentence production: it is not talking about generating sentences in real time – any more than the syntactocentric theory is.¹³ Since the theory is stated in terms of constraints rather than algorithmic ordered derivations, there is no implied order of derivation: the well-formedness of sentences is simply characterized in terms of a network of simultaneous constraints. We will see how the architecture relates to a processing model in Section 7.

Phillips and Lau concede the possibility of a generative system for conceptual structure (which they allow may be “possibly even hierarchical” – how could it not be?), but maintain the traditional view that linguistic semantics is separate from conceptual structure and therefore may still be entirely derived from syntax. However, they offer no arguments or citations of arguments against the position of FL (Section 9.7) that linguistic semantics *is* conceptual structure (see also Jackendoff 1983 and the evergreen classic Bolinger 1965). Recent psycholinguistic evidence (Hagoort 2005) confirms this position at the neural level: there is no distinction in the brain areas and brain signatures involved in detecting semantic anomaly and in detecting factual violations. Thus Phillips and Lau’s appeal to a separate linguistic semantics in order to save the

competence theory accounts for it. They cite in their favor several studies in which prosodic contours have been derived from syntactic structure. But of course these studies were conducted under the assumptions of the syntactocentric architecture, so the authors never considered alternative accounts. See Frazier, Carlson, and Clifton 2006 for some recent experimental work on the partially homomorphic relation of intonational phrasing to syntactic structure and how it influences parsing. See also Jackendoff 2007a for further remarks on prosody in parsing within the parallel architecture.

13. This criticism typically comes from people who have no problem claiming that the order of a traditional syntactic derivation bears no relation to the order of processing: as we were always taught, one doesn’t begin producing a sentence by starting with the symbol S, then expanding it, doing transformations, and finally coming out simultaneously with a pronunciation and what it means. Why do they now revert to interpreting the parallel architecture this way?

syntactocentric architecture is without basis other than tradition. (And even if Phillips and Lau were correct, a parallel-architecture style interface would be required between semantic and conceptual structure.)

Looking further afield, my work with Fred Lerdahl in the 1970s on a grammar for tonal music began by trying to develop an algorithmic grammar that would generate “all and only” the “grammatical” pieces of music. This rapidly proved impossible, and the architecture we arrived at (Lerdahl and Jackendoff 1983) was in fact a parallel architecture with four independent components connected by interfaces. In some respects it drew encouragement from the concurrent developments that led to phonology being conceived of in terms of independent tiers. In particular there were important parallelisms in the treatment of rhythm in music and phonology.

For another example of parallel components, consider the relation of language to vision, such that we can talk about what we see. It is hard to imagine that there is a single “generative engine” that governs both linguistic and visual representations. However, it is quite natural to imagine a structured correlation between linguistic and visual representations, i.e., an interface of precisely the sort the parallel architecture proposes *within* language (Jackendoff 1987, 1996a). In other words, multiple “generative engines” connected by interfaces seem to be commonplace elsewhere in cognition. This seems to me to count as *prima facie* evidence in favor of the parallel architecture and against the syntactocentric approach.

In the parallel architecture, each kind of structure is segregated into its proper component. For instance, the phonological aspects of words belong to the same component as the phonological aspects of phrases. In the syntactocentric architecture, by contrast, the phonology of words is a matter for the lexicon, while the phonology of phrases is part of the PF component of a syntactic derivation – two unrelated components. This fails to capture the fact that, whatever differences between word and phrase phonology, the well-formedness of phonological entities such as syllables and stress grids constitutes a unified formal system and should not be divided among unrelated components. Similarly in semantics: consider that the verb *enlarge* and the phrase *make larger* both contain a causative component whose effect is an event of something getting larger. In the parallel architecture, this is a natural consequence of mapping the same formal semantic structure into two possible phonological/syntactic realizations. In the syntactocentric architecture, it is a coincidence that a word, built in the lexicon, should have much of the same semantic structure as a phrase, built in LF.¹⁴

14. Mainstream theory has sometimes attempted to account for this fact by claiming that *enlarge* is built “in the lexicon” by syntactic movement (e.g., Hale and Keyser 2002). But this is theoretically incoherent. If the lexicon is supposed to be where words are stored, how can

Because the parallel architecture relates its components through a partially homomorphic interface rather than through a derivation, it offers straightforward accounts of mismatches between components. For example, *I'm* is a single phonological word, but in syntax it is a combination of a proper noun, a verb, a tense, and an agreement marker. The parallel architecture encodes this simply as a noncanonical relation between a phonological word and a syntactic nonconstituent. But in a derivational theory, what syntactic node should dominate *I'm*? It can be neither a noun that contains an affixed verb nor vice versa – both are syntactic monsters. Similarly, the syntactocentric architecture predicts that uniform semantics is expressed by uniform syntax (there was a time when this was called the “Universal Base Hypothesis”); thus crosslinguistic differences have to be localized in the mapping between LF and the surface. In the parallel architecture, though, it is no surprise that different languages map the same meanings into radically different syntactic constructions, as illustrated by Myachykov, Tomlin, and Posner (2005: 357) for information structure in English and Russian.

The parallel architecture pays a price for keeping each kind of structure segregated from the others: one can no longer insert lexical items as a whole into syntactic structure, because their phonological and semantic structures do not belong in syntax. The solution is to think of “lexical insertion” as involving all three components at once. “Inserting a lexical item” means establishing a local correspondence among fragments of phonological, syntactic, and semantic structure licensed by the lexical item. This makes a lexical item a sort of interface rule. It is perhaps the most intuitively disconcerting feature of the parallel architecture, but it proves to have its benefits, as will be seen.

Marantz quite rightly observes that *something* in the grammar has to govern the combination of stored structures into larger constituents. He advocates the Minimalist Program’s Merge: “two elements A and B are ‘merged’ to create a constituent C” (2005: 441). The corresponding device in the parallel architecture, as in all the constraint-based variants of generative grammar such as LFG, HPSG, and Construction Grammar, is unification. Merge and unification are not equivalent. Unification allows for a constituent to consist of two elements that partially overlap. For instance, the unification of (1a) and (1b) is (1c), in which the *V* of the two elements overlaps. By contrast, Merging these two elements yields (1d).

movement rules apply within it? In addition, the Hale/Keyser hypothesis is severely flawed on syntactic and semantic grounds that have been known since the 1960s in the context of the Generative Semantics debates (Culicover and Jackendoff 2005: 53–56).

FL argues (Section 12.1) that the combinatorial principles for phrasal meaning are a *subset* of those involved in word meaning – which is why phrasal definitions are rarely complete, and why the Hale/Keyser hypothesis is inadequate, despite its initial appeal.

- (1) a. [V, +past]
 b. [_{VP} V NP]
 c. [_{VP} [V, +past] NP]
 d. [[V, +past] [_{VP} V NP]]

Formally, Merge can be characterized as the unification of two elements with a piece of tree structure whose terminal elements are variables:

- (2) Merge A and B = [A B]
 Unification of A, B, and [x y] = [A B]

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 goal of unifying cognitive science urges selection of unification as the fundamental combinatorial operation.¹⁵

A linguistic theory with unification rather than Merge as its combinatorial device also permits a natural formalization of contextual features such as sub-categorization. For instance, an obligatorily transitive verb such as *devour* can be formalized syntactically as (3), where the *NP* in italics stands for a typed variable that must be satisfied by unification.

- (3) V *NP*

In a theory based on Merge, lexical items are supposed to consist of X^0 categories alone. Such a theory does not countenance a lexical item like (3) with internal syntactic structure. Moreover, if it did, the result of merging it with an NP would be [[V *NP*] NP] rather than the desired [V NP]. Mainstream theory is therefore forced to encode contextual features in terms of additional theoretical machinery such as abstract Case and Case-checking. These are “virtual conceptual necessities” only because Merge has been recognized as the only “simple” possibility for an account of combinatoriality.

15. Another case to consider is song, where linguistic and musical representations are involved simultaneously. Unification offers a natural account in terms of superimposition of syllables and fixed pitches; Merge offers only the curious possibility of a sequence consisting of a syllable and a pitch.

Phillips and Lau complain that FL contains few detailed arguments for the parallel architecture. By this I presume they mean arguments of a traditional syntactic sort, for FL offers many arguments based on the prospects of unifying linguistics, both internally and within the context of cognitive science, as well as many arguments about the structure of semantics and its relation to syntax. The requisite syntactic arguments now have appeared in *Simpler Syntax* (Culicover and Jackendoff 2005), which demonstrates that all the standard syntactic phenomena – phrase structure, argument structure, passive, raising, control, ellipsis of all sorts, long-distance dependencies, and binding – find more adequate analyses within the parallel architecture than in mainstream accounts. At the same time, syntactic structure can be drastically simplified in terms of the number of nodes in standard trees and in the number of abstract elements; and the notion of movement can be eliminated altogether. To be sure, there is a price in the complexity of the syntax-semantics interface component, but Culicover and I show that in many respects this complexity is necessary even for mainstream theory, and it is still far less than the savings incurred in the syntactic component.

It should be stressed that this approach to syntax is possible only because of the highly articulated theory of meaning, Conceptual Semantics, embedded in the parallel architecture. This independently motivated formalism for semantic structure is what enables one to see where semantics parallels the form of syntax, or fails to. Jackendoff 1983 stressed the way Conceptual Semantics parallels syntax, in a fashion impossible for standard treatments from formal logic. This set a standard within which Jackendoff 1990 was able to pick out *mismatches*, leading to the partial independence of syntax and semantics advocated in FL and developed in *Simpler Syntax*. Mainstream theory, lacking an independent theory of semantics, has no way to evaluate the degree of match between syntax and semantics, and Interface Uniformity forces a match at whatever cost in the covert component.¹⁶

The use of unification as the essential generative operation leads to the possibility of the fourth innovation in the parallel architecture. Consider first a VP idiom like *kick the bucket*. Merge requires that this be built up from its individual words, leading to the question of where the idiomatic meaning ‘die’ is situated in the lexicon. Because of recursive compositionality, its meaning must be localized in one of the words of the idiom. Marantz (2005: 443) proposes that it’s localized in the lexical entry of *bucket*, which is therefore polysemous; this meaning is licensed when *bucket* is in the context of *kick* and *the*. He ne-

16. In fact, Construction Grammar (Goldberg/Del Giudice) and usage-based grammar (Tomasello and Bybee/McClelland) suffer from the same problem: instead of presuming that semantics follows syntax, they presume the reverse. But because they do not have an explicit and independently motivated theory of semantics, their claims cannot be rigorously evaluated.

glects to observe that *kick* and *the* also require special (null) readings when they are in the context of each other and this special meaning of *bucket*. Hence his account actually encodes all the idiomatic information in all three words, redundantly. (And why is the actual meaning encoded in *bucket* rather than *kick* or even *the*?)

By contrast, the parallel architecture basically adopts HPSG's approach to idioms. *Kick the bucket* is encoded as a lexical VP that links to the meaning 'die'; here a semantic unit corresponds to a composite syntactic constituent, violating recursive compositionality in Marantz's sense. The idiom can unify with an entire VP in syntactic structure, and its verb can be unified with tense along the lines sketched in (1a–c).

This approach extends naturally to VP constructions that carry special meanings over and above the meanings of the words. For instance, the meaning of (4a) is 'go past the house screeching', and the meaning of (4b) is 'spend/waste 3 hours knitting', even though *screech* is not a verb of motion and *knit* is not a verb that takes a time period as an argument.

- (4) a. screech past the house
b. knit 3 hours away

Because of recursive compositionality, mainstream theory is obliged either to attribute these constructions to systematic polysemy of some unlimited number of verbs (e.g., *screech* means either 'emit screeching noise' or 'move while emitting screeching noise'; *knit* means either 'knit' or 'spend time knitting'), or to derive them from syntactic structures closer to those of their paraphrases, via some peculiar stratagem. The parallel architecture, here in concurrence with Construction Grammar, encodes the constructions as VP idioms in which the surface verb is interpreted as a manner expression and the surface argument structure is a consequence of the meaning of the construction. (5) is a rough encoding; for more detail see Jackendoff 1990: Chapter 11; Goldberg 1995; FL: Section 6.6; and Goldberg and Jackendoff 2004.

- (5) a. [_{VP} V PP] = 'go PP, [V, emitting sound]-ing'
b. [_{VP} V NP away] = 'spend [NP, time period] V-ing'

These lexical items unify with the verb and the arguments in standard fashion, adding their own special piece to the meaning. There is no extra complexity in the formalism used to encode such items. The only price is giving up strict recursive compositionality.

From here it is only a small step to allow syntactic structures in the lexicon that have no particular association with meaning, for instance (6a). This is a notational variant of the standard phrase structure rule (6b); it stipulates a

“treelet” that can be unified with others to build (or license) larger structures.¹⁷

- (6) a. $[_{VP} V NP]$
 b. $VP \rightarrow V - NP$

Thus the parallel architecture ends up being able to say that words, idioms, constructions, and phrase structure rules are all encoded in a common formalism, permitting a reduction in the sorts of formal entities in the theory.¹⁸

This conclusion is altogether at odds with mainstream theory, which, following traditional grammar, considers words and rules to be entirely different theoretical entities. As far as I know, idioms have received only cursory mainstream treatment, and there are no general proposals at all concerning constructional idioms like (5), which bridge the gap between idioms and phrase structure rules.

Moreover, all the mainstream arguments about learnability of grammar have been directed toward the learning of rules, with no relation drawn to the learning of words. For instance, Boeckx and Piattelli-Palmerini (2005), in speaking of the optimality of narrow syntax, ignore the lexicon entirely. Similarly, Fitch, Hauser and Chomsky (2005) characterize the narrow language capacity as “an abstract core of computational operations.” This excludes the lexicon, since lexical items on the standard view are not “computational operations.” In fact, mainstream theory has no overall view of the structure of the lexicon – quite a stunning gap. In view of this, it is difficult to take seriously Boeckx and Piattelli-Palmerini’s expressions of optimism toward the Minimalist Program.

4. The Parallel architecture is not radical enough: Critiques from the left

Some TLR authors criticize aspects of generative grammar without observing that FL offers an alternative. For instance, Lieberman (2005: 128) objects to the

17. Moreover, within a constraint-based formalism, it is possible to separate constraints on constituency from constraints on constituent order, as advocated by GPSG several decades ago. Culicover and Jackendoff 2005 work out some details of English phrase structure and word order variation in these terms, incorporating semantic and pragmatic factors, thereby answering some of the objections to mainstream theory mounted by Myachykov, Tomlin, and Posner (2005: 353).

18. Pinker’s (1999) distinction between words and rules survives in the theory in somewhat different form. For Pinker, irregular inflectional forms are stored, and the same is true in the present story. Pinker’s treatment of regular inflection is a rule: “Add *-d* to form past tense.” In the present approach, the regular past tense affix is stored in the lexicon as such; it has contextual features to say what it attaches to. Consequently, regular past tenses are formed by unification of this affix with a stem. That is, there are still two different ways to get past tenses, but they both fall out of the common formalism. See FL: 163–165.)

algorithmic formalization of mainstream theory and its emphasis on formal elegance, both addressed in FL. Tomasello, Bybee and McClelland, Goldberg and Del Giudice, and Myachykov, Tomlin, and Posner all bewail the syntactic bias of mainstream theory, without acknowledging that a main motivation behind the parallel architecture is to give proper weight in the grammar to phonology, semantics, and pragmatics as well as syntax.

4.1. *Arguments from Construction Grammar*

Goldberg and Del Giudice in particular attack the hypothesis that there are autonomous syntactic principles that are not somehow connected to semantics. Their specific target is English subject-auxiliary inversion, where they show that there is a family resemblance among the semantic/pragmatic situations in which it is used: questions, counterfactual conditionals (*had he left, ...*), after initial negatives with sentential scope (*nothing could he see*),¹⁹ wishes (*may you plotz!*), exclamatives (*does it ever stink!*), *so*-inversion (*so do I*), and optionally in comparatives (*... than does Harry*); one might add also *so-Adj/Q* inversion (*so tall was Harry that ...*). They argue that these are all situations in which either the clause has non-assertive force or the subject is non-topical.

I don't want to dwell on the empirical questions raised by their analysis.²⁰ The more important point is that the paper's agenda, like that of Goldberg 1995, 2005, is to defuse arguments for an autonomous syntax. Goldberg's version of Construction Grammar rejects the possibility of rules like (6a), which are not paired with a meaning or with a (somewhat arbitrarily related) family of meanings. The TLR paper is directed toward Newmeyer's (2000) argument that SAI is an example of autonomous syntax. However, there exist more convincing examples than SAI, where the syntax of the language offers no choices that can be put to use to express semantic differences. These include the fixed position of the determiner in noun phrases; the position of the verb in sentences (if it *is* fixed); what kind of agreement system the language has; whether there are object clitics, and if so, whether they are attached to the verb, as in Romance, or fall in second position, as in some Slavic languages, and whether they occur only in the absence of full arguments or must be doubled.

19. Goldberg and Del Giudice attribute the observation of the scope differences between *For no money would she quit* and *For no money, she would quit* to Lakoff and Brugman 1987. In fact this contrast is discussed earlier by Klima 1964 and Jackendoff 1972, for instance.

20. For example, they do not explain why SAI is obligatory in some of these constructions but not others, or why in standard English it does not occur in indirect questions. They do not explain why *yes-no* exclamatives invert but *wh*-exclamatives in modern English do not (*What big eyes you have!*). They also do not explain why there are two other constructions in English used to express non-topical subjects: *there*-insertion and locative inversion (*into the room ran Harry*), and why SAI can't be used instead of the latter (**into the room did Harry run*).

Another example of autonomous syntax is English *do*-support. Goldberg and Del Giudice may be right that the semantics of assertive force helps account for why the tense-bearing element is singled out for inversion, but this does not explain why English inverts *do* in clauses bearing no other auxiliary, rather than just inverting the finite verb like most reasonable languages. This is just a stupid fact of English syntax. Yet another case is the English verb-particle construction, which exhibits the same syntactic pattern for at least five different kinds of meaning (Jackendoff 2002c):

- (7)
- a. John put the book down. (Particle is a locative argument alternating with a PP)
 - b. John looked the answer up. (Verb + particle is an idiom)
 - c. John ate the sandwich up. (*Up* is an aspectual particle)
 - d. John sang his heart out. (*X's heart out* is an idiom and the verb is independent)
 - e. John knitted the night away. (NP + *away* is an idiom and the verb is independent)

Unlike Construction Grammar, the parallel architecture allows both for autonomous syntactic constraints like (6a) and for interface constraints that connect syntactic structure to meanings, like (5). There is nothing theoretically unnatural about the former: autonomous phrase structure constraints are paralleled by autonomous phonological constraints such as phonotactic rules and by autonomous semantic constraints such as proper conditions for binding variables by quantifiers.

Goldberg and Del Giudice also are concerned to show that the meanings of constructions, like all human categories, have a family resemblance character. This is an expected result in Conceptual Semantics (FL: Section 11.6; summarizing discussion in Jackendoff 1983: Sections 3.1, 5.2.2, 7.3, 7.4, and 8.1–8.7); it is an important concurrence with Cognitive Linguistics (Lakoff 1987; Langacker 1987) and an important break from the logical tradition.²¹

4.2. Arguments from phonetics

Far more radical are the TLR critiques which, observing that language cannot be analyzed in terms of classical categories, therefore claim that the attempt to

21. Bybee and McClelland (2005: 387) praise cognitive linguistics for connecting meaning with more general properties of human categorization, especially in the domain of spatial relations and its metaphoric extensions. They do not mention that Conceptual Semantics also embraces these connections, which were treated in detail in Jackendoff 1983 prior to their popularity in cognitive linguistics.

formalize linguistic theory in terms of algebraic symbolic systems (or “mental representations”) is misguided.

Two of these, MacNeilage/Davis and Lieberman, involve phonology and phonetics. MacNeilage and Davis argue that many tendencies in phonology can be attributed to the effects of controlling the vocal tract, and therefore that statements of these tendencies in terms of formal structure miss the point. They argue that the basic form of the syllable emerges as an elaboration of the “mandibular cycle”, rooted in chewing movements, and that phonological structure is an emergent property of programming motor movements, “the epigenetic result of normal experience” (2005: 177). While they may be right about constraints on syllable structure being imposed by the need to control the vocal tract, this does not explain why animals, who also chew and suck, do not develop phonology; their chewing motions do not get digitized into discrete categories. Nor, as stressed by Goldin-Meadow (2005: 201), does this approach explain why, phonetics aside, language in the signed modality shares so much structure with vocal-auditory language. The pressures on phonetics from the anatomy and basic control of the vocal tract are undoubtedly crucial to explaining language – they are perhaps of the nature of the “interface conditions” to which Chomsky and Boeckx/Piattelli-Palmerini appeal. But they cannot explain all of language by any means.

Lieberman offers a related argument: the detailed phonetics of languages cannot be reduced to a finite set of innate distinctive features; and even within languages, different speakers have different ways of producing the same sounds, precluding a direct relation of phonetics to motor control. From this he jumps to the conclusion that the entire formal analysis of phonological theory and possibly even linguistic theory as a whole is to be rejected. I have some sympathy for the facts he points out and their implications for SPE-style distinctive feature theory; similar arguments have been offered by Port and Leary 2005; Ladefoged 2005; and Mielke 2006. But the conclusion is hyperbolic. It offers no acknowledgment of the insights behind traditional formal analysis, and it offers no way to approach phonology, morphology, and syntax, much less meaning. So the question is how to reconcile these observations with linguistic theory.

Suppose the language acquisition device does not come equipped with a set of innate phonological distinctive features. What it still needs as a prerequisite for learning – and which is present in no other natural communication system – is the presumption that the signals are organized into a structured and digitized space. That is, a child comes to language acquisition expecting that the sounds being emitted by others can be structured into syllables that in turn can be decomposed into speech sounds, and that the speech sounds and their combinations into syllables form a structured taxonomy. Part of the child’s job then is to discover this taxonomy – and how to produce it. (By contrast, the theory

of innate features claims that children already *know* the taxonomy and how to produce it.) If this is the case, languages (and dialects) can specify very precisely the phonetic values of sounds and can contain unusual sounds that don't fit nicely into the standard feature repertoire, and these can be learned from experience. Still, the sound system of a language *is* a system, because that's how learners construct it. The fact that phonological systems are not collections of arbitrary sounds then becomes a consequence of the well-understood functional pressures of acoustic and articulatory distinctness (as well as the sorts of pressures discussed by MacNeilage and Davis), rather than being built into the feature system from the start. In short, it may be possible to weaken UG in a way that copes with Lieberman's observations. But that does not mean UG can be eliminated. Something must be there to attune children to the organization of the system they are trying to achieve.

4.3. *Arguments from irregularity and noncompositionality*

Bybee and McClelland (2005) construct an argument similar to Lieberman's, but on a larger scale. They emphasize the vast number of irregularities and semiregularities in phonology and morphology, the family resemblance character of linguistic categories, the dependence of many syntactic and morphological principles on semantics, and above all the role of historical change in producing closed-class morphemes through grammaticalization. They correctly point out that no system of totally general syntactic rules can possibly account for this overall texture in linguistic structure. Like Lieberman, they jump from these observations to a wholesale rejection of the standard tools of formal linguistics – the quote in section 1.3 gives the essence of their conclusion. Unlike Lieberman, they advocate an alternative: connectionist modeling, which lacks discrete representations. Language for them is an emergent property of training a connectionist network.

Bybee and McClelland do not hesitate to follow their argument to its astonishing conclusion (2005: 402):

It is true that the coding of inputs and outputs in some form is always necessary, but the form can be a relatively raw sensory representation as provided by peripheral sense organs such as the ear or a relatively raw motoric representation such as the time series of neural signals to the muscles controlling oral articulators. In principle, at least, all other forms of representation can be derived through the learning process itself, gradually adjusting the strengths of connections that determine how particular patterns of activation in the sense organs are recoded to allow the network to assign its own metric of similarity relevant to capturing the structure present in the set of items it is tasked to learn.

This is associationism of the most radical sort, well beyond that of Tomasello

(2005: 184–185) and Elman et al. (1996), who acknowledge that there are innate biases in the learning machinery of the brain, and who deny only that there are any such for language. Bybee and McClelland are rejecting the possibility of *any* innately structured behavior, from sneezing to birdsong to sex to parental care to complex social interaction, in disregard of the voluminous animal literature (e.g., Wilson 1980; Gallistel 1990; Hauser 2000).

In particular, they are ruling out the possibility that anyone could develop language (or anything else!) without a “set of items [they are] tasked to learn.” This position is immediately called into question by the well-known work on Home Sign by Susan Goldin-Meadow and associates (including her article in TLR), which shows that deaf children with *no* linguistic input nevertheless develop remarkably language-like systems. To be sure, these systems are impoverished compared to standard languages. But what can you expect of a single child working alone? In particular, Goldin-Meadow shows evidence of closed-class morphology despite the absence of opportunity for grammaticalization due to historical change.

Bybee and McClelland lump FL with mainstream generative grammar in its commitment to the possibility of free combinatoriality in language, and therefore take FL to be disproven by the sorts of evidence they adduce. Yet chapter 6 of FL is devoted to incorporating much of this very evidence into linguistic theory: the existence of vast tracts of irregularity and semiregularity in the lexicon, and in particular the existence of vast numbers of memorized combinations – not only idioms but also fixed collocations such as *catcher’s mitt*, *baby-blue eyes*, *faster than a speeding bullet*, *weapons of mass destruction*, and *pick on someone your own size*, not to mention proper names, movie titles, and song lyrics. Unlike Bybee and McClelland, FL does not conclude that combinatoriality therefore can be dispensed with in the foundations of linguistic theory. Rather, the problem is how to allow for *both* regularity and irregularity.

The theory of the lexicon sketched in Sections 2 and 3 has the requisite property. A lexical item is more word-like to the extent that it contains idiosyncratic information, and more rule-like to the extent that it contains more variables and less specified information, and to the extent that its variables are of broad types. Thus the item (6a) that encodes the English VP construction is more general and productive than the constructional idioms in (5), which are in turn more general and productive than a fully specified VP idiom like *kick the bucket*. A lexical item is also more regular to the extent that it falls into an inheritance hierarchy (or family resemblance) with other items, so *kick the bucket*, a normal VP, is more regular than syntactic outliers like *by and large* and *day in day out*.

Bybee and McClelland somehow want to say that stored combinatorial forms don’t exactly count as combinatorial. In particular, on p. 396, they claim that such items “lose their compositionality if they are of high absolute or relative

frequency.” But if they initially have compositionality that they can lose, how can Bybee and McClelland dispense with compositionality?

The position that Bybee and McClelland are arguing against (2005: 398) is that the only thing stored in memory is primitives, and everything else is built up on line. But surely a theory that includes primitive categories and combinatorial rules need not claim that only primitives are stored. For example, the phonology of words is clearly composite, and it is clearly stored, not built on line. Similarly, taking a cue from Culicover’s TLR article, consider jazz improvisation. Culicover argues that in the course of learning to play jazz, players accumulate a repertoire (or lexicon) of musical fragments that can be combined to create novel improvisations. The process of becoming a better player involves (in part) acquiring a richer set of patterns and more flexible ways to apply them and combine them. Frequency of occurrence helps: that’s why musicians practice. But this does not mean that what they are playing (and their mental representations of what they are playing) ceases to contain individual notes and chords that the player has to produce. A player’s memory of one of these learned patterns still has to be built out of relatively primitive categories like *c-sharp* and *third degree of major scale* – plus composite categories like *subdominant flatted ninth chord* (even if, for musicians who only play by ear, the categories don’t have these conventional *names*). These categories are deeply involved in performing an improvisation as well as in intuitively judging its felicity and/or originality.

Similarly for (morpho)syntax. Of course, sometimes grammaticalization changes the syntactic status of an item. For instance, Bybee frequently presents the development of English future *gonna* from *going to* + purpose clause. It is clear that *gonna* is no longer a verb, since it cannot be inflected and doesn’t occur where other present participles can (*I like going to/*gonna swim*). But this isn’t the case with *kick* in *kick the bucket*. Even if *kick the bucket* is stored as a unit and has lost its *semantic* compositionality, the unit *kick* is still *syntactically* identified as a verb, as we can tell by the fact that tense and agreement go there. So syntactic compositionality in this idiom has remained (partially) intact. Similarly for the dozens of idioms with *take* and *give*: their past tenses all use *took* and *gave*, showing that the irregular verb has not been swallowed up in the larger gestalt.

Although it is undeniable that grammaticalization through historical change is an important source of closed-class morphology, I find a lot left unexplained. What is it about the semantics of *go*+purpose that lends itself to being bleached out into a future? And why can something that means ‘want’ bleach out to practically the same thing (English *will*)? More generally, what explains the “possible trajectories of change” (Bybee and McClelland 2005: 386) in the semantic domain? For instance, why couldn’t *going to* reduce phonologically to *gonna* without bleaching its original meaning? Why do certain sorts of meanings get

encoded as closed-class morphemes while others don't? Without a substantive semantic theory that includes treatments of both open-class and closed-class items and the semantic relations between them (which Bybee and McClelland admit they lack (2005: 404)), answers based on grammaticalization cannot be the end of the story. A substantive semantic theory, at least according to the evidence with which I am familiar (FL: Chapters 9–12), will have to presuppose just the sort of innate grain to human thought and motivation that Bybee and McClelland's radical associationist approach to mind denies.

4.4. *Arguments from connectionism*

Finally let us turn to Bybee and McClelland's embrace of connectionism. FL demonstrates that there is a cline of generality in language from idiosyncratic to rule-like, acknowledges that frequency effects, priming, and competition play a role in processing and acquisition, and agrees that acquisition is usage-based, all of which make sense in a connectionist framework. And I do think these are all important considerations that are revealing something deep and true and new about language, something foreign to mainstream theory.

But appealing to connectionist models makes it hard to retain something deep and true and *old* about language. The job of language isn't to produce past tenses of verbs given the present tense, and it isn't to predict the next word of a sentence in an artificial language with a vocabulary of 50–100 words, the sorts of tasks that connectionist models have been built to accomplish. Such abilities are purely ancillary to what linguistic ability really comprises, namely the ability to map between an unlimited class of structured sounds and structured meanings, based on a large but finite class of stored structures. Meanings in turn must be sufficiently structured to support inference, integrate with discourse context, and interact with perception and action. These fundamental observations motivate every variety of modern linguistic theory. Furthermore, all standard linguistic theories give us a handle on how to analyze sentences like (8), which is (to be just a little cruel) the opening sentence of Bybee and McClelland's article:

- (8) There is a range of views about the relationship between language as an abstract system and what people actually say when they talk – what de Saussure called 'langue' and 'parole'.

Despite over twenty years of research on connectionist modeling, no connectionist model comes close.²²

22. See FL (32: N. 7; 59, N. 17; 62–63, N. 21) on the purported attempts to get networks to deal with embedding. To be sure, there are always new models on the market, and Bybee

There are principled reasons for this, detailed in FL (Section 3.5) as “four challenges for cognitive neuroscience.” These are not just problems for language; FL argues that they recur in many other cognitive domains. Boiled down the essentials, they are:

- A neural network has no way to represent the relationships among items in a sentence being perceived – to build structure on line – because there is no independent working memory. The activation corresponding to a sentence is basically the activation of the sentence’s words (plus, in Elman’s recurrent network, some extra representation of the most recent words). This makes it impossible for a standard network to distinguish *The dog chased the rat* from *The rat chased the dog*, much less from **The the dog rat chased*. The recurrent network allows for linear order effects, but still not unlimited long-distance dependencies such as the subject-verb agreement in *The man in the yellow shirt with red and green horizontal stripes that are covered with paint stains and political buttons likes/*like ice cream*. (These *do* occur in real speech.)
- The problem of 2: A neural network has no way to represent multiple occurrences of the same item in a sentence. For example, in *My cat can beat up your cat*, the two occurrences of *cat* have to be semantically linked to different possessors. But this is impossible if the only effect of *cat* being in the sentence is that the lexical entry for *cat* (whether holistic, compositional, or distributed) is activated.
- The problem of variables: A standard neural network cannot encode a general relation such as *X is identical with Y*, *X rhymes with Y*,²³ or *X is the past tense of Y*. Connectionists (including Bybee and McClelland 2005: 403) claim that there are no such general relations – there are only family resemblances among memorized items, to which novel examples are assimilated by analogy. But to show that there is less combinatoriality than you thought is not to show there is *no* combinatoriality.²⁴ Gary Marcus (1998, 2001),

and McClelland cite some. Still, if someone really had an account for the sort of complexity illustrated in (8), Bybee and McClelland would certainly tell us about it. Instead they recycle the usual examples of Rumelhart and McClelland’s (1986) past tense learner and Elman’s (1990) Simple Recurrent Network.

23. Note that rhymes cannot be all memorized. One can judge novel rhymes that cannot be stored in the lexicon, because they involve strings of words. Examples are Gilbert and Sullivan’s *lot o’news/hypotenuse*, Ira Gershwin’s *embraceable you/irreplaceable you*, and Ogden Nash’s *to twinkle so/I think so*. Moreover, although *embraceable* is a legal English word, it is probably a coinage; and *thinkle* is of course a distortion of *think* made up for the sake of a humorous rhyme. Hence these words are not likely stored in the lexicon (unless one has memorized the poem).

24. Ferreira (2005: 378) observes of connectionist models: “there are reasons to worry that the architecture of such models requires certain simplifying assumptions to be made about representations (e.g., recursion is problematic, as is any type of identity rule).” In a note she adds,

in important work that has been met with deafening silence by the connectionist community,²⁵ demonstrates that neural networks *in principle* cannot encode the typed variables necessary for instantiating general relations, including those involved in linguistic combinatoriality.

- In neural networks, long-term memories are encoded in terms of connection strengths among units in the network, acquired through thousands of steps of training. This gives no account of one-time learning of combinatorial structures, such as the meaning of *I'll meet you for lunch at noon*, a single utterance of which can be sufficient to cause the hearer to show up for lunch.

A few people in the connectionist paradigm have addressed these questions of combinatoriality, for example Pollack 1990, Shastri and Ajjanagadde 1993, van der Velde and de Kamps 2006, Smolensky and Legendre 2006, and Feldman 2006. I must confess I don't fully understand their solutions. But evidently mainstream connectionists don't either, as none of these proposals have achieved currency in that community.

I think it is no accident that Bybee and McClelland's examples come primarily from phonology and morphology, where free combinatoriality plays a less predominant role than in syntax and semantics. In general, connectionist models have been popular only in dealing with noncombinatorial phenomena such as word recognition, parsing of small finite artificial languages, and certain kinds of pattern recognition; these are areas of psychology and psycholinguistics where the capabilities of networks provide some foothold. But as soon as one scales up to combinatorial structure in language or in any other cognitive domain such as complex visual fields, planning, or social cognition, these mechanisms are inadequate. By contrast, FL (Chapter 6) proposes a synthesis, on which Bybee and McClelland unfortunately do not comment.

"Some connectionists have dealt with this weakness by asserting that recursive structures are rare in natural speech [reference omitted], a claim which is certainly false."

The ongoing "past tense debate" is telling. It has typically been couched (by both sides) in terms of whether or not one needs a combinatorial account of the regular past tense in addition to the associationist account of irregularities and semiregularities: are there two mechanisms or only one *for the past tense*? What this misses is that the *regular* past tense, the one the connectionists wish to do away with, is the tip of a combinatorial iceberg which includes verb phrases, relative clauses, argument structure, massively regular morphology in languages such as Turkish, and so on. Stamping out the regular past tense does not eliminate the need for a combinatorial mechanism for the rest of these phenomena. See FL: 154–167.

25. For instance, Bybee and McClelland do not cite Marcus; nor do any of the papers in a 1999 special issue of *Cognitive Science* entitled "Connectionist Models of Human Language Processing: Progress and Prospects"; nor was he cited other than by me at the 2006 LSA symposium entitled "Linguistic Structure and Connectionist Models: How Good is the Fit?"

5. More on acquisition and the nature of Universal Grammar

As might be expected, the major outlier among the contributors on the issue of acquisition is Boeckx and Piattelli-Palmerini's paper. They sing the virtues of the innate parameters of Principles and Parameters theory, and seem to (I can't be sure) liken the effects of parameters in language acquisition to those of regulatory "master" genes in epigenetic development. There are at least three problems with taking this analogy too seriously. First, the "rich principles that UG makes available" (2005: 450) have to be coded on the genome somehow (as noted in FL: 190). Their richness is at odds with the Minimalist Program's hypothesis that language doesn't need much special built in, and that the language capacity may even be "the result of a deep restructuring, started by a sudden genetic transposition, and then submitted to one of the ubiquitous computation-optimizing process acting on the organization of the brain" (Boeckx and Piattelli-Palmerini 2005: 461). Second, Boeckx and Piattelli-Palmerini's account of language acquisition is completely silent on the acquisition of words. If words and rules form a continuum, as FL has shown, then rules cannot be reduced to a small set of possibilities.

Third, Boeckx and Piattelli-Palmerini conjecture that the set of parameters will come to resemble the periodic table: a relatively compact structured set. This view was promising during the 1980s, but has come to seem less likely. Newmeyer 1998, Culicover 1999, and Ackerman and Webelhuth 1999 point out that as the variations among languages have been more deeply explored, and as more and more phenomena have been discovered that arguably fall outside canonical grammar (Culicover's "syntactic nuts"), the explanatory value of parameters has steadily diminished. They argue that no small set of parameters will be adequate, and perhaps not even a large finite set will.

Nor can the mainstream escape this last argument by arguing that parameters apply only to "core" grammatical properties. First, the core is never defined in a non-question-begging fashion. Second, as Culicover 1999 points out, once one has developed a non-parametric acquisition theory that is adequate for learning the periphery, there is no reason it cannot be extended to the core – or at least no argument has been given that it cannot. Boeckx and Piattelli-Palmerini do not engage with these arguments, preferring to remain on the high ground of philosophy of science.

Tomasello (2005: 186–189) mounts similar arguments against parametric acquisition. He presumes this is the only possible version of Universal Grammar, so he thinks that by demolishing parameter theory he has demolished UG as well. He further doubts that there is really a problem with poverty of the stimulus, and argues against it on the basis of Chomsky's old off-the-cuff argument about the ungrammaticality of **Is the man who smoking is tall* (Tomasello 2005: 189–191), which has also come under attack by Pullum and

Scholz (2002). Like Bybee and McClelland, Tomasello overlooks far more robust arguments for Universal Grammar (perhaps because the one who gave them wasn't Chomsky), many of which are discussed briefly in FL: 82–90. In particular, Goldin-Meadow's (2005) TLR paper, based on decades of well-publicized research, discusses the creation of systems of Home Sign, about the most extreme version of poverty of the stimulus. Home Sign systems are quasi-languages with words, morphology, (possibly) grammatical categories, concatenation, hierarchical structure, function-argument organization, and topic-focus organization – all developed with *no* linguistic input. This is poverty of the stimulus to the max. As far as I can tell, it completely undermines Tomasello's argument.

Tomasello, Bybee/McClelland, and Goldberg/Del Giudice advocate a “usage-based” theory of acquisition, in which children acquire language based on statistical patterns of input. Rules (or rule-like generalizations) are the outcome of generalizing over the behavior of words, rather than the result of setting parameters. To some degree this is a no-brainer: there's no way that a child can learn where verbs go in a sentence without first learning some words. None of them observe that FL (187–190) comes out in favor of usage-based learning.

Because the FL theory codes words and rules in the same format, an account of rule acquisition emerges rather naturally. Suppose there are a number of similar items in the lexicon, acquired through usage. Their similarity can prompt the learning mechanism to construct a new item which (a) incorporates the parts these items have in common and (b) substitutes a variable for the parts in which they differ from each other. Because of the variable, the item will be more rule-like, and can combine with novel new items. The original items are not thereby eliminated from the lexicon, so there is some redundancy between words and rules (an issue stressed by Walenski and Ullman in TLR). Iteration of this process over similar rule-like items can lead to still more general rules in the lexicon.

Under this account, an important problem for acquisition is to determine which linguistic phenomena are productive and which are not.²⁶ Here, as usage-based theorists emphasize, statistical patterns play an important role; and it is not because the brain is actually counting occurrences, but because neural activation and strength of connection depend on frequency. Thus these aspects of acquisition theory rest ultimately on the neural level of analysis and perhaps play no direct role in the formal level.

26. For Bybee and McClelland, this problem doesn't arise, of course, because they believe there are no truly productive phenomena.

As seen above, Bybee and McClelland think nothing more is necessary for learning language than usage-based generalization (*without* variables). Tomasello thinks there is more, but nothing specific to language: only intention-reading and cultural learning, schematization (which *might* be the process of building items containing variables, though he doesn't put it this way), analogy, entrenchment and preemption, and functionally (i.e., semantically) based distributional analysis (Tomasello 2005: 193). Language universals, for him, result from "universal processes of human cognition, communication, and vocal-auditory processing" (189), plus the process of grammaticalization. Nevertheless, his explanations presume that learners *do* know something specific about language: they know about focus and presupposition (191), they can identify things like adverbial and restrictive relative clauses (191), and above all, since grammar is supposed to be semantically bootstrapped, they have semantics. He claims that children face no linking problem between syntax and semantics because "semantic and pragmatic function are part of the definition of the construction" (192). But this presupposes, first, that semantic and pragmatic function are sufficient to determine syntax, which they are not (see Myachykov, Tomlin, and Posner's point about expression of information structure in English and Russian above); second, that there are no purely syntactic facts of language, which there are (see Section 4.1 above); and third, that children somehow have access to "the definition of the construction", i.e., its semantics, in advance of learning the construction.

More basically, Tomasello implicitly takes for granted that children identify language as having *words*, made up of phonological units, which, as is well known, are not present in any direct way in the signal. Moreover, it seems unlikely that the child's interpretation of noises as words is predicted by a general "communicative function" in cognition, since no other human or animal communicative functions have the same rich expressive power as language. Again, the wrong quantifier has been substituted: just because *some* aspects of language can be explained through more general cognitive devices doesn't mean *all* of them can.

FL acknowledges that many aspects of language come from general cognition and communicative function, and stresses the influence of semantics on grammatical structure (chapters 5 and 12; see also Culicover and Jackendoff 2005, which amplifies the point considerably). But FL suggests that there have to be at least two features specific to language, which might be called UG or the narrow faculty of language (see also Pinker and Jackendoff 2005; Jackendoff and Pinker 2005). One feature is the basic architecture. Children come expecting that something adults do will be a combinatorial expression of intended meaning. While children have to discover the phonology (and perhaps even the phonological features) of their language, they don't have to discover that there *is* phonology. While they have to discover the morphological and

syntactic structure of their language, they don't have to discover that there *are* morphology and syntax, which offer canonical ways for them to express meaning. While they have to discover how their language realizes semantic arguments and information structure, they don't have to discover that these aspects of semantics exist.

In some respects, the basic architecture of the language capacity is like a toolkit: languages may or may not have fixed word order, morphological case, phonological tone, or noun incorporation. These are tools that children come prepared to use if the input warrants it and to discard otherwise. Their existence is not predictable from general cognitive function. Perhaps they play roughly the role that parameters play in mainstream theory, only far more flexibly.

The other important feature of UG, not necessarily separate, concerns the acquisition of rules. As mentioned above, FL conceives of rule-learning as a process that adds to the lexicon new items containing variables, which generalize over existing items – an approach that is possible only because rules and words are coded in a common format. However, given a collection of items over which generalizations are to be made, there are bound to be conflicts or ambiguities about what constitute the proper generalizations. FL proposes that UG contains a set of “attractors” that favor certain generalizations over others – highly general pieces of structure towards which language learning is drawn. Any particular language may or may not observe these generalizations, or it may contain subsystems that violate them, but the existence of the attractors will lead to statistical tendencies among languages. In a way these attractors play the role assigned to markedness theory and the evaluation metric in early generative grammar: they do not dictate the grammar of a language, but they say which grammar consistent with the data is to be favored. FL (192–193) suggests some candidates for principles of UG along these lines; Culicover and Jackendoff (2005: 40–41) offer slightly more detail.

6. “Resilient” and “fragile” layers of UG

Chapter 8 of FL suggests that the language capacity is “layered”: some of its components are (relatively) more essential to communication, and others are refinements that improve communication but could not function alone as a communication system. The more essential components appear first in acquisition, and they appear in undeveloped or impaired systems such as pidgins, agrammatic aphasia, and language acquisition after the sensitive period. These components include conceptual structure (and the presumption that the communication system expresses it), phonology, the large lexicon, and the use of linear order to express semantic relations. The refinements include syntactic

categories, phrase structure, functional categories, and inflectional morphology.

This rough division is based on Bickerton's (1990) "protolanguage" hypothesis, but is intended as a less hard and fast division. Like Bickerton, FL proposes it in the service of a story about the possible evolutionary course of the language faculty: the more robust components would have to have developed first, but they could have functioned adequately on their own prior to the advent of the modern language faculty. In a sense, this layering parallels our understanding of the eye: some sort of vision would be possible with a retina and a lens but without muscles to move the eyeball, but not vice versa. Similarly, useful communication would be possible with concatenated words lacking syntactic organization, but syntactic organization is useless without words. Section 9 returns to evolutionary issues.

TLR offers two further confirmations of this layering. Goldin-Meadow (2005) compares the inventory of features in FL with the features of Home Sign, and discovers that Home Sign aligns for the most part with the "robust" features of FL's UG. Home Sign has symbols (2005: 203), some morphology (203), linear order of concatenated symbols correlated with thematic roles (205), grammatical categories such as noun/verb that are not identical to semantic categories such as object/action (215), clauses that can be conjoined (211), time reference (but not tense per se), expressions of illocutionary force, and negation (214). (It may not have phonology in the standard sign language sense (203).) She calls these the "resilient" features of language – those that children can develop without access to input. She suggests that development of the more "fragile" features may require true interaction with other speakers and/or a time depth of more than one generation, as has occurred in Nicaraguan Sign Language, for instance. She leaves open whether these features are due only to cultural/historical processes or whether, as FL suggests, they also require some influence from UG.

Goldin-Meadow also argues (2005: 205–213) that the form of Home Sign is driven in part by performance constraints, in particular a pressure to keep the length of utterances down. She shows that this constraint interacts with the argument structure of the proposition being expressed. In expressing an action with only one argument, such as running, the probability that its Agent will be expressed is higher than the probability that the Agent will be expressed with an action with many arguments, such as giving. Evidently, the semantic arguments are competing for expression, and the competition is governed by the constraints on utterance length. In order for this to be the case, the children must understand the structure of actions – again something not present in the input – even if they can't (or don't) express the structure in full.

To be sure, the correspondence between Goldin-Meadow's "resilient" features and FL's "robust" features is not one-to-one. But the two accounts do

agree that UG is not a monolithic entity, and that successful quasi-linguistic communication does not require that every feature of adult language be utilized.

A second and unexpected manifestation of the layering of the language faculty emerges from the paper by Baynes and Gazzaniga (2005), who investigate the linguistic capacities of the right (non-language) hemisphere. Their studies are based primarily on split-brain patients. Although there is tremendous variability among individuals, the overall results are eerily familiar. Patients do acquire a right-hemisphere lexicon “that appears to be commensurate with their dominant hemisphere” (2005: 321). They typically show auditory and orthographic word recognition and semantic priming of words in the right hemisphere. The right hemisphere “does have some sensitivity to syntax, but it is not able to consistently use syntactic information to guide comprehension” (319), although, surprisingly, it is able to make some grammaticality judgments. It is difficult to test phonological abilities in the right hemisphere, but typically patients have difficulty judging whether printed words presented to the right hemisphere rhyme. Baynes and Gazzaniga cite a hypothesis of Zaidel 1978, that “the right hemisphere is an acoustic processor whereas the left hemisphere is a phonetic processor” (320), a distinction that is no doubt a bit overdrawn but suggests a direction of bias.

The biggest difference between right-hemisphere language and Goldin-Meadow’s “resilient” features is that the right hemisphere is markedly limited in its ability to *produce* speech. This is presumably because this ability has been inhibited by the left hemisphere, and indeed some split-brain patients over time begin to initiate limited speech in the right hemisphere.

The overall conclusion is that “right hemisphere language may be more widely distributed in the normal population than is represented by figures derived from attempts to establish language dominance. Modern imaging methods suggest a broader role for the right hemisphere in language processing although the functional significance remains to be defined” (322). Baynes and Gazzaniga suggest (323) the “the right hemisphere may support the remnants of a biologically older protolanguage” with “a dependence on a semantically-based word order and a lack of morphology”; “conceptual/semantic knowledge appears to be the most well-developed of the major language components.” In short, right hemisphere language provides a further piece of evidence for the layering of the language faculty.

Returning for a moment to architectural concerns: it is significant that protolanguage has at best rudimentary phrase structure and syntactic categories, its linear order being driven directly by semantics and pragmatics. In a syntactocentric architecture, the existence of protolanguage is therefore something of a mystery: it would seem otiose to assume that its phonological and semantic structures are generated by a standard syntactic component. In the parallel

architecture, by contrast, phonology and semantics can be linked by a direct interface, which of course includes the words, plus nonsyntactic principles of linear order, e.g., “place the Agent first”, “place modifiers adjacent to what they modify.” Full language can be seen then as adding syntax and its interfaces as a further refinement or “supercharger” on this more primitive linking (FL: 261–264). Psycholinguistic evidence is beginning to emerge suggesting that this primitive linking goes on even in normal adult speakers, creating processing difficulty when it conflicts with syntactically-determined composition (Ferreira 2003, who calls it “shallow processing”; Kuperberg 2007).

FL (249–250, 255–256) suggests that some aspects of grammar in “full” languages retain the flavor of protolanguage, in that they are governed more by semantics, pragmatics, and linear order than by any explicit elaboration of syntactic structure. One case is English noun-noun compounds, where the semantic head is linearly to the right, and the rest of the semantic relation between the nouns is determined by finding a pragmatically appropriate relation between them – with no further syntactic reflex. Another case is sentence adverbials, which can be of practically any syntactic form (adverb, PP, AP, participial, full clause, or small clause) and which can appear sentence-initially, sentence-finally, or after the subject; their semantic relation to the main clause is purely a function of the semantics of the adverbial. That is, there is very little syntactic constraint on either the form of the adverbial or its position, and the semantics and pragmatics are doing most of the work. Compared to the canonical case of argument structure, where both the syntactic category and position of arguments is highly regulated by syntax, the organization of these cases far more closely resembles protolinguistic principles. Thus the layering of grammar is not confined to abnormal cases such as pidgins, aphasia, and right hemisphere language. It is very much alive within the grammar itself.

7. The parallel architecture and language processing

When we turn to processing, the papers in TLR present the by now expected dichotomy. Marantz (2005: 439) says that of course theoretical linguistics is relevant to studies of processing, and all linguists have to take the Derivational Theory of Complexity seriously (despite its apparent failure in the 1970s). He maintains that the Minimalist Program “allows for more straightforward testing and falsification of linguistic hypotheses”; “the syntactic computations described in the theory are necessary to the representations that they derive and thus speakers and listeners must carry out these computations” (440). This is all to the good; however, he then turns to a discussion of compositionality and never shows how a Minimalist derivation is to be interpreted in processing terms.

The other side of the dichotomy is represented most clearly by Ferreira (with support from Walenski/Ullman and Lieberman). She observes (2005: 369) that processing theories based on Government-Binding Theory have been primarily concerned with finding gaps of movement – that is, reconstructing S-structure – rather than with running syntactic derivations in reverse to derive D-structure. As for the Minimalist Program (MP), she finds it “highly unappealing from the point of view of human sentence processing” (370). One reason is that in the MP, the derivation begins with the most deeply embedded lexical items (in English, usually on the right) and works up to the topmost node. This “obviously is difficult to reconcile with left-to-right incremental parsing”, which decades of research have confirmed to be the way people do in fact parse. Ferreira further argues that the MP architecture makes it difficult to account for syntactic reanalysis based on semantic anomaly, because by the point in the derivation where semantic information can be evaluated, syntactic information has been purged from the tree. She takes this to be “a case where a basic mechanism of minimalism is completely incompatible with known facts about human processing (which were published in mainstream journals more than a decade ago)” (371).

In a sense, Phillips and Lau’s (2004) discussion of processing from a Minimalist perspective amplifies Ferreira’s point. They assert (with caveats) that “it is certainly possible in principle to run current transformational grammar derivations in either direction” (10) – though notice that in parsing it is necessary to run derivations backward from phonetics to syntax and then forward from syntax to semantics (and vice versa in production).

Phillips and Lau acknowledge the necessity of incremental processing, but are worried about how this can be achieved. “If real-time processes build only well-formed structure, and if the grammar defines well-formedness in terms of possible constituents, then we might expect that structure-building should occur only when a string is identified that forms a possible constituent” (11). That is, they see anticipatory or predictive structure building as theoretically problematic. This is because in MP, the rules of grammar can create structure only through Merge of two constituents: there can be no open structure waiting to be filled by elements yet to come. Phillips and Lau cite experimental evidence that speakers do indeed build anticipatory structure, and find most theoretical approaches wanting in this respect. They conclude (2004: 16) that

... the main challenge for unification in this area involves the question of how to build structures accurately and incrementally in real time. This challenge could be viewed as the ‘Logical Problem of Language Processing’, and it remains somewhat mysterious under most theoretical approaches.

In other words, they more or less grant Ferreira’s point, though they find the glass half full rather than mostly empty.

In FL, the organization of the lexicon developed in Chapter 6 is the key to the model of processing proposed in Chapter 7. Words and rules of grammar are stored in memory in a common format, namely as pieces of structure. In processing a sentence either in perception or production, the goal is to construct in working memory a complete triplet of phonological, syntactic, and semantic/conceptual structures, linked in well-formed fashion. There are two parts to the procedure: retrieving candidate pieces of structure from long-term memory (lexical access), and integrating the pieces into a larger structure. The pieces are integrated by unification: “clipping” pieces of structure together in such a way that one piece satisfies contextual variables of the other. When alternative possibilities present themselves, whether in words (lexical ambiguity) or in larger structures (e.g., syntactic ambiguity), competing candidate structures are maintained in working memory until one wins the competition. This leaves room for feedback from semantics or even from visual perception (e.g., Tanenhaus et al. 1995) in establishing syntactic and even phonological structure.

Because the grammar is stated simply in terms of pieces of structure, it imposes no inherent directionality: in production it is possible to start with conceptual structure and use the interfaces to develop syntactic and phonological structure; and in perception it is possible to start with phonological strings and work toward meaning. More specifically, in language perception, auditory/phonetic processing creates a string of phonetic structure in working memory that leads to a call to the lexicon for items with that phonetic structure. Because these items are interface rules, they bring with them pieces of syntactic and semantic structure that can be immediately integrated with any syntactic and semantic structure already present (including material from semantic context). If a word cannot yet be integrated, its syntactic category in working memory motivates a call to the lexicon for syntactic structures in which it can be embedded.

For example, hearing the phonetic string (9a) calls up the word *the* (9b), which includes syntactic and semantic structure. (9a) and (9b) unify to form (9c), since the phonetic string overlaps entirely with the word’s phonology. The syntactic *Det* feature in (9b) calls the lexicon for a phrase structure in which *Det* is a variable, and (9c) is retrieved. (9c) unifies with (9b) to form (9d). This in turn calls to the lexicon for a phrase structure in which *the* can be embedded, yielding (9e), which unifies with (9d) to form (9f).

- (9)
- | | | | |
|----|----------|---|--------------------------------------|
| a. | ðə | | |
| b. | Phon: ðə | Syntax: Det | Sem: DEF |
| c. | | Syntax: [NP <i>Det</i> . . . <i>N</i>] | |
| d. | Phon: ðə | Syntax: [NP <i>Det</i> . . . <i>N</i>] | Sem: [. . . DEF . . .] |
| e. | | Syntax: [_S NP <i>Aux</i> VP] | |
| f. | Phon: ðə | Syntax: [_S [NP <i>Det</i> . . . <i>N</i>] <i>Aux</i> VP] | Sem: [. . . [. . . DEF . . .] . . .] |

Notice how this model of processing incorporates predictive parsing in precisely the sense described by Phillips and Lau. The anticipated constituents consist entirely of well-formed pieces of structure, but they terminate with variables instead of words, creating constraints on how subsequent material is to be parsed. For another case, recall how a transitive verb encodes its subcategorization: the syntactic structure of the verb is not simply *V*, but *V NP*, where the *NP* is a contextual variable. Thus when the verb is retrieved into working memory, it brings along anticipatory structure that has to be filled by subsequent material.²⁷ Likewise, a *wh*-word is lexically marked as having the syntactic structure *wh...t*, where the trace must be satisfied at a later point at some unspecified position in the tree.

This model parallels Ferreira's description of parsing using Lexicalized Tree-Adjoining Grammar (2005: 375–376), which likewise pastes together elementary trees listed in the lexicon. I believe the parallel architecture's lexicon is a bit more flexible than L-TAG in accommodating mismatches between phonology, syntax, and semantics, but the spirit is the same. In any event, it does incorporate the integration of semantics, syntax, and phonology that Ferreira desires (378), and there is nothing at all "mysterious" about how the competence theory is related to the performance model.

The parallel architecture's detailed semantic theory helps make the processing model more precise than models based solely on syntactic parsing and an inexplicit semantics. In addition, as mentioned in Section 3, this semantics makes it possible to reduce the complexity of syntax to the sorts of structures that people working on processing have assumed for decades, eliminating the elaborate structures posited in late GB and the Minimalist Program.

Walenski and Ullman's (2005) paper is devoted to the balance between storage and computation in language processing (though they do not observe that Chapter 6 of FL is devoted to this very topic). Like the mainstream, they see a distinction in formal type between words and rules; they call the lexicon a kind of "declarative memory" and the grammar "procedural" or "computational." In the parallel architecture, by contrast, there is no such distinction. Rather, the "procedural" aspect of a rule is the integration process in working memory by which its variable is satisfied, i.e., the procedural manifestation of unification. As far as I can tell, all their discussion can easily be reinterpreted from this perspective (see FL: 164–165).

From the point of view of integration of linguistic theory, FL makes a larger point. The question of what is stored in long-term memory and what is computed online would normally be considered an issue of performance. For instance, di Sciullo and Williams 1987, a work on the theoretical notion of

27. Culicover and Jackendoff 2005 (Chapter 6) deal with some of the complications that arise because of passive and raising.

“word”, considers this question to be “of no interest to the grammarian.” Yet this issue proves to be of utmost importance in developing the basic architecture of a theory of competence, as it motivates the form of the parallel architecture’s lexicon, in particular the continuum between words and rules.

The relationship is in fact a two-way street. The parallel architecture makes it possible to run experiments that actually test linguistic hypotheses. For example, consider the phenomenon of aspectual coercion, which is responsible for the sense of repeated jumping in *Joe jumped until the bell rang*. A syntactocentric architecture requires this sense of repetition to be present either in the syntax or in the lexical meaning of *jump*, which therefore must be polysemous (along with every other point-action verb in the lexicon). In contrast, the parallel architecture/Simpler Syntax approach treats the sense of repetition as a bit of semantics necessary to make the meaning well-formed, but it lacks any syntactic reflex. It turns out that the difference between the two approaches is not just a matter of theory. Piñango, Zurif, and Jackendoff 1999 test for processing load in this type of example, contrasting it with non-coerced examples such as *Joe slept until the bell rang*. Extra processing load does indeed show up in the coerced examples, in a time frame consistent with semantic rather than syntactic or lexical processing, just as predicted by the parallel architecture.²⁸ Preliminary results on light verb constructions (*John gave Bill an order*) similarly show that they create extra semantic processing load compared to ordinary verb-argument constructions (*John gave Bill an orange*), again in confirmation of the parallel architecture/Simpler Syntax account (Piñango, Mack, and Jackendoff 2006).

Thus the parallel architecture is more responsive than mainstream approaches to Marantz’s desire for integrating linguistic theory with psycholinguistics. Unlike mainstream theory, it does not call for ungainly structures that cannot be verified experimentally, nor does it lead to fundamental mysteries about how parsing takes place. Faced with this disconnect, many psycholinguists have relied on semantic network or connectionist approaches for theoretical grounding, even though, as Ferreira notes (2005: 378), these approaches leave much to be desired in terms of rigorous linguistic description. The parallel architecture, it seems to me, achieves the correct balance. It permits perspicuous description of linguistic structure, and on its model of processing, the

28. Piñango, Zurif, and Jackendoff test processing load during auditory sentence comprehension by measuring reaction time to a lexical decision task on an unrelated probe. The timing of the probe establishes the timing of the processing load. Another strand of psycholinguistic research on coercion, e.g., Pytkäinen and McElree 2006, also finds evidence of increased processing load with coercion. However, their experimental technique, self-paced reading, does not provide enough temporal resolution to distinguish syntactic from semantic processing load.

kind of incremental lexically driven parsing that has achieved broad consensus among the psycholinguistic community falls out as a natural consequence.

8. What's special about language?

An issue that comes up repeatedly in the papers in TLR and that has already come up here several times is the degree to which the language capacity is simply an assemblage of mental faculties that serve other purposes besides language. As Wilkins puts it (2005: 273), "we do not know *a priori* which features of the cognitive system are necessarily language specific." Hauser, Chomsky, and Fitch (2002) have stated the issue in terms of a dichotomy between the "broad language faculty", which includes everything in the brain necessary for learning and using language, and the "narrow language faculty", namely those parts of the broad language faculty that are specific to language. All things being equal, one would like to be able to attribute as little of language as possible to the narrow language faculty, both for the sake of integration with the rest of cognitive science, and also so as not to place too great a burden on evolution for developing the distinctive aspects of language.

Hauser, Chomsky, and Fitch conjecture that the narrow faculty of language consists exclusively of syntactic recursion (Merge), and that all other properties of language come from the interaction of recursion with independent properties of sound production and auditory processing (the "sensorimotor interface") and of conceptualization (the "conceptual-intentional interface"). In TLR, Boeckx and Piattelli-Palmerini enthusiastically endorse this conjecture. Four objections have already arisen here. First, the existence of "syntax-challenged" protolanguages – which other animals do not develop – shows that recursion is not necessary to a human communicative system. Second, the operation responsible for recursion is not Merge but unification, which is a widespread process in cognition, hardly unique to language. Third, this view excludes words from the narrow language faculty, since they are not "computational mechanisms." It claims that the presence of words in language comes for free from pre-existing ways of structuring sound and meaning (although Hauser, Chomsky and Fitch do offer some mitigating discussion on this point). This seems unlikely: if the mechanisms are present, why don't other animals have words? Fourth, since words and rules form a formal continuum, it is impossible to isolate the general recursive properties of grammar from the words, idioms, and constructions.²⁹

29. For critiques of many other aspects of Hauser, Chomsky, and Fitch's position, for example that language is not "for" communication and that language may be "perfect", see Pinker and Jackendoff 2005; Jackendoff and Pinker 2005. For a response to the former of these, see Fitch, Hauser, and Chomsky 2005.

As has already been noted, several papers in TLR advocate eliminating the narrow language faculty altogether. Tomasello attributes language acquisition to universal properties of cognition, generalization, and cultural learning (whatever *that* is – see Jackendoff 2007b); Myachykov, Tomlin, and Posner advocate a similar position. Goldberg and Del Giudice attribute acquisition to general properties of categorization. Lieberman attributes it to general properties of temporal sequencing in motor control. MacNeilage and Davis attribute it to practiced refinement of a particular motor pattern, the mandibular cycle. Bybee and McClelland seem to deny that there is anything special about *anything* in the mind beyond primitive sensory and motor representations and general principles of Hebbian learning. The basic problem with all these arguments is that even if these more general faculties play a role in language, they do not account for *all* of language, any more than recursion does.

Another aspect of language, not mentioned in TLR, also apparently draws on more general capacities: our memory for tens of thousands of idiosyncratic and general lexical items, as well as the fast learning processes that create these memories. I know of no estimates of the number of tunes people know (or can at least recognize) – folk songs, popular songs, advertising jingles, religious chants, and (for some people) 30-minute symphonies. But informal estimates mount up very fast, and my suspicion is that the total is in the thousands (consider the number of tunes people have on their iPods that they actually know). Think also of all the artifacts that anybody knows how to use: buttons, zippers, shoes, shoelaces, socks, hats, beds, pillows, forks, various kinds of knives, can openers, cans, tinfoil, paper, doorknobs, shower stalls, faucets, toilets, cars, computers, books, televisions, VCRs, DVDs, and on and on. It's hard to know how to measure this knowledge, but it's vast. Also think of all the ways you know to get around cities: all the streets and landmarks. In light of these other cases, the magnitude of lexical learning seems less special.

Nor need language be the sole user of “language areas” in the brain. Evidence is mounting that much temporally sequenced hierarchical structure is constructed by the same part of the brain – roughly Broca's area and subcortical areas connected to it – whether the material being assembled is language, dance (Lieberman 2005: 297), hand movements (Lieberman 2005: 294; Wilkins 2005: 279), or music (Patel 2003). Nor, within language, is Broca's area confined to syntax, as often asserted: it also appears to play a role in phonological and semantic combinatoriality, possibly in distinct though overlapping subareas (Hagoort 2005; Poeppel and Embick 2005).

A more nuanced account of the balance between general and special has already been hinted at several times above. Language acquisition of course rests on social skills such as theory of mind (rudimentary in chimpanzees) and understanding pointing (not found in chimpanzees), as well as on more general perceptual machinery such as attention. These provide scaffolding for language

acquisition but do not themselves provide the material out of which knowledge of language is built. Similarly, the use of language requires long term memory and working memory, both of which are general faculties that extend over all cognitive domains. Language use also requires temporal sequencing and the online construction of hierarchical structure, both of which appear also in motor control, the planning of action (see Jackendoff 2007b: Chapter 4), and probably visual action perception – not to mention music. Thus *all* higher mental capacities make use of the same sorts of basic machinery: memory, attention, and the construction of structure.

What differentiates these capacities from each other, however, is the character of the structures they build and how these structures interact with the rest of the mind. What makes language *language* is that it has phonology – a distinctive way of encoding sound and sound production (or in the case of sign language, gesture) – plus a distinctive interface between phonology and conceptual structure. In turn, within this interface is yet another unique form of mental structure, syntactic structure. It is these structures and interfaces that make possible the productivity of language through what Hockett (1960) called “duality of patterning.” As far as we know, there is no other cognitive capacity among animals that codifies sound and sound production the way phonology does,³⁰ and there is no other cognitive capacity that productively associates complexes of sound with complex concepts the way the linguistic lexicon and syntactic structure do.

The point of repeating these well-worn observations is that new kinds of mental structure don’t come for free in the brain. Knowledge of streets and landmarks builds on preexisting capacities for spatial navigation. Knowledge of how to use artifacts builds on preexisting capacities for manipulating objects and constructing action patterns. The mental representations involved in both of these are at the moment unknown. Knowledge of music indeed builds on the auditory capacity, but the digitization of pitch into the patterns of musical scales requires something added to the cognitive repertoire, and the recursively hierarchical structure of musical meter is (as far as we know) shared only with language (Jackendoff and Lerdahl 2006). Similarly, knowledge of phonological structure builds on the auditory capacity and patterns of motor control for the vocal tract, but it is not just a more sensitive auditory system or a more efficient way of controlling the vocal tract: it requires representational innovations, as

30. Some birdsong and cetacean song has internal structure, sometimes quite complex. But (a) its decomposition is in many respects different from phonology; (b) its combinatorial structure does not interface with combinatorial meaning; and (c) even so, these vocalizations are not in a related species, so their evolution must be independent (though perhaps driven by similar functional considerations of how to produce a large number of distinct messages). For a summary of apparent differences between animal vocalizations and human phonology (many crying out for experimental verification), see Yip 2006.

does the large-scale productive linkage between this encoding of sound and encoding of thought.

In short, what is special about language is the collection of mental structures it employs – phonology and syntax – and the interfaces between them and conceptual structure. The processing and acquisition of these structures may be accomplished by brain-general mechanisms of long-term memory, integration in working memory, learning (including statistically-based learning), and attention, and may rely as well on understanding of social interaction and theory of mind.³¹ But unless the specific unique building blocks for phonology, syntax, and their connection to concepts are present, language and language acquisition are not possible. It is these building blocks that constitute the narrow language faculty.³²

9. Evolutionary issues

We finally come to the issue that everyone wants to know about the most, but about which there is the least to say: the evolution of the language capacity. There is no direct evidence for the course of language evolution. The emergence of the modern vocal tract in anatomically modern humans (Lieberman 1984) doesn't mean language didn't exist before then – it only means that language with modern phonetics didn't exist before then. The appearance of art and ritual objects between 100,000 and 50,000 years ago probably means that *something* important changed in human cognition. But was it language? We don't know. In the absence of direct evidence, speculation runs rampant, and for the most part, people's theories of language determine their theories of language evolution.

31. This conclusion is not new. For example, MacDonald, Pearlmutter, and Seidenberg (1994: 700) conclude: "... our approach suggests that whereas there may be distinctly linguistic forms of *representation*, the *processing* principles that account for language comprehension and ambiguity resolution are not specific to language at all. Rather they seem to reflect general properties of memory, perception, and learning, properties that are involved in nonlinguistic domains as disparate as concept learning, pattern recognition, and decision making." (However, MacDonald, Pearlmutter, and Seidenberg's connectionist account of language processing is not without its serious problems; see Jackendoff 2007a.)

32. Notice how this view differs from Fodorian modularity. For Fodor (1983), language perception is an isolated automatic process that has no connection with the rest of the mind, except that somehow the "shallow representations" it produces (surface structure? LF?) are picked up by domain-general mechanisms of central cognition. Moreover, Fodor claims that modules are innate, neglecting that one doesn't have a functioning language module until one has learned a language. The present view is that the mental representations that constitute language are processed (and possibly acquired) by domain-general mechanisms. However, information flow to and from language, and within language itself, is constrained by the character of linguistic representations and the interfaces among them. See FL: 218–230.

Those who believe there is nothing special about language (including several authors in TLR) of course believe there was no “evolution of the language faculty.” Rather, all that needs to be explained is the evolution of the capacities from which language is cobbled together: voluntary control of the vocal tract, theory of mind, cultural learning, and so on, depending on the theorist. Since all these other capacities must be accounted for anyway, this approach certainly places the least burden on evolution and on the genome. The problem is not with this approach to evolution, which is the null hypothesis: it’s with this approach to *language*, as I have been emphasizing throughout the discussion.

The Minimalist Program ups the ante only slightly. It claims that the only special part of language is recursion, and the rest is exaptation of existing capacities plus epigenetic consequences in the brain of adding recursion into the mix. Thus, as noted earlier, Boeckx and Piattelli-Palmerini suggest that the language capacity may be the product of a single mutation. I have sketched here many of the reasons why the Minimalist Program is inadequate as a theory of synchronic language, especially in the context of the rest of cognitive science. These would not be redeemed by a highly attractive approach to language evolution (see the previous paragraph). Nevertheless, it’s worth taking a moment to examine the conjectures that Boeckx and Piattelli-Palmerini put on the table.³³

The relevant passage is this (2005: 461):

Whether this new component [narrow syntax, consisting exclusively of recursive Merge – RJ] started out by allowing the articulation of an inner soliloquy (as Chomsky suggested in recent lectures), and/or by allowing the deployment of the series of natural numbers, and/or by allowing the recursive merge of sets into sets of sets, it’s hard to say at present. Social communication is, once again, out of this evolutionary picture, and N[arrow] S[yntax] does not seem to be at all the outcome of progressive adaptation.

Each of these possibilities is dubious. It is unclear how “the deployment of the series of natural numbers” could facilitate the genesis of language. If anything, the reverse is the case: as is well known, there are plenty of languages that lack the infinite series of natural numbers; in contrast, there are no known cultures that have all the natural numbers but lack language. As for the “recursive merge of sets into sets of sets”, it’s not even clear what this could mean cognitively.

33. These should not be construed as claims. Boeckx and Piattelli-Palmerini’s discussion, like the rest of their paper, is highly derivative of Chomsky’s recent work. Chomsky typically phrases evolutionary discussion in rhetorical terms that maintain deniability, such as: “Such-and-such a hypothesis [the one he’s attacking] is obviously a fairy tale. It’s no better than any other fairy tales that might be told, for instance [the hypothesis he’s (sort of) proposing].” All terms such as “hypothesis” and even “conjecture” that I use in the following paragraphs must be understood under this very cautious epistemic modality. I should add that I disagree with Chomsky: some fairy tales *are* better than others.

Going back to the first of their suggestions: Pinker and Jackendoff 2005 (who Boeckx and Piattelli-Palmerini cite in other connections) point out that an “inner soliloquy”, the inner voice or Joycean stream of consciousness, requires above all *words*. It cannot be the product of pure syntactic structure. And people *acquire* words to use in their inner soliloquies either through social communication or, as in Home Sign, by inventing them in the service of social communication. In the case of Home Sign, recursion demonstrably plays no role at all. Moreover, anecdotal evidence suggests that deaf people who have never been exposed to sign language do not experience an “inner soliloquy.” So overt language for social communication is essential for the ability to experience an “inner soliloquy,” and recursion is not. Jackendoff 2007b (Chapter 3), based in part on Jackendoff 1987 and Jackendoff 1996b, shows how the “inner soliloquy” emerges from the fact that phonological images (or the gestural equivalent for signers) are the most prominent conscious manifestations of thought, and so present themselves to experience as the act of thought itself. In other words, the soliloquy hypothesis too is a nonstarter, though slightly more subtly so.

The soliloquy hypothesis is connected with a broader hypothesis about the relation of language to thought (Boeckx and Piattelli-Palmerini 2005: 461): “Even if we could imagine some hominid ancestor possessing the same sensorimotor system as modern humans, and the same conceptual-intensional apparatus, but lacking recursive Merge, it is very doubtful that such a creature could be capable of thought and expression in anything remotely resembling the human sense.” Well, yes, it couldn’t learn to express itself in English or any other “full” language. But it might well be able to learn a protolanguage such as a pidgin or a Home Sign, which doesn’t require syntactic recursion.

As for “thought . . . in . . . the human sense”, this hints at the rather Cartesian view that combinatorial thought is possible only through having language. Of course, it depends on what you mean by “thought in the human sense.” Cheney and Seyfarth (whose TLR paper I can at last bring to bear) argue that monkeys have a rudimentary theory of mind, and that (2005: 149) “monkeys and apes have many concepts for which they have no words.” In particular, they have combinatorial concepts with dyadic argument structure, especially particularly in social predicates such as “X threatens Y” and “X is dominant to Y”, which they put to use in chains of inference. This may not be thought in the fully human sense, but it is combinatorial thought nevertheless. Boeckx and Piattelli-Palmerini, who along with Chomsky offer no characterization of the “conceptual-intensional system,” give us no hint of how they would account for the combinatoriality of primate thought.

Finally, Boeckx and Piattelli-Palmerini’s line “N[arrow] S[yntax] does not seem to be at all the outcome of progressive adaptation” simply reiterates the hypothesis that language evolved in one step. It’s nevertheless a little over-

stated to say it wasn't an adaptation. In Darwinian theory, *every* evolutionary change emerges from a chance mutation, not from adaptive pressures. However, if the mutation spreads through the population, then we tend to ask which of the resulting phenotypic changes facilitated that spread, and we call those changes adaptive. Are Boeckx and Piattelli-Palmerini claiming that some trait other than language was responsible for the spread of the gene that made language possible? Lacking any more definitive text from them, I abandon this line of argument, and conclude that the Minimalist Hypothesis leads to no viable conjectures about the evolution of the language faculty.

The parallel architecture has more special parts to it than the Minimalist Program, and therefore places greater demands on evolution. As mentioned in the previous section, FL (Chapter 8) addresses the problem by means of reverse engineering: which parts of the language capacity could exist in the absence of others, and still constitute a better communication system than if they were not there? The following partial ordering emerges from this logical exercise. (Numbers are ordered, letters are unordered, for instance there is no logical dependence between 3a and 3b.)

1. Conceptual structure: there must be something for the system to express
2. Use of vocalizations and/or gestures in a voluntary, symbolic, non-situation specific fashion to express concepts
3.
 - a. Digitization of vocalizations and/or gestures in order to permit the invention of unlimited repertoire of symbols
 - b. Concatenation of symbols to express combinations of concepts
 - i. Using linear position to signal semantic relations
4. Phrase structure, whose phrases are projected from lexical heads
5.
 - a. Vocabulary for relational concepts: spatial relations, time, marks of illocutionary force and modality, markers of discourse status, quantification, purposes, reasons, and intermediate causes, and more general discourse connectors (*but, however, therefore, etc.*)
 - b. Grammatical categories, starting by distinguishing verbs from everything else
 - i. Inflectional morphology
 - ii. Grammatical functions

The most significant cut is between 3b.i and 4: the components up to this point give us protolanguage, which, as has been seen, emerges as a prominent "layer" within the modern language faculty.

This is only a logical decomposition, but it provides conjectures about the order in which different aspects of language could have been introduced in the course of evolution. In particular, it is consistent with (a suitably modulated version of) Bickerton's (1990) hypothesis that the hominid line had protolanguage for a long time before modern language developed.

Though FL is not entirely explicit on the matter, it gives the impression that each of these features of language is the product of a separate genetic mutation. As Wilkins (2005) and Marcus (2004) observe, this is probably too much to ask of evolution. So let's see how much can be attributed to other sources.

Step 1, conceptual structure, is presumed to exist prior to the evolution of language. In particular, many aspects of the content of Conceptual Semantics, especially predicates having to do with spatial cognition and features concerned with individuation (such as the type/token and mass/count distinctions), are motivated by their necessity for nonlinguistic cognition.

Step 3b, the concatenation of symbols, is conceivably just the product of stretching the capacity of working memory a little. As seen in the discussion of Home Sign above, producing sequences of symbols is hard, but practice at it could be entrenched through use in the community. Step 3b.i, using linear order to express semantic relations, may well be the product of natural relations of iconicity, along lines often urged by functional grammarians. For instance, the most prominent semantic elements in a sentence have the perceptually most prominent position, either first or last. Step 5a, the vocabulary for relational concepts, may be a natural coinage once there is enough syntactic structure in place that one can tell what the relations pertain to. So these steps could well be cultural inventions rather than requiring genetic changes.

This leaves Step 2, the invention of situation-specific symbols; step 3a, the invention of phonology; Step 4, the invention of phrase structure; Step 5b, the invention of grammatical categories³⁴; Step 5b.i, the invention of inflectional morphology; and Step 5b.ii, the invention of grammatical functions (subject, object, and indirect object).³⁵ I have already given reasons here why I think phonology cannot just be a cultural invention; I will return to Step 2 in a moment. The remaining four are essential parts of the architecture, but fall into the "fragile" layer of language; they are the things that tend not to be present in agrammatism, pidgins, late language acquisition, possibly right hemisphere language, and (on some people's interpretation) certain cases of Specific Language Impairment, including those due to mutation of the FOXP2 gene. This

34. I separate grammatical categories from phrase structure because one could conceivably have phrase structure based solely on semantic categories. The sentential adjuncts mentioned above are a case in point: their semantic function and their position in the sentence are correlated, but their syntactic category is irrelevant.

35. Culicover and Jackendoff 2005 (chapter 6) develop a theory of grammatical functions as an extra tier of the syntax-semantics interface. Grammatical functions appear in some guise in every halfway adequate theory of syntax: in LFG as f-structure, in Relational Grammar as the whole content of the theory, in HPSG as the complement hierarchy, and in mainstream theory as Abstract Case. They have broad ramifications crosslinguistically (despite Tomasello's complaint about the abstractness of the notion subject (2005: 185)): in argument structure, in agreement and case systems, in binding, and in the formulation of passive and raising.

suggests that they have a genetic basis. Whether they are due to separate mutations or consequences of a single mutation is certainly an open question.³⁶

Let's now go back to Steps 1 and 2, the existence of conceptual structure and the ability to express it through symbols. Wilkins (2005: 285) takes FL to task for not paying more attention to these. The reason for doing so was only one of focus. The FL hypothesis was developed in response to proposals like Deacon 1997 and Donald 1991, who quite rightly see step 2, the onset of symbol use, as the crucial breakthrough leading to the evolution of language. The problem is that they think that from there the full apparatus of language can develop as a matter of course. FL elaborates all the logical steps in order to demonstrate that there is a lot more to the language capacity than the use of symbols plus cultural evolution. But Steps 1 and 2 are certainly legitimate subjects of inquiry.

Wilkins focuses on the possible changes in the brain during human evolution that made Steps 1 and 2 possible. She pinpoints two changes from the ape brain as important. First, the parietal lobe has undergone major expansion, to the point where there is a new convergence of parietal, occipital, and temporal association cortex, the so-called parietal-occipital-temporal junction or POT – which includes Wernicke's area. Second, new major tracts of fibers have developed, the arcuate and superior longitudinal fasciculi, which link the POT with Broca's area.

Wilkins sees the development of the POT as crucial in the development of conceptual structure in the human sense: she takes it to be an association area in which all the perceptual modalities come together: visual information from the occipital lobe, auditory information from the temporal lobe, and somatosensory information from the parietal lobe. The result is a modality-neutral representation, necessary for conceptual structure: we conceptualize objects (and refer to them) as independent of our sensory experiences of them (part of the issue that motivates Gross's paper in TLR). Wilkins claims that other primates lack such modality-neutral representations. I am not sure how to reconcile this claim with Cheney and Seyfarth's evidence for primate concepts such as social predicates and some rudimentary theory of mind, which are at a considerable abstraction from perceptual modalities. However, exploring the apparent conflicts between these claims may help sharpen some of the questions about differences between primate and human cognition.

I find myself more interested in the role of the fasciculi that connect the POT to Broca's area. Wilkins sees their development as creating a "connectivity between motor and somatosensory cortices necessary for, among other things, appropriate control of the hand, and very importantly the thumb" (2005: 279–

36. I have more serious concerns about the "attractors" that UG needs in order to make acquisition possible. Do these all have a functional basis, or are they a result of further genetic shaping that primes acquisition to deal with more complex grammars?

280), for which there was selective pressure in the interest of toolmaking and tool use. The result is greatly enhanced voluntary control, organization, and modulation of finely structured motor movements – in the service of meaningful actions.

Let's put this together with an observation of Cheney and Seyfarth: monkeys are pretty good at learning auditory perceptual patterns and understanding what they are about. However, monkeys are severely limited in their ability to produce novel calls or to combine existing calls. The communication bottleneck, therefore, is on the production end. Now suppose that a major enhancement of voluntary organized fine hand function were to generalize to some extent across the motor system to include the vocal tract. The result *might* be a greatly enhanced ability to produce voluntary finely structured vocalization in the service of meaningful actions. Since the auditory system is already ready to interpret these vocalizations, we are *pretty* close to step 2, a facility with inventing and using symbols.

A further piece of circumstantial evidence comes from Baynes and Gazzaniga's paper. There has always been a question of why language is predominantly in the left hemisphere. If the evolutionary source of language emerges from control of the hand, then the correlation of hemispheric dominance for language and for handedness is rather natural. In particular, the variation in degree of dominance for language and its correlation with degree of hand dominance is no accident. Neither are the ways in which degree of dominance manifests itself in the two domains, for example the ability of the nondominant hemisphere to take over functions from an impaired dominant hemisphere.

This story is at best suggestive, of course – and at best it only fills in one of the many steps necessary to the modern language capacity (on the FL account at least). What I like about it, though, is the way it unexpectedly integrates evidence from primate behavior, comparative brain anatomy, and split-brain experiments with a detailed functional decomposition of the language capacity to form a coherent thread that in turn may inspire new directions of inquiry. To me this is what it means to integrate linguistics with cognitive science, and being able to put these papers together was the most pleasurable aspect of reading TLR.

10. Wrapping up

In this paper, I have defended FL's approach to integrating linguistics with cognitive science, and I have been rather critical of the attempts in TLR to do, more critical than I had hoped. There is a common theme among my criticisms. In nearly every case, the problem lies in a framework too narrow to accommodate the full scope of linguistic phenomena – a focus on semantics to the detriment

of phonology or vice versa, a focus on syntax to the detriment of both phonology and semantics, a focus on high-level generalization to the detriment of irregularity or vice versa, a focus on neural implementation to the detriment of formal description or vice versa, and so on. Many of the schools of thought represented here show no serious engagement with each other's work. As a result, their proposed integration of linguistics with cognitive science may make sense locally but fails to scale up to a globally coherent picture.

I believe that FL succeeds better at integration because it aspires to incorporate every possible perspective on the language faculty: the balance of phonology (and its tiers), syntax, and semantics (and *its* tiers); the role of the lexicon; the relation of all of these to Universal Grammar, processing, acquisition, and neural implementation; the relation of language to thought; and finally evolutionary considerations – with additional nods in the direction of the social manifestations of language. In addition, it seeks to acknowledge the insights of many distinct research traditions, rather than treating any particular framework as all right or all wrong.

At the time I wrote FL in 1999–2000, I had the sinking feeling that this was quite possibly the last time such a comprehensive undertaking could be attempted. The universe of cognitive science has been expanding so fast and in so many dimensions that the full frontier is fast receding from any single person's viewpoint. By now, less than a decade later, the situation is incomparably more difficult. TLR is therefore valuable because it brings together so many points of view in a single place. I hope that this response to it can in part play the role of a simulated conversation among its authors, and that it can stimulate *actual* conversation among them and among their respective research communities. It's hard to resist invoking the hackneyed metaphor of the blind men and the elephant: my goal is to encourage people to explore others' place on the elephant, and even to feel up a hippopotamus or giraffe now and then.

Center for Cognitive Studies, Tufts University

References

- Ackermann, Farrell, and Gert Webelhuth (1999). *A Theory of Predicates*. Stanford: CSLI.
- Baynes, Kathleen, and Michael S. Gazzaniga (2005). Lateralization of language: Toward a biologically based model of language. *The Linguistic Review* 22: 303–326.
- Bickerton, Derek (1990). *Language and Species*. Chicago: University of Chicago Press.
- Bloom, Paul, Mary A. Peterson, Lynn Nadel, and Merrill F. Garrett (eds.) (1996). *Language and Space*. Cambridge, MA: MIT Press.
- Boeckx, Cedric, and Massimo Piattelli-Palmerini (2005). Language as a natural object – linguistics as a natural science. *The Linguistic Review* 22: 447–466.
- Bolinger, Dwight (1965). The atomization of meaning. *Language* 41: 555–573.

- Boroditsky, Lera (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition* 75: 1–28.
- Bybee, Joan, and James L. McClelland (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* 22: 381–410.
- Cheney, Dorothy L., and Robert M. Seyfarth (2005). Constraints and preadaptations in the earliest stages of language evolution. *The Linguistic Review* 22: 135–160.
- Chomsky, Noam (1957). *Syntactic Structures*. The Hague: Mouton.
- (1965). *Aspects of the Theory of Syntax*. Cambridge, MA.: MIT Press.
- (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
- Culicover, Peter W. (1999). *Syntactic Nuts: Hard Cases in Syntax*. Oxford: Oxford University Press.
- (2005). Linguistics, cognitive science, and all that jazz. *The Linguistic Review* 22: 227–248.
- Culicover, Peter W., and Ray Jackendoff (2005). *Simpler Syntax*. Oxford: Oxford University Press.
- Deacon, Terrence W. (1997). *The Symbolic Species*. New York: Norton.
- di Sciullo, Anne-Marie, and Edwin Williams (1987). *On the Definition of Word*. Cambridge, MA.: MIT Press.
- Donald, Merlin (1991). *Origins of the Modern Mind*. Cambridge, MA: Harvard University Press.
- Elman, Jeffrey (1990). Finding structure in time. *Cognitive Science* 14: 179–211.
- Elman, Jeffrey, Elizabeth Bates, Mark Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Feldman, Jerome A. (2006). *From Molecule to Metaphor: A Neural Theory of Language*. Cambridge, MA: MIT Press.
- Ferreira, Fernanda (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47: 164–203.
- (2005). Psycholinguistics, formal grammars, and cognitive science. *The Linguistic Review* 22: 365–380.
- Fitch, W. Tecumseh, Marc D. Hauser, and Noam Chomsky (2005). The evolution of the language faculty: Clarifications and implications. *Cognition* 97: 179–210.
- Fodor, Jerry A. (1983). *Modularity of Mind*. Cambridge, MA.: MIT Press.
- (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
- (1990). *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
- (2000). *The Mind Doesn't Work That Way*. Cambridge, MA: MIT Press.
- Frazier, Lyn, K. Carlson, and Charles Clifton (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences* 10.6: 244–249.
- Gallistel, Charles R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.
- Goldberg, Adele (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- (2005). *Constructions at Work: Constructionist Approaches in Context*. New York: Oxford University Press.
- Goldberg, Adele, and Alex Del Giudice (2005). Subject auxiliary inversion: A natural category. *The Linguistic Review* 22: 411–428.
- Goldberg, Adele, and Ray Jackendoff (2004). The English resultative as a family of constructions. *Language* 80: 532–568.
- Goldin-Meadow, Susan (2005). What language creation in the manual modality tells us about the foundations of language. *The Linguistic Review* 22: 199–226.
- Gross, Steven (2005). The nature of semantics: On Jackendoff's arguments. *The Linguistic Review* 22: 249–270.

- Hagoort, Peter (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences* 9: 416–423.
- Hale, Ken, and Samuel Jay Keyser (2002). *Prolegomena to a Theory of Argument Structure*. Cambridge, MA: MIT Press.
- Hauser, Marc D. (2000). *Wild Minds: What Animals Really Think*. New York: Henry Holt.
- Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298: 1569–1579.
- Higginbotham, James (2003). Jackendoff's conceptualism. *Behavioral and Brain Sciences* 26: 68–681.
- Hockett, Charles F. (1960). The origin of speech. *Scientific American* 203: 88–111.
- Jackendoff, Ray (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.
- (1983). *Semantics and Cognition*. Cambridge, MA: MIT Press.
- (1987). *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press.
- (1988). Why are they saying these things about us? (Topic-Comment column) *Natural Language and Linguistic Theory* 6: 435–442.
- (1990). *Semantic Structures*. Cambridge, MA: MIT Press.
- (1992). *Languages of the Mind*. Cambridge, MA: MIT Press.
- (1994). *Patterns in the Mind*. New York: Basic Books.
- (1996a). The architecture of the linguistic-spatial interface. In Bloom et al. (1996), 1–30.
- (1996b). How language helps us think. *Pragmatics and Cognition* 4: 1–24. (Slightly different version in Jackendoff (1997), chapter 8).
- (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.
- (2002a). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- (2002b). Review of Jerry Fodor, *The Mind Doesn't Work That Way*. *Language* 78: 164–170.
- (2002c). English particle constructions, the lexicon, and the autonomy of syntax. In *Verb-Particle Explorations*, N. Dehé, R. Jackendoff, A. McIntyre, and S. Urban (eds.), 67–94. Berlin: Mouton de Gruyter.
- (2007a). A parallel architecture perspective on language processing. *Brain Research* 1146: 2–22.
- (2007b). *Language, Consciousness, Culture: Essays on Mental Structure*. Cambridge, MA: MIT Press.
- Jackendoff, Ray, and Fred Lerdahl (2006). The capacity for music: What's special about it? *Cognition* 100: 33–72.
- Jackendoff, Ray, and Steven Pinker (2005). The nature of the language faculty and its implications for the evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition* 97: 211–225.
- Klima, Edward S. (1964). Negation in English. In *The Structure of Language*, J. A. Fodor and J. J. Katz (eds.), 246–323. Englewood Cliffs, N. J.: Prentice-Hall.
- Kuperberg, Gina (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research* 1146: 23–49.
- Ladefoged, Peter (2005). Featureless phonetics. Plenary lecture at LSA annual meeting, Oakland, Calif. Written version: Features and parameters for different purposes. <http://www.linguistics.ucla.edu/people/ladefoged/PLfeaturesParameters.pdf>
- Lakoff, George (1987). *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.
- Lakoff, George, and Claudia Brugman (1987). The semantics of aux-inversion and anaphora constraints. Paper presented at LSA annual meeting, San Francisco, CA.
- Landau, Barbara, and Ray Jackendoff (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences* 16: 217–238.
- Langacker, Ronald (1987). *Foundations of Cognitive Grammar*, vol. I. Stanford: Stanford University Press.

- Lerdahl, Fred, and Ray Jackendoff (1983). *A Generative Grammar of Tonal Music*. Cambridge, MA.: MIT Press.
- Levinson, Stephen C. (2003). *Space in Language and Cognition*. Cambridge: Cambridge University Press.
- Li, Peggy, and Lila R. Gleitman (2002). Turning the tables: Language and spatial reasoning. *Cognition* 83: 265–294.
- Lieberman, Philip (1984). *The Biology and Evolution of Language*. Cambridge, MA: Harvard University Press
- (2005). The pied piper of Cambridge. *The Linguistic Review* 22: 289–302.
- MacDonald, Maryellen C., Neal J. Pearlmutter, and Mark S. Seidenberg (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review* 101: 676–703.
- MacNeilage, Peter F., and Barbara L. Davis (2005). Functional organization of speech across the life span: A critique of generative phonology. *The Linguistic Review* 22: 161–182.
- Marantz, Alec. (2005). Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review* 22: 429–446.
- Marcus, Gary (1998). Rethinking eliminative connectionism. *Cognitive Psychology* 37: 243–82.
- (2001). *The Algebraic Mind*. Cambridge, MA: MIT Press.
- (2004). *The Birth of the Mind*. New York: Basic Books.
- Marr, David (1982). *Vision*. San Francisco: Freeman.
- Mielke, Jeff (2006) Emergent feature theory. *CLS* 41. Chicago: Chicago Linguistic Society.
- Miller, George A., and Noam Chomsky (1963). Finitary models of language users. In *Handbook of Mathematical Psychology*, vol. ii, R. D. Luce, R. R. Bush, and E. Galanter (eds.), 419–492. New York: Wiley.
- Myachykov, Andriy, Russell S. Tomlin, and Michael I. Posner (2005). Attention and empirical studies of grammar. *The Linguistic Review* 22: 347–364.
- Newmeyer, Frederick J. (1998). *Language Form and Language Function*. Cambridge, MA: MIT Press.
- Partee, Barbara, Alice ter Meulen, and Robert E. Wall (1993). *Mathematical Methods in Linguistics*. Dordrecht: Kluwer.
- Patel, Aniruddh D. (2003). Language, music, syntax, and the brain. *Nature Neuroscience* 6: 674–681.
- Phillips, Colin, and Ellen Lau (2004). Foundational issues (review article on Jackendoff 2002a). *Journal of Linguistics* 40: 1–21.
- Pinker, Steven (1999). *Words and Rules*. New York: Basic Books.
- Pinker, Steven, and Ray Jackendoff (2005). The faculty of language: What’s special about it? *Cognition* 95: 201–236.
- Piñango, Maria M., Edgar Zurif, and Ray Jackendoff (1999). Real-time processing implications of enriched composition at the syntax–semantics interface. *Journal of Psycholinguistic Research* 28: 395–414.
- Piñango, Maria M., Jennifer Mack, and Ray Jackendoff (2006). Semantic combinatorial processes in argument structure: Evidence from light verbs. *Proceedings of Berkeley Linguistics Society*.
- Poehpel, David, and David Embick (2005). Defining the relation between linguistics and neuroscience. In *Twenty-first century psycholinguistics: Four cornerstones*, A. Cutler (ed.), Hillsdale, NJ: Erlbaum.
- Pollack, Jordan B. (1990). Recursive distributed representations. *Artificial Intelligence* 46: 77–105.
- Port, Robert F., and Adam P. Leary (2005). Against formal phonology. *Language* 81: 927–964.
- Pullum, G. K., and Barbara Scholz (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19: 9–50.
- Pylkkänen, Liina, and Brian McElree (2006). The syntax-semantics interface: On-line composition of sentence meaning. In *Handbook of Psycholinguistics* (2d ed.), M. Traxler and M. A. Gernsbacher (eds.), 537–577. New York: Elsevier.

- Rey, Georges (2006). The intentional inexistence of language – but not cars. In *Contemporary Debates in Cognitive Science*, R. J. Stainton (ed.), 237–255. Malden, MA.: Blackwell.
- Ritter, Nancy (2005a). On the status of linguistics as a cognitive science. *The Linguistic Review* 22: 117–134.
- Ritter, Nancy (ed.) (2005b). Special theme issue: The Role of Linguistics in Cognitive Science. *The Linguistic Review* 22(2–4).
- Rumelhart, David E., and James L. McClelland (1986). On learning the past tense of English verbs. In *Parallel Distributed Processing*, vol. ii, J. McClelland, D. E. Rumelhart, and the PDP Research Group (eds.), 216–71. Cambridge, MA: MIT Press.
- Seyfarth, Robert, and Dorothy Cheney (2003). The structure of social knowledge in monkeys. In *Animal Social Complexity: Intelligence, Culture, and Individualized Societies*, F. de Waal and P. Tyack (eds.), 207–229. Cambridge, MA: Harvard University Press.
- Shastri, Lokanda, and V. Ajjanagadde (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables, and bindings using temporal synchrony. *Behavioral and Brain Sciences* 16: 417–494.
- Slobin, Dan I. (1996). From "thought and language" to "thinking to speaking." In *Rethinking Linguistic Relativity*, J. J. Gumperz and S. C. Levinson (eds.), 70–96. Cambridge: Cambridge University Press.
- Smolensky, Paul, and Géraldine Legendre (2006) *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge, MA: MIT Press.
- Tanenhaus, Michael, M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268: 1632–1634.
- Tomasello, Michael (2005). Beyond formalities: The case of language acquisition. *The Linguistic Review* 22: 183–198.
- van der Velde, Frank, and Marc de Kamps (2006). Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences* 29: 37–70.
- Walenski, Matthew, and Michael T. Ullman (2005). The science of language. *The Linguistic Review* 22: 327–346.
- Wilkins, Wendy K. (2005). Anatomy matters. *The Linguistic Review* 22: 271–288.
- Wilson, Edward O. (1980). *Sociobiology: The Abridged Edition*. Cambridge, MA: Harvard University Press.
- Yip, Moira (2006). The search for phonology in other species. *Trends in Cognitive Sciences* 10: 442–446.
- Zaidel, Eran (1978). Lexical organization in the right hemisphere. In *Cerebral Correlates of Conscious Experience*, P. Buse and A. Rougeul-Buser (eds.), 177–197. Amsterdam: Elsevier.