C:/ITOOLS/WMS/CUP/463980/WORKINGFOLDER/LSO/9780521516457C03.3D 63 [63–72] 1.8.2009 11:19AM

3 Your theory of language evolution depends on your theory of language

Ray Jackendoff

This paper is more about the questions for a theory of language evolution than about the answers. I'd like to ask what there is for a theory of the evolution of language to explain, and I want to show how this depends on what you think language is.

So, what *is* language? Everybody recognizes that language is partly culturally dependent: there is a huge variety of disparate languages in the world, passed down through cultural transmission. If that's all there is to language, a theory of the evolution of language has nothing at all to explain. We need only explain the cultural evolution of languag*es*: English, Dutch, Mandarin, Hausa, etc. are products of cultural history.

However, most readers of the present volume probably subscribe to the contemporary scientific view of language, which goes beneath the cultural differences among languages. It focuses on individual language users and asks:

Q1: Structure and acquisition of language competence) What is the structure of the knowledge that individual language users store in their brains, and how did they come to acquire this knowledge?

The question of acquisition leads to an important corollary question:

Q2: (Structure of capacity to learn language) What is the structure of the knowledge/ ability in the child that makes language acquisition possible?

This latter knowledge is independent of what language the child actually learns in response to the environment. It is closer to what is generally called the language capacity (or the language instinct or universal grammar or the language acquisition device). Because it's prior to learning, it has to be built into the brain by something in the genetic code, plus standard processes of biological development – which of course we don't understand very well. Ultimately, then, a theory of language evolution has to account for the presence of the genetic code that is responsible for the development of the language capacity in every normal human.

Figure 3.1 summarizes this chain of causal and theoretical dependence between the process of evolution and the linguistic competence of native

		Structure of adult knowledge
		↑ ↓
Informational input	→	Learning process
		↑ ↓
		UG / LAD
		↑ ↓
Input	→	Brain structure
		↑ ↓
		Genetic structure
		↑ ↓
		Evolution from ancestral genome

where ↑ = "upper is a result of lower";
↓ = "upper constrains/places theoretical demands on lower"

Figure 3.1. Inferential relation between theory of linguistic knowledge and evolutionary theory

speakers. The structure of adult knowledge of language is a product of a learning process, and therefore the theory of the learning process ultimately has to account for how the adult knowledge is structured. The learning process is based on two sources: (a) information the child takes in from the environment, and (b) the child's prior knowledge about how to structure and generalize that information, coming from the innate universal grammar. Therefore a theory of the learning process constrains the theory of universal grammar: the latter had better be rich enough to account for learning.

In turn, universal grammar is the product (or functional realization) of innate brain structure. Therefore our theory of brain structure has to be rich enough to account for the ability to learn and process language. The language learner's brain structure is a consequence of developmental processes that arise from the interaction of the human genome with environmental input such as nutrition and informational stimulation. Our theory of the genome and its role in the developmental process has to be sufficient to account for the brain structure that supports language learning. And finally, the human genome is a product of evolutionary processes acting on ancestral species, and our account of these evolutionary processes must be sufficient to account for the presence of the brain structure that supports language acquisition. Thus there is a long chain of reasoning between evolutionary processes and modern-day speakers' use of language.

Another dimension of the problem is not addressed by figure 3.1:

Q3: (Special vs. general) What aspects of the language capacity in the mind/brain are special for language, and what aspects make use of more general capacities?

For any aspect of the language capacity, there are four logical possibilities, which for reasons of terminological neutrality I'll call Departments 1, 2, 3, and 4.

Department 1 contains things necessary to language that have required *no* changes from the ancestral genome. Examples would be the lungs and the basic auditory system, which as far as I know are pretty much unchanged from primate prototypes.

Department 2 consists of innovations in the human lineage that are essential to language or language acquisition but that serve purposes more general than language. Examples might be the use of pointing for drawing attention, the capacity for detailed imitation of others' actions, and a fully developed theory of mind (that is, the ability to conceptualize the beliefs and intentions of others, including their beliefs about one's own beliefs and intentions). At least according to my reading of the literature (e.g. Povinelli 2000; Tomasello *et al*. 2005), these are all unique to humans, but not restricted to use in language or language acquisition.

Department 3 contains aspects of language that are unique to humans, that are used exclusively for language or language acquisition, and that resulted from some alteration or specialization of preexisting primate structures or capacities. A clear example would be the shape of the human vocal tract and the neural structures used for controlling it in speech.

Department 4 consists of aspects of language that require something altogether new and unprecedented in the primate lineage.

Roughly, Departments 1 and 2 are what Hauser, Chomsky, and Fitch (2002) call the broad faculty of language; Department 4 is their narrow faculty of language. It is not clear to me where Department 3 falls according to their criteria (see discussion in Jackendoff and Pinker 2005).

This division is crucial for a theory of the evolution of language. Other things being equal, a theory of language that puts more in Departments 1 and 2, less in Department 3, and as little as possible in Department 4 is preferable on evolutionary grounds, because this gives evolution less work to do to get to the present stage.

Given all these dependencies and possibilities, and the near absence of any telling evidence in the archaeological record, how are we to study the evolution of language? To me, the most productive methodology seems to be to engage in reverse engineering:

We attempt to infer the nature of universal grammar and the language acquisition device from the structure of the modern language capacity, using not only evidence from normal language, but also evidence from language deficits, language acquisition, pidgins/creoles (Bickerton 1981; DeGraff 1999), and language creation à la NSL (Kegl, Senghas, and Coppola 1999) and Israeli Bedouin Sign Language (Sandler *et al*. 2005). This is what linguists and psycholinguists normally do.

We can correlate this with comparative evidence from other cognitive capacities to separate the special parts of language from the more general parts. To do this properly, we need to have analyses of other capacities in sufficient detail that we can compare them to language to see what's more general. At present there is little in the way of such analyses. My work with Fred Lerdahl on music (Lerdahl and Jackendoff 1983; Jackendoff and Lerdahl 2006) is the only comparable analysis that I know of; David

Marr's approach to vision (Marr 1982) was on the right track but has been largely abandoned; Jackendoff 2007 offers a sketch of the structure of complex action that may show some promise. This means that cross-capacity comparison, though in principle valuable, cannot be very telling right now.

We can correlate what is known about language and other human capacities with comparative evidence from other organisms, in order to separate what is uniquely human from what is more general across organisms, especially apes, and to see which of the uniquely human aspects of cognition are specializations of preexisting capacities and which are total innovations. This is the program of research advocated by Hauser, Chomsky, and Fitch (2002). To do this properly, we need analyses of animal capacities comparable to our best analyses of language; here I think the situation is even worse than comparing language to other *human* capacities.

After all this comparative work, it might turn out that Departments 3 and 4 are null: that all aspects of language acquisition can be accounted for in terms of more general human cognitive capacities, with no further tuning of existing capacities and no unprecedented innovation. If this is the case, then there's nothing *special* needed for evolution of language, just evolution of more general capacities. And in principle this would be a fine result if we could bring it off. Many people think this is the way it will turn out. For example, those who think language is purely a cultural phenomenon (e.g. Tomasello 2005) are in effect taking this view – though they still have to account for the evolution of human *culture*. Even some linguists, such as Joan Bybee (e.g. Bybee and McClelland 2005) and Adele Goldberg (2005), say "There's nothing innate about language," meaning "There's nothing innate that's *special* for language." Some people – I think Terrence Deacon (1997) and Michael Arbib (2005) might fall into this class – think that the only thing specific to language is the capacity for symbolic use of sound and/or gesture; Departments 3 and 4 are otherwise empty.

Towards a more detailed attempt to sort this out, let's first consider *phonology*. Phonological representations require a digitization of the auditory signal into sequences of discrete speech sounds grouped into syllables. In turn, speech sounds themselves are digitized into a discrete structured space of phonological features. This aspect of phonology is clearly adaptive for language: it is what makes possible a very large number of distinguishable signals – that is, a large vocabulary, an essential feature of human language. We can ask two questions:

Would the acquisition of phonological representations be possible with *no* special capacities in the brain beyond precise vocal imitation (which itself seems fairly special to language and perhaps music)? That is, is phonological structure just a cultural invention (as Arbib claims), or do children come specially prepared to detect digitized vocal signals?

If children do come with such special preparation, are there antecedents elsewhere in human or animal cognition for digitization of sounds either in sequences or in the structuring of the sound repertoire, and if so, how and when did these antecedents arise? There are two different cases to consider: first, in close human relatives, indicating possible homology and little or no evolution in the descent of humans from primate lineage – this looks unlikely. Second, it might be found in distant species such as birds or whales, indicating independent evolution. Here I suppose the jury is out. (See Yip 2006 for a summary of questions and results on animal counterparts of phonology.)

Now let's turn to *syntax* and *semantics*. An important part of the innate language capacity has to be the overall architecture of language, since the child can't be expected to discover whether language has a structure along the lines of, say, the standard theory, the minimalist program, lexical-functional grammar, cognitive grammar, or some correct theory we haven't thought of yet. I want to contrast two sorts of possible architectures for language, which differ particularly in how syntax and semantics are articulated, to see what questions each raises for the evolution of the language capacity. These two architectures are:

1. a *syntactocentric* architecture – the architecture assumed by mainstream generative grammar (e.g. Chomsky 1965, 1981, 1995b);
2. a *parallel* architecture, which is developed in my books *Foundations of Language* (Jackendoff 2002) and *Simpler Syntax* (Culicover and Jackendoff 2005).

Let me take them up briefly in turn.

In a syntactocentric architecture, the generative capacity of language is localized in the syntactic component. Lexical items are associations of phonological, syntactic, and semantic features, and they are embedded into a syntactic structure in the course of derivation. The combinatorial properties of sound and meaning are derived from or read off of different levels or stages or phases of a syntactic derivation. Because everything is read off of syntax, there can be no correspondences between phonology and semantics without some sort of syntactic mediation. Figure 3.2 shows an overall flow diagram for one version of this architecture, Government-Binding Theory of the 1980s and early 1990s (Chomsky 1981).

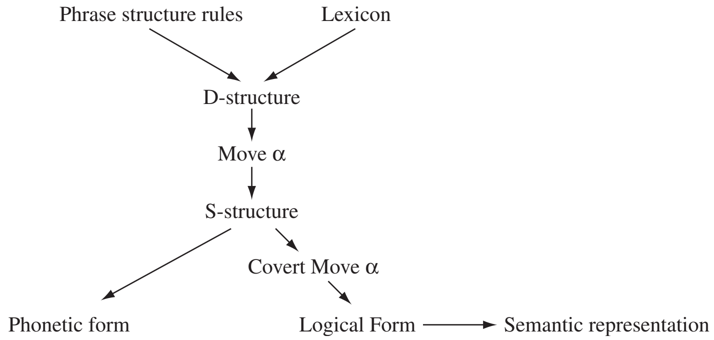

Figure 3.2. Architecture of Government-Binding Theory

A parallel architecture, by contrast, incorporates independent principles of combinatoriality in phonology, syntax, and semantics, each restricted to its proprietary structure. The structures from the three components are linked with each other by *interface rules*, which are of the form "Substructure X in level L1 may correspond to/be linked to substructure Y in level L2." Lexical items are still associations of phonological, syntactic, and semantic features, but in this architecture, these associations are an active part of the interfaces among the three types of structure, rather than being passively manipulated by syntactic derivations. Also among the interface rules may be rules that accomplish direct linkages between phonology and semantics, independent of syntax. Figure 3.3 lays out the organization of the parallel architecture.

Of course, I have a strong preference for the parallel architecture, but I acknowledge that there are numerous proponents of the syntactocentric theory in the world. This is not the place to argue about which theory is correct on grounds internal to linguistics. I've presented my side in fairly gory detail in Jackendoff (2002) and Culicover and Jackendoff (2005). Here I only want to ask:

Q4: What are the implications of these opposing theories for evolution of language?

At least as I understand it, the implication of syntactocentrism is that the combinatorial structure of human thought is derived from (or is a consequence of) the combinatoriality of syntactic structure. Thus syntactic structure is what makes thought/reasoning possible, at least reasoning that involves concepts that cannot be expressed in a single word. In addition, syntax provides the vehicle for connecting thought with vocalization, since it is a necessary link between semantics and phonology. In a way, this is a very Cartesian view, tying together the human capacities for language and reason.

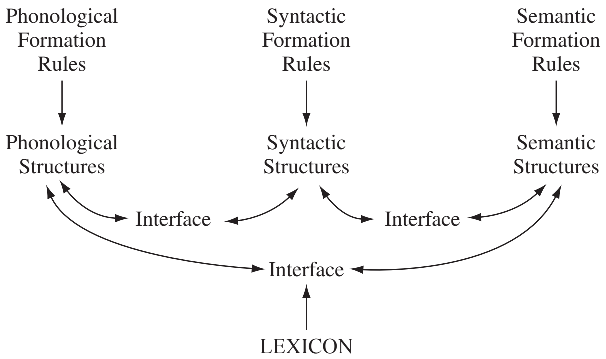

Figure 3.3. The parallel architecture

In terms of evolution, though, it is hard to think of any sort of preadaptation on top of which this sort of language capacity could have been an innovation. The whole generative syntactic system and the mappings to phonetic and logical form have to spring into existence more or less out of the blue – though once they come into existence they're obviously adaptive. This theoretical situation therefore creates a motivation for theorists to try to reduce the evolution of the language capacity to a single step. For example, Hauser, Chomsky, and Fitch (2002), assuming a version of the syntactocentric theory, speculates that all that was needed to get the language capacity up and running was the insertion of recursion into a preexisting system. This preexisting system was either the standard-issue ape cognitive capacity or some more general-purpose cognitive innovation in humans; that is, these aspects of language would belong to Departments 1 or 2. Recursion, in contrast, is for them a uniquely human innovation, possibly *sui generis* (Department 4) or possibly a borrowing from some non-communicative cognitive capacity such as navigation, in which case it would fall into Department 3. Culicover and Jackendoff (2005) and Pinker and Jackendoff (2005) show that recursion is in fact a component of visual cognition, as illustrated by the array in Figure 3.4, which is perceived as built recursively out of elementary x's and o's, assembled into lines of five elements, which are then assembled into nine-element squares. These are arranged into rows of three, and the three rows together form a square which could be further embedded. Thus the answer is that recursion belongs to Department 3.

However, Hauser, Chomsky, and Fitch's proposal does not address the evolutionary source of the lexicon, where all the features specific to particular languages are coded. These features of course include phonological features. As I mentioned earlier, phonological structure is digitized in two respects, and therefore at least on the face of it, phonology differs from the structure of vocal signals in other species, in particular those of our closest relatives. Thus these features seem to belong to Departments 3 or 4, requiring some special innovation in our species.

xxxxx ooooo xxxxx xxxxx ooooo xxxxx xxxxx ooooo xxxxx
ooooo xxxxx ooooo ooooo xxxxx ooooo ooooo xxxxx ooooo
xxxxx ooooo xxxxx xxxxx ooooo xxxxx xxxxx ooooo xxxxx

xxxxx ooooo xxxxx xxxxx ooooo xxxxx xxxxx ooooo xxxxx
ooooo xxxxx ooooo ooooo xxxxx ooooo ooooo xxxxx ooooo
xxxxx ooooo xxxxx xxxxx ooooo xxxxx xxxxx ooooo xxxxx

xxxxx ooooo xxxxx xxxxx ooooo xxxxx xxxxx ooooo xxxxx
ooooo xxxxx ooooo ooooo xxxxx ooooo ooooo xxxxx ooooo
xxxxx ooooo xxxxx xxxxx ooooo xxxxx xxxxx ooooo xxxxx

Figure 3.4. Recursion in cognition of a visual array

Lexical entries also include syntactic features, which determine possibilities for syntactic combinatoriality – feature complexes such as *transitive verb*. Since these are specifically syntactic features, practically by definition they belong to Department 4, since there aren't nouns, verbs, prepositions, tense markers, agreement markers, or case markers in any other cognitive capacity. These syntactic aspects of the lexicon are cognitively useful only if there are syntactic trees to insert lexical items into, so it is hard to imagine why or how they should have evolved prior to the advent of syntax. The other possibility, that syntactic categories and syntactic features emerged automatically with the introduction of recursion, seems equally hard to imagine, since syntactic features are neither properties of the sensorimotor interface nor properties of the conceptual-intentional interface (though some of them are *related* to the latter – see Jackendoff (2002), section 5.9; Culicover and Jackendoff (2005), section 5.6 for my take on this). Of course, I'll be the first to admit that finding something hard to imagine doesn't make it false. On the other hand, it's also hard to imagine how syntax could work without these features, so it looks like they have to join recursion as part of the proposed single step.

I might point out as well that some people, including some linguists, think all these syntactic features either don't exist or arose through cultural innovation (see references above). This would make life better for the syntactocentric view of evolution, but on the other hand most practitioners of the syntactocentric view would fervently reject such a view of syntactic features.

Consider now the parallel architecture. In this architecture, semantic/conceptual structure is the product of a combinatorial capacity that is independent of syntax. This allows the possibility that thought was highly structured in our prelinguistic ancestors – they just couldn't express it. Such combinatorial thought could serve as a crucial preadaptation for the evolution of combinatorial expression: our ancestors had combinatorial thoughts that it would be useful to share. Moreover, the parallel architecture permits correspondences to be established directly between phonology and semantics, without syntactic mediation. It therefore leaves open the possibility of a hominid paleo- or protolexicon, storing associations of pieces of thought to vocal or gestural expression, without a syntactic component at all. Such a mechanism would not necessarily have to be digitized in the modern sense, but it would create the opportunity for evolution to discover the modern digitized form.

Such a system would lend itself to the production of multiple vocalizations to express combinatorial thought. (It is not clear to me whether such a development would require additional mutations, or whether it might be just a consequence of training up working memory.) An elaborate modern syntax is not necessary for this: since such vocalizations are necessarily in some linear order, the opportunity is present for using linear order to crudely express semantic relations, for example agents first, and modifiers adjacent to what they modify.

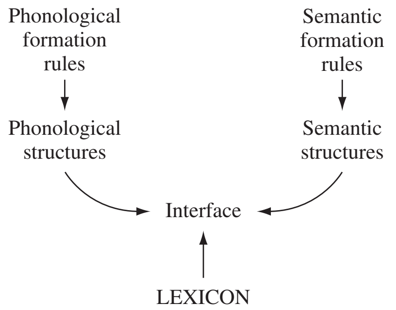

Figure 3.5. Architecture of protolanguage

This sort of organization is arguably found in present-day pidgins, which may lack any hierarchical organization of concatenated words. The system at this stage, roughly Bickerton's (1990) "protolanguage", looks like figure 3.5: the parallel architecture minus the syntactic component and its interfaces.

One can then see syntactic structure as the final capstone innovation, providing a more regimented way to conventionalize semantic relations among words, where the semantic relations and the words are already present in the pre-existing system, unlike in the syntactocentric theory.

In other words, with the parallel architecture, one can imagine various scenarios in which the language capacity evolves in stages, each adding an increment to the system's communicative efficiency and flexibility. There is a fair amount in Departments 3 and 4, but it is motivated by its incremental adaptivity. And the stages I've mentioned here can be articulated into still more gradual stages, as I've proposed in Jackendoff (2002). Some of these stages may be cultural inventions, but others, such as the digitization of phonology, the adaptation of headed hierarchical structure for syntactic purposes, and probably the capacity for dividing words into syntactic categories, seem to require fundamental innovations in the brain's representational capacity.

Thus the two theories of the architecture of language not only have important differences in how they account for the modern language capacity. They also lend themselves to quite different hypotheses about the evolution of language. Again, though, it's worth stressing that these hypotheses are more along the lines of plausibility arguments than evidence for one approach over the other.

What sort of evidence would one seek to help make these stories more than plausibility arguments? One issue I would see as crucial is the degree to which other animals can be shown to exhibit combinatorially structured thought. This need not be anything especially elaborate, such as recursive theory of mind. It just has to be complex enough to require some embedding and to allow some flexibility of behavior. Places I would be inclined to look for such a capacity would be in spatial cognition (as we've seen in figure 3.4), in action planning (Whiten 2002), and especially in social cognition, which involves fairly sophisticated integration of factors such as group membership, dominance, kinship, alliance, and history of reciprocation and conflict (Seyfarth and Cheney 2003). To the extent that combinatorial thought is possible in an organism without language, we might be correspondingly skeptical of the assumption that syntactic generativity is the source of thought.

Another question to ask is whether there are analogues elsewhere in human or animal cognition for the lexicon, which is an extensive store of associations of different types of structure. A possibility in human cognition is our vast knowledge of artifacts and how to use them – everything from buttons and doorknobs to steering wheels, tennis rackets, computer keyboards, and guitars. This involves an integration of visual and motor representations. Should this or some other capacity prove to be analogous to the lexicon, we can then ask if there are principles that permit the combination of its elements into larger complexes of thought and action. If so, are these principles more like the principles of mainstream generative grammar, which derive structures by rules extrinsic to the lexicon, or are they more like those of the parallel architecture, in which lexical items are an integral part of the interface components? Again, if a clear answer emerges, it tells us something about the overall texture of cognitive capacities. To the extent that the texture of the language capacity falls in with the other capacities, its components are at worst in Department 3, rather than in Department 4. And perhaps this other capacity will offer clearer clues about the course of *its* evolution that will shed some light on the evolution of language. You never know.

I note yet again that these are not questions that can be answered from the paleontological record. They are strictly arguments from reverse engineering. Maybe someday, when we understand more fully how genes build brains, we will be able to tighten the argument still further. But for the moment I think the best we can do is set a direction, be clear about the questions, and not succumb to premature confidence or discouragement. What I hope to have shown here is that depending on what your theory of syntax and semantics and rules and lexicon looks like, you're going to seek different sorts of comparative evidence, both across the rest of human cognition and across the animal kingdom.