

CHAPTER 4

Universal Grammar

4.1 The logic of the argument

Over the years, the most controversial aspect of generative grammar has been Chomsky's hypothesis that humans have a cognitive specialization for learning language. This hypothesis is what connects linguistic theory most closely to biology, cognitive development, ethology, and evolutionary psychology. It also has been a main driving force in research on language typology, language acquisition, and linguistic change, not to mention day-to-day research on the structure of language. Hence, whatever controversy surrounds this hypothesis, its importance cannot be overestimated. This chapter will therefore be devoted to a fairly careful exegesis of the Universal Grammar hypothesis, the evidence for it, the arguments against it, and the tensions and challenges it presents to linguistic theory and the other disciplines on which it impinges.

The language learner first enters the scene in *Aspects* in a passage worth quoting and discussing at some length.

[I]t is useful to consider the abstract problem of constructing an "acquisition model" for language, that is, a theory of language learning or grammar construction. Clearly, a child who has learned a language has developed an internal representation of a system of rules that determine how sentences are to be formed, used, and understood. (Chomsky 1965: 25)

Our discussion in sections 2.1 and 3.4 recommended purging intentional vocabulary from the theory. This suggests that Chomsky's phrase "has developed an internal representation of a system of rules" is better expressed as "has internally developed a system of rules." The rules are not *represented* in the learner's mind, they are just there.

Similarly, "learn" has to be handled with care, because of its closeness in meaning to "come to know," an intentional term. In this passage and those to be quoted below, Chomsky speaks of "inventing," "constructing," "developing," "devising," and "acquiring" a grammar, all circumlocutions designed to

get away from the idea that the child's activity is anything like learning facts. Unfortunately, many of these terms still carry overtones of conscious activity, which we certainly do not want to ascribe to the child, given that the grammar itself is unconscious.¹ In other words, we must understand "learn" in the functional sense of "come to have f-knowledge" in the sense of Chapter 2.

Whatever locutions are adopted, though, Chomsky's point is clear: if speakers have a grammar in their f-minds, then it is important to ask how the grammar got there—how children come to acquire it.

[W]e can say that the child has developed and internally represented a generative grammar [better: "has internally developed a generative grammar"—R]. . . . He has done this on the basis of observation of what we may call *primary linguistic data*. (Chomsky 1965: 25)

A footnote explains the notion of primary linguistic data better than the main text:

It seems clear that many children acquire first or second languages quite successfully even though no special care is taken to teach them and no special attention is given to their progress. It also seems apparent that much of the actual speech observed consists of fragments and deviant expressions of a variety of sorts. Thus it seems that a child must have the ability to "invent" a generative grammar that defines well-formedness and assigns interpretations to sentences even though the primary linguistic data that he uses as a basis for this act of theory construction [substitute "f-rule construction"—R] may, from the point of view of the theory ["grammar"] he constructs, be deficient in various respects. (Chomsky 1965: 200-1)

That is, environmental evidence alone is an insufficient basis for the child to f-construct a grammar. This is the essence of the "poverty of the stimulus" argument, to which we will return in section 4.6.

Returning to the main text (with my proposed emendations in brackets):

To learn a language, then, the child must have a method for [f-]devising an appropriate grammar, given primary linguistic data. As a precondition for language learning, he must possess, first, a linguistic theory ["a functional prespecification"] that specifies the form of the grammar of a possible human language, and second, a [n unconscious] strategy for [f-]selecting a grammar of the appropriate form that is compatible with the primary linguistic data. (Chomsky 1965: 25)

The term "Universal Grammar" first appears in *Aspects* in a 1788 quotation from James Beattie, where it is defined as those features that all languages have

¹ This does not preclude the child's consciously *practicing* language, as documented as early as Weir 1962. But practicing constructing sentences is distinct from having conscious access to the rules by which the sentences being practiced are constructed.

in common and that therefore do not need to be mentioned in grammars of particular languages (Chomsky 1965: 5–6). This usage apparently derives from the term *grammaire générale* of the 1660 Cartesian “Port-Royal Grammar.” Soon afterward in Chomsky’s writings (1972b; 1975), “Universal Grammar” comes to be used to denote the “initial state” of the language learner; it thus is conceived of as the aspect of the human mind that *causes* languages to have the features in common that they do. More precisely, Chomsky often uses this term to refer to the child’s initial prespecification of the form of possible human grammars. He uses the term “Language Acquisition Device” to refer to the child’s strategy for constructing or “inventing” a grammar based on primary linguistic data, using Universal Grammar as the starting point. (Alternatively, “Universal Grammar” is sometimes used more loosely to encompass both of these.)²

Another passage further conveys the flavor of what Chomsky has in mind:

[I]t seems reasonable to suppose that a child cannot help constructing a particular sort of transformational grammar to account for the data presented to him, any more than he can control his perception of solid objects or his attention to line and angle. (Chomsky 1965: 59)

This observation is what lies behind Pinker’s (1994b) calling the ability to learn language the “language instinct.” It is part of being human that a child, in response to language in the environment, learns to speak. In a long passage of *Aspects* (pp. 47–52), further developed in Chomsky (1966; 1972a), the idea of an “instinctive” cognitive structure underpinning the acquisition of knowledge is referred back to rationalist forebears including Descartes, Lord Herbert, Cudworth, Arnauld, Leibniz, and particularly Wilhelm von Humboldt.³

Finally, Chomsky brings the issue back to the problems faced by linguists:

As a long-range task for general linguistics, we might set the problem of developing an account of this innate linguistic theory [“innate structure” or “prespecification”] that provides the basis for language learning. (Chomsky 1965: 25)

² Chomsky proposes the simplifying idealization that the child stores up a lot of data and then selects a grammar instantaneously. As far as I know, this idealization has not played any significant role in research. Essentially everyone sees the problem as describing what stage the child is at such-and-such a point, and how the next stage is achieved.

³ A more recent tradition which Chomsky rarely cites is the gestalt psychology of the 1920s and 1930s (e.g. Wertheimer 1923; Köhler 1927; Koffka 1935), which pursued arguments about the innate basis for perception and learning along lines remarkably similar to Chomsky’s (see Macnamara 2000 for discussion). Based on my recollections of discussion with Chomsky, I suspect that he disregarded the gestalt psychologists because their later speculations about brain mechanisms (e.g. Köhler 1940) had brought them into considerable disrepute in American psychology by the 1950s (Lashley 1956). However (as Pim Levelt has pointed out to me), nativist thinking did remain influential in Europe much longer, e.g. in the work of Michotte (1954).

What are the initial assumptions ["functional prespecifications"] concerning the nature of language that the child brings to language learning, and how detailed and specific is the innate schema (the general definition of "grammar") that gradually becomes more explicit and differentiated as the child learns the language? For the present we cannot come at all close to making a hypothesis about innate schemata that is rich, detailed, and specific enough to account for the fact of language acquisition. Consequently, the main task of linguistic theory must be⁴ to develop an account of linguistic universals that, on the one hand, will not be falsified by the actual diversity of languages and, on the other, will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning. (pp. 27-8)

I think it fair to say that the task set in these passages—describing the character of Universal Grammar—lies behind the exuberant flowering of linguistic theory and language acquisition research over the past thirty-five years, including much of the work done by resolute opponents of Universal Grammar, who believe there is no such specialized prespecification. Generative linguists will recognize here the mantra that precedes every exposition of linguistic theory; what they may not be aware of is how widely it is rejected and even reviled, not only in other schools of linguistics (such as Cognitive Grammar) but in the vast reaches of psychology and neuroscience.⁵

I happen to think that the idea of Universal Grammar makes a great deal of sense and deserves the influence it has had. On the other hand, it requires a certain amount of polishing, repair, and retrofitting in order to get it into appropriate shape for the Age of Cognitive Neuroscience. That is the task of this chapter; Pinker (1994b) and Jackendoff (1994) have more extended discussion of many of the points raised here.

4.2 Getting the hypothesis right

In order to get to the substance of the hypothesis, we need first to clear away some common misunderstandings of what it says. The quotes below are generalized from many different sources in the literature, as well as from personal conversations over the years.

- "Chomsky claims that grammar is innate." No. Children have to acquire the grammar of whatever language is present in their environment. Universal

⁴ Notice how within two pages, an account of Universal Grammar has been upgraded from "we might set as a long-range task" to "the main task must be"; the latter is clearly what is intended all along. Rhetorical strategy aside, there is no denying the task's interest and importance.

⁵ See Jackendoff (1988) for some illustrations; the situation has if anything become worse in the meantime.

Grammar is not the grammar of any single language: it is the prespecification in the brain that permits the learning of language to take place. So the grammar-acquiring capacity is what Chomsky claims is innate. If the child is not exposed to language, language will not develop (though see section 4.9.4). Perhaps the term "Universal Grammar" is misleading and Chomsky should have called it "metagrammar" or "the seeds of grammar." But in order to preserve historical continuity, we are more or less forced to stick with this term, whatever incorrect connotations it invites.

The term "innate" also requires comment. For a first approximation, it means "present at birth." However, it is customarily used more broadly to denote a characteristic that appears automatically in the course of an organism's development, whether before or after birth. For instance the number and organization of human teeth, which develop after birth, can be said to be innate. The term is normally contrasted with "acquired" characteristics, which are due to the influence of the environment. It is now widely understood that most characteristics of organisms result from an interaction of innate and environmental influences. The strength of one's muscles depends on exercise and nutrition. But the fact that humans develop muscles in the places they have them is innate. So it is meant to be with speaking a language: the capacity for acquiring a language may well develop in the brain over the first two or three years of life. And the child's actual achievements at speaking and understanding result from the interaction of this capacity with the input in the environment.

- "Chomsky claims that there is a universal, innate Deep Structure." Not quite. As seen in the previous chapter, Deep Structure denotes the level of a syntactic derivation prior to the application of derivational (transformational) rules. The content of this level might or might not be universal.⁶ However, even if Deep Structure is universal, this cannot exhaust the scope of Universal Grammar, which must also prespecify the possibilities for derivational rules (if any) and for constraints in syntax, not to mention many important facets of phonological structure.

On the other hand, in the *Aspects* view, Universal Grammar does specify that there is Deep Structure, even it does not specify the exact content of Deep Structure in any particular language. This is part of the overall form of grammar, one that conditions the sorts of (f-)expectations children will have in trying to make sense of the incomprehensible noises the people around them are making. We return to this in section 4.4.

⁶ However, the incorrect interpretation is encouraged by such statements as "In general, it should be expected that only descriptions concerned with deep structure will have serious import for proposals concerning linguistic universals" (Chomsky 1965: 209-10). There was a brief interest in the late 1960s in the "Universal Base Hypothesis," in which the Deep Structure of all languages was taken to be the same; this idea has been revived in more recent work such as Kayne (1994).

• "Chomsky claims that there is a universal, innate Deep Structure that encodes the meaning of language." This reading of Chomsky was responsible for the powerful appeal of generative grammar to philosophers and psychologists (as well as the general public) in the 1960s: it promised that generative grammar would at last provide a key to meaning, the holy grail of the study of mind. Unfortunately, Chomsky intended no such thing. As just observed, Deep Structure itself is not necessarily universal. In addition, it is not meaning: it is just a level of syntactic structure. The misinterpretation was encouraged by the fact that in the *Aspects* theory, following a proposal by Katz and Postal (1964), Deep Structure was hypothesized to be the level of linguistic structure that, though not meaning itself, *determines* meaning.⁷

Thereby hangs a tale, if I may be permitted a digression. The *Aspects* theory inspired considerable research into whether Deep Structure could be brought still closer to meaning. One of the outcomes was the very popular theory of Case Grammar (Fillmore 1968), in which semantic relations such as Agent, Patient, and Goal were coded directly as "case-markers" in Deep Structure; this formed the basis for much subsequent research in computational linguistics and cognitive psychology. More radical was the theory of Generative Semantics (McCawley 1968; Lakoff 1971; Postal 1970a), in which Deep Structure was posited to be *identical* with meaning. This led to a startling expansion of the overall size of posited underlying syntactic structures, plus a much more elaborate set of derivational rules and constraints. Outside of Chomsky's immediate circle, Generative Semantics became phenomenally attractive, and with good reason: if it were correct, the holy grail would have been achieved. However, Chomsky (1972a) and his students (e.g. Akmajian 1973; Culicover 1972; Jackendoff 1972) attacked it vigorously; the ensuing intellectual *melée* has with

⁷ This misconstrual is further encouraged by occasional passages like these:

"In place of the terms 'deep structure' and 'surface structure,' one might use the corresponding Humboldtian notions 'inner form' of a sentence and 'outer form' of a sentence. . . . The terms 'depth grammar' and 'surface grammar' are familiar in modern philosophy in something roughly like the sense here intended (cf. Wittgenstein's distinction of '*Tiefengrammatik*' and '*Oberflächengrammatik*,' 1953, p. 168) . . . The distinction between deep and surface structure, in the sense in which these terms are used here, is drawn quite clearly in the Port-Royal *Grammar*." (Chomsky 1965: 198-9)

"The deep structure that expresses the meaning is common to all languages, so it is claimed [by the Port-Royal grammarians], being a simple reflection of the forms of thought." (Chomsky 1966: 35)

One can be fairly certain that the authors cited really did not have a level of syntax in mind when they spoke of "inner" or "deep" form; they meant "meaning." The problem for these earlier grammarians was that, lacking the tools of formal logic and the like, the only way they had to talk about the meaning of a complex sentence was in terms of simpler related sentences. Chomsky chooses to interpret this practice, for better or for worse, as an implicit version of his own theory of syntax.

tion in
mmar-
xposed
aps the
alled it
storical
correct

tion, it
adly to
anism's
id orga-
ate. The
e due to
charac-
nmental
ion. But
nate. So
anguage
And the
he inter-

quite. As
e deriva-
e content
ucture is
must also
its in syn-

ecify that
of Deep
of gram-
in trying
are mak-

general, it
s import for
f interest in
guages was
94).

some justification come to be known as the Linguistic Wars. In the end, for reasons detailed by Newmeyer (1980), Harris (1993), and Huck and Goldsmith (1995), Chomsky emerged victorious—but with a theory, the so-called Extended Standard Theory, in which Deep Structure no longer had the privilege of determining meaning: rather, this role was shared by Deep and Surface Structure. And then he turned his interest away from meaning, to constraints on syntactic derivations (e.g. Chomsky 1973; 1977).

The reaction in the wider community was one of disillusionment, above all at the bad behavior displayed by both sides in the dispute (including the present author). But the disillusionment was intellectual as well: Chomsky's theory turned out not to reveal meaning after all, at least in the sense that had been anticipated. The consequence was that many researchers felt as though they had been seriously misled by linguistics, and they lost all trust in the field.⁸ Many psychologists who had been intrigued with generative grammar and its nativist underpinnings came to reject both. Many philosophers interested in formal theories of meaning turned from Generative Semantics to formal logic (e.g. Montague Grammar, Partee 1975; 1976), with its explicitly apsychological underpinnings. Nor was anyone outside linguistics impressed (if they were even paying attention) when some years later, Chomsky (1981) proposed a new level of syntax, Logical Form, that again was supposed to determine meaning. They had all been there before. In short, this painful episode was an important factor in the alienation of linguistics from the rest of cognitive science.

4.3 Linguistic universals

Returning to our main theme, the claims of the hypothesis of Universal Grammar, let me deal with some common questions about linguistic universals. As mentioned above, the term "Universal Grammar" is sometimes used interchangeably with "linguistic universals." This suggests that Universal Grammar is to be found uniformly in the structure of all languages, leading to the following sorts of questions:

⁸ This turn away from linguistics is documented in the following quote from Walter Kintsch: "... the work on syntax was the showpiece of psycholinguistics. Linguists were going to tell psychologists what to look for, and psychologists were going to find it sooner or later.

"But this linguistics-gone-psychological disappeared from the scene almost as fast as it came. This was partly because the psychologists wanted to share the fun of theorizing and partly because many linguists between MIT and Berkeley kept changing their minds and contradicting each other. But the chief reason was that the linguists talked syntax and, from a psychological point of view, there seemed to be so many more significant, more pressing problems [*i.e.* *meaning—R*] about which our linguistic mentors had much less to say." (Kintsch 1984: 112)

In conversation, other prominent researchers have reported similar experiences.

id, for rea-
Goldsmith
so-called
ie privilege
id Surface
straints on

bove all at
the present
y's theory
t had been
gh they had
eld.⁸ Many
its nativist
formal the-
logic (e.g.
ychological
y were even
a new level
ining. They
rtant factor

f Universal
: universals.
used inter-
al Grammar
the follow-

Valter Kintsch:
ing to tell psy-

ast as it came,
partly because
ing each other:
point of view,
ng—RJ] about

• “How can Universal Grammar claim to be universal, when (at least at the beginning) it was applied only to English?” Answer: In fact, Chomsky’s very earliest work (1951) was on Modern Hebrew; other early work in syntax concerned German (Lees 1960; Bierwisch 1963), Turkish (Lees 1960), Latin (Lakoff 1968), Japanese (Kuroda 1965), and the Native American languages Hidatsa (Matthews 1964) and Mohawk (Postal 1962). The number of languages now studied is vastly larger. Generative phonology from the start embraced a wide range of languages: the “language index” in Chomsky and Halle’s (1968) *Sound Pattern of English* lists references to over 100 languages.

Still, the syntactic machinery most heavily investigated in Chomskyan versions of generative grammar does indeed betray a pedigree grounded in the study of English-like languages. In reaction, other generative theories, notably Lexical-Functional Grammar (Bresnan 1982a; 2001), Autolexical Syntax (Sadock 1991), and Role and Reference Grammar (Van Valin and LaPolla 1997), have developed syntactic machinery more explicitly designed to speak to the varieties of syntactic phenomena in the languages of the world. This leads to the next question:

• “If languages differ so much from each other, how can there be any universals? And if there aren’t, how can Universal Grammar have any content?” Remember, Universal Grammar is not supposed to be what is universal among languages: it is supposed to be the “toolkit” that a human child brings to learning any of the languages of the world. If we find that a certain aspect of linguistic structure is indeed universal, then it is a good candidate for part of Universal Grammar, though other options must also be considered (see section 4.5).

However, non-universal aspects of linguistic structure may be candidates for Universal Grammar as well. When you have a toolkit, you are not obliged to use every tool for every job. Thus we might expect that not every grammatical mechanism provided by Universal Grammar appears in every language. For instance, some languages make heavy use of case marking, and others don’t; some languages make heavy use of fixed word order, and others don’t. We would like to say that Universal Grammar makes both these possibilities available to the child; but only the possibilities actually present in the environment come to realization in the child’s developing grammar.

One prominent version of the “toolkit” approach is Principles and Parameters theory (Chomsky 1981), in which all grammatical variation among languages is localized in a set of universal parameters, whose settings are triggered by environmental input. Learning a language can then be thought of roughly as like customizing the settings in a software package. But there are other, less rigid theories of Universal Grammar as well. In any event, it is commonly understood that Universal Grammar provides possibilities, not just certainties, for the structure of the grammar the child is to develop.

Such an approach is subject to a certain justified skepticism. It is tempting to fall into the trap, upon encountering a new grammatical phenomenon, of positing it as a new component of Universal Grammar—just one that happens to be restricted to this single language. This potentially leads to an inflated estimate of what the child brings to language learning. Although of course researchers do sometimes fall into this trap, in general I think it has been avoided.

Rather, the dialectic goes roughly like this: When one looks at language after language and finds that the tools one has proposed for Universal Grammar are sufficient, one begins to get the feeling that one has the right toolkit. On the other hand, when the same difficulty starts cropping up time after time, one begins to consider proposing a new tool, or revising the tools one has previously proposed. A good example might be the widespread phenomenon of Noun Incorporation, which came to the attention of generative theorists in the middle 1980s. There are many languages in which it is possible to express a direct object by attaching a noun closely to the verb ("incorporating it into the verb"), leaving the modifiers of the noun still expressed in direct object position. (1) is an example from Southern Tiwa, quoted in Sadock 1991; note that 'cat' is part of the verb form, sandwiched between the agreement marker and the verb itself.

- (1) Wisi ibi- musa-tuwi-ban
 two AGR-cat - buy -PAST 'They bought two cats.'

It was clear to everyone that this phenomenon did not lend itself easily to the tools of the then-current theories of syntax, and that some addition had therefore to be made to the toolkit posited by Universal Grammar. The character of the proper mechanism has been subject to lively discussion (Mithun 1984; Baker 1988; Rosen 1989; Sadock 1991; and many others). Is it a new kind of derivational rule, a new kind of lexical formation rule, or a lexical redundancy rule? Does the same tool account for English nouns like *man-eater* and *language learner*, which are built out of a verb and its object? This is not the right place to go into details; the point here is only to give the flavor of the conversation.

At the same time, there is a constant re-evaluation of the inventory of elements posited in Universal Grammar. The goal is to posit the smallest toolkit that can still account for the data. For instance, under the conception of derivational rules in early generative grammar, an account of Noun Incorporation was altogether straightforward; Postal (1964) in fact used Noun Incorporation in Mohawk to construct an important argument for transformational grammar. However, in the interests of constraining the possibilities offered within Universal Grammar, the theory was changed in order to rule out many movement phenomena that had not yet been observed. As it happened, these changes also ruled out the possibility of Noun Incorporation. The problem then faced in

the 1980s was to formulate an account that allowed for the observed phenomena without reintroducing the excessive promiscuity of the earlier hypothesis. More generally, the frequent readjustments in Chomsky's theories over the years have been an attempt to propose as lean a version of Universal Grammar as possible; the Minimalist Program (Chomsky 1995) is by far the leanest.

Some of this constant re-evaluation has been driven by explorations into what forms of grammar are mathematically learnable. In particular, many of the early constraints on transformational rules were in part inspired by learnability considerations (Gold 1967; Wexler and Culicover 1980), which applied specifically to the *Aspects* model. And later research has often averted to learnability considerations as well (Baker and McCarthy 1981; Pinker 1989; Gleitman and Landau 1994; Van Valin 1994; and many others).

What I hope the reader can appreciate from this brief discussion is that hypotheses about the content of Universal Grammar constantly raise complex empirical issues about how a multitude of linguistic phenomena are to be described. Continued examination of more and more linguistic phenomena, with attention not only to what happens but also to what does *not* happen, has led to the many reformulations of linguistic theory over the years, with concomitant rearticulation of the content of Universal Grammar.

4.4 Substantive universals, repertoire of rule types, and architectural universals

Another aspect of the problem of universals is addressed in a section of *Aspects* called "Formal and Substantive Universals" (Chomsky 1965: 27–30), which distinguishes two different facets of Universal Grammar. By "substantive universals," Chomsky means the basic building blocks of linguistic structure: phonological distinctive features and the notion of syllable in phonology, and parts of speech and the notion of syntactic tree in syntax. These parts are used differently in different languages, but one cannot construct a human language without them. Chomsky therefore wishes to attribute them to the brain's prespecification.

By "formal universals," Chomsky means the overall organization of the grammar. These might be divided into two subcategories. First, the child has to have a repertoire of rule types—what kinds of rules a language might have for combining the basic units into complex structures. Chapter 3 distinguished phrasal formation rules, derivational rules, several varieties of constraints, lexical formation rules, lexical redundancy rules, and inheritance hierarchies. All of these rule types have to be in the child's repertoire, ready to be filled with content so they can develop into the rules of the child's very own grammar.

Also in the repertoire of rule types belong particular constraints on derivations that have been proposed over the years. For instance, section 3.2.3 mentioned the Sentential Subject Constraint, which prohibits certain kinds of relation between underlying and surface forms. Insofar as every movement rule (or its equivalent in alternative non-movement theories) in every language obeys this constraint—and children don't violate it in the course of learning language—we would like to be able to say that this constraint and others like it come prespecified in the child's toolkit.

As we observed in Chapter 3, a great deal of theoretical dispute concerns which kinds of phenomenon fit under which kind of rule—and whether certain sorts of rules (in particular derivational rules and inheritance hierarchies) exist at all. These disputes concern specifically what repertoire of rule types should be ascribed to the child's prespecification.

Still more basically, the child needs to f-know in what overall linguistic structures the basic building blocks can be arranged. The basic architectural outline of linguistic structure sketched in Fig. 1.1—interconnected phonological, syntactic, and conceptual/semantic structures, each containing substructures and tiers of particular sorts—is common to all languages; this sort of formal universal might be called an "architectural universal." Languages can differ considerably in how this architecture is realized, but at the largest scale there is little or no deviation.

Some of the major changes in linguistic theory have concerned architectural universals. For instance, the level of Deep Structure proposed in *Aspects* is meant as an architectural universal: it is a syntactic level that is input to both the transformational component and semantic interpretation. The proposed architecture changed when the role of Deep Structure in semantics was altered in subsequent versions of the theory. Chapter 5 will develop a more extensive revision of the overall architecture.

4.5 The balance of linguistic and more general capacities

Michael Tomasello (1995) voices a common complaint with linguists' hypotheses about formal universals, in the context of a critique of Steven Pinker's (1994b) exposition of Universal Grammar.

[T]he list [of innate aspects of language] contains things that no nonlinguist would ever recognize—such things as the projection principle, the empty category principle, the subadjacency constraint, and the coordinate structure constraint. . . . [A]ll of these universals are described in linguistically specific terms such that it is very difficult to relate them to cognition in other psychological domains. (Tomasello 1995: 135–6)

However, Tomasello begs the question: he presupposes that everything innate in language should be explicable in terms of more general psychological phenomena.

One would be laughed at for such a complaint in the case of an undeniably specialized system, say visual stereopsis (perceiving depth from the disparity of images in the two eyes). What makes a system specialized is in part that it performs processes not found elsewhere in the f-mind. If language is indeed a specialized system, one should expect some of its functional principles to be *sui generis*.

Similarly, Andy Clark (1996) wonders how much brain specialization is necessary for language. The formal structure of language provides part of the answer. Using language requires the ability to process syntactic and phonological structures—two structures qualitatively different from anything else in the brain—in the course of mapping between thoughts and external signals. This seems to me to require a brain specialization of some sort, certainly more than “some additional feedback loops” or the “freeing up of some resources” thanks to a larger cortex, as Clark speculates. In particular, Clark asks whether language could be due just to “some small tweak” in brain architecture. I would counter, “Small compared to what?” It’s a small tweak compared to the differences between insects and fish, or to the differentiation of the basic body plan. But then, so is the machinery in the bat’s brain that permits echolocation, or the machinery in the elephant’s brain that permits it to use its trunk. I doubt that Clark would attribute echolocation just to “some additional feedback loops” or the “freeing up of some resources.” As Chomsky (1975) and Pinker (1994b) emphasize, there is no reason to view language any differently.

Part of the trouble is that Universal Grammar has often been construed—by both sides in the dispute—in terms of a cartoonish “grammar box,” cut off from the rest of the mind. This is clearly an oversimplification. Tomasello and Clark are correct to the extent that, insofar as possible, we should be conservative in how much linguistic structure we ascribe to an innate UG. We should welcome explanations of linguistic universals on more general cognitive grounds. For example, if there prove to be hierarchical structures and/or instantiation of variables elsewhere in perception, cognition, and action (as argued in Chapter 3), we then need not ascribe these characteristics to a specifically linguistic “toolkit.” This point is clearly prefigured in *Aspects* (if only in a footnote):

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. (Chomsky 1965: 207)

On the other hand, even granted the broader applicability of hierarchical structure and variable instantiation, general principles alone cannot explain the character of the specifically linguistic hierarchies and linguistic categories: the

ints on deriva-
ion 3.2.3 men-
rtain kinds of
movement rule
anguage obeys
e of learning
id others like it

spute concerns
whether certain
erarchies) exist
types should be

inguistic struc-
tural outline of
gical, syntactic,
res and tiers of
universal might
derably in how
no deviation.
d architectural
d in *Aspects* is
put to both the
roposed archi-
s altered in sub-
tensive revision

ities

guists' hypothe-
Steven Pinker's

guist would ever
ry principle, the
l of these univer-
ult to relate them

anything innate in
cal phenomena.

units out of which they are built (the substantive universals), the repertoire of rule types that govern them, and the overall architecture. Therefore, in order to see what aspects of language *can* be reduced to more general principles of the f-mind, we cannot just observe, for instance, that hierarchical structures are found in motor control (e.g. tool construction), and thereby claim that these are directly related to the hierarchical structures of language (such an argument seems to be offered by Corballis 1991; see the discussion by Bloom 1994a). Rather, it is necessary to develop a functional description of the cognitive structures in question, parallel to that for language exemplified in Fig. 1.1, so we can look for finer-scale commonalities.

I take David Marr (1982) to have been developing such a functional description for the visual system; unfortunately this goal seems to have receded since Marr's death, shortly before publication of his book. Lerdahl and Jackendoff (1983) develop a functional description for musical cognition. Within this system one can see detailed similarities to both vision and language. In particular, the similarities between musical rhythm and linguistic prosody are striking—although neither is reducible to the other. The two are related more or less like fingers and toes.

At a different scale, connectionist neural network modeling (Rumelhart and McClelland 1986a) has suggested some very general principles of f-mental computation and learning that demonstrably extend over many capacities. These are interestingly echoed in Lakoff's (1987) and Jackendoff's (1983) arguments at the level of functional description for the existence of flexible, violable (and therefore nonalgorithmic) constraints involved in many domains of cognition and perception, including language, and they re-emerge in contemporary linguistics to some degree in Optimality Theory. These general principles may be conceived of as formal universals of the f-mind. They undoubtedly constrain and shape formal universals specific to language. But they do not determine the particular content of the linguistic universals, precisely to the extent that language is a distinct specialization.

On still another scale, the one area where a great deal is known about neural instantiation is the low-level visual system (e.g. Hubel and Wiesel 1968). This system and how it maps to higher visual areas or to visuo-motor coordination are sometimes invoked (e.g. Churchland 1986) as a model for how the rest of cognition works. The trouble is that low-level vision shares very little in the way of functional properties with any aspect of language. For example, as pointed out in Chapter 3, low-level vision does not require freely combinatorial structure mediated by typed variables. Hence we cannot expect arguments about low-level vision to carry over very decisively to language.

On the whole, linguists have taken more interest in establishing universals than

repertoire of
e, in order to
iples of the f-
ructures are
hat these are
an argument
om 1994a).
gnitive struc-
.1, so we can

onal descrip-
eced since
d Jackendoff
ithin this sys-
In particular,
re striking—
re or less like

amelhart and
-mental com-
cities. These
3) arguments
violable (and
of cognition
emporary lin-
ciples may be
dly constrain
determine the
tent that lan-

about neural
:l 1968). This
coordination
ow the rest of
ttle in the way
le, as pointed
atorial struc-
ments about

niversals than

in reducing them to more general cognitive capacities, and to this degree Tomasello's complaint is justified.⁹ This is an area where sympathetic cooperation with researchers in other areas of perception and cognition would be extremely helpful: an attempt to find detailed functional parallels to linguistic phenomena is usually beyond the professional competence of either an unaided linguist or an unaided non-linguist.

One other possible source for linguistic universals ought to be mentioned here. It is quite possible that there are some constraints that apply to *any* communicative system operating in a community of organisms, and therefore apply to language. Such "systems" effects might not be explicitly present in any single organism but might arise as "emergent properties" of the community as a whole. An example is the architecture of termite mounds, which arises presumably not because any single termite has a grand plan in *f*-mind, but because each termite is programmed to perform certain far simpler actions which collectively add up to the communal construction of an elaborate structure. I am unaware of any research that demonstrates such effects in aspects of linguistic structure, but it is important to leave the possibility open, especially in the light of interesting work now being done on mathematical and computational modeling of communities of communicating organisms (e.g. Kirby 1998; Steels 1998; Batali 1998; Nowak et al. 2000).

A related line of argument for linguistic universals, which I find less persuasive, appeals to the historical development of languages. For instance, Terrence Deacon says:

Human children appear preadapted to guess the rules of syntax properly, precisely because languages evolve¹⁰ so as to embody in their syntax the most frequently guessed patterns. The brain has co-evolved with respect to language, but languages have done most of the adapting. (Deacon 1997: 122)

But this puts the cart before the horse. Deacon is correct that human languages do not push the envelope of Universal Grammar very much. But our question is: What is this envelope anyway, such that languages, however they evolve over time, must conform to it? Given all the differences among the languages of the world, what is it about them that enables children to "guess the rules of

⁹ A notable exception is Cognitive Grammar (Lakoff 1987; Langacker 1987). However, the psychological principles invoked by Cognitive Grammar rarely extend beyond the figure-ground distinction and Rosch's (1978) theory of categorization, so the depth of the reduction is somewhat limited.

¹⁰ Note that "evolve" here refers to historical change in languages, e.g. the changes from Middle English to Modern English. This is quite a different sense from the evolution of the brain, and it takes place over a different timescale.

syntax" so well? This something, whatever it is, is what is meant by Universal Grammar.

4.6 The poverty of the stimulus; the Paradox of Language Acquisition

Aspects continually returns to the assertion that the primary linguistic data available to the language learner underdetermine the choice of grammar, and therefore are insufficient for inducing the grammar without the aid of a specialized Universal Grammar. Chomsky frequently adds assertions along the following lines (a similar passage was quoted earlier):

It is, for the present, impossible to formulate an assumption about initial, innate structure *rich enough* [my italics—RJ] to account for the fact that grammatical [f-]knowledge is attained on the basis of the evidence available to the learner. (Chomsky 1965: 58)

That is, he says, the problem facing us is not to reduce the tools we ascribe to the child: it is to give the child enough tools to do the job. This argument has been used to justify major phases of elaboration in the theory of Universal Grammar.

On the other hand, opponents of Universal Grammar have argued that the child has much more evidence than Chomsky thinks: among other things, special modes of speech by parents ("Motherese") that make linguistic distinctions clearer to the child (Newport et al. 1977; Fernald 1984), understanding of context, including social context (Bruner 1974/5; Bates and MacWhinney 1982), and statistical distribution of phonemic transitions (Saffran et al. 1996) and of word occurrence (Plunkett and Marchman 1991). All these kinds of evidence are indeed available to the child, and they do help. Chomsky makes a telling slip here, when he says (1965: 35), "Real progress in linguistics consists in the discovery that certain features of given languages can be reduced to universal properties of language, and explained in terms of these deeper aspects of linguistic form." He neglects to observe that it is also real progress to show that there is evidence enough in the input for certain features of languages to be *learned*.

What the critics do not demonstrate, however, is that these kinds of evidence alone are enough to vault the child into the exalted realm of structures like Fig. 1.1. Bates and Elman (1996), for instance, argue that learning is much more powerful than previously believed, weakening the case for a highly prespecified Universal Grammar. I agree that learning which makes more effective use of the input certainly helps the child, and it certainly takes some of the load off Universal Grammar. But I do not think it takes *all* the load off. It may allow Universal Grammar to be less rich, but it does not allow UG to be dispensed

c data avail-
nd therefore
ialized Uni-
wing lines (a

, innate struc-
f-knowledge
1965: 58)

scribe to the
ent has been
al Grammar.
ued that the
things, spe-
distinctions
iding of con-
inney 1982),
1996) and of
s of evidence
a telling slip
sts in the dis-
iversal prop-
of linguistic
that there is
learned.

ls of evidence
ures like Fig.
s much more
prespecified
ive use of the
the load off
It may allow
be dispensed

with altogether. (More detailed discussion of this point appears in Gleitman and Wanner 1982 and Shatz 1982.)

To be sure, Bates and Elman end by saying "Even if we assume that a brain . . . contains no innate knowledge at all, we have to make crucial assumptions about the structure of the learning device, its rate and style of learning, and the kinds of input that it 'prefers' to receive." Still, it is my impression that many advocates of such arguments either are not aware of or explicitly wish to deny the complexity of linguistic structure; and a less complex structure naturally requires a less elaborate learning theory. As in Chapter 1, I insist that we cannot adequately assess theories of language learning without understanding the character of what is learned: Fig. 1.1 represents a bare minimum that all linguists (not just unreconstructed Chomskyans) agree upon. Vision seems intuitively simple too, yet no one in cognitive science believes any more that the visual system has a simple structure or that the brain just "learns to see" without any specialized genetic support.

It is useful to put the problem of learning more starkly in terms of what I like to call the Paradox of Language Acquisition: The community of linguists, collaborating over many decades, has so far failed to come up with an adequate description of a speaker's *f*-knowledge of his or her native language. Yet every normal child manages to acquire this *f*-knowledge by the age of ten or so, without reading any linguistics textbooks or going to any conferences. How is it that in some sense every single normal child is smarter than the whole community of linguists?

The answer proposed by the Universal Grammar hypothesis is that the child comes to the task with some *f*-preconceptions of what language is going to be like, and structures linguistic input according to the dictates (or opportunities!) provided by those expectations. By contrast, linguists, using explicit reasoning—and far more data from the world than the child—have a much larger design space in which they must localize the character of grammar. Hence their task is harder than the child's: they constantly come face to face with the *real* poverty of the stimulus. No child has to decide the sorts of issues we have been sketching here: whether there is a separate prosodic tier in phonology; whether or not there are derivational rules; whether such-and-such a phenomenon belongs to derivational rules, lexical formation rules, or interface constraints; and what kind of a rule is responsible for Noun Incorporation. And surely no child has to choose among major architectural alternatives such as GB, LFG, HPSG, Cognitive Grammar, OT, and many other yet-to-be-devised alternatives. The Universal Grammar hypothesis supposes that at some level (we might say intuitively or instinctively), the child *f*-knows the right choices. It may take children a while to sort the phenomena out, but, as stressed in Chapter 2, they all come up with essentially the same solution.

It is worth pointing out that the Paradox of Language Acquisition finds parallels in every cognitive domain. All normal children learn to see, navigate, manipulate objects, and engage in rich social interaction; but we are far from being able to describe the f-mental processes that constitute these abilities, and even farther from being able to specify a simple learning process that leads to these abilities without support from a rather richly specified initial state. There seems no reason why language should be singled out for different treatment.

Nevertheless, sometimes it is objected¹¹ that positing an innate basis for language acquisition is a counsel of despair or resignation, just pushing the problem downstairs into the genes (which we discuss in section 4.8). But one might justifiably have said the same of the theory of gravitation in Newton's time: it postulated an occult, invisible, inexplicable force that physicists are still trying to explain. There is nothing wrong with such a theoretical move if it is done with care and it leads to interesting conclusions. While there is scientific virtue in desiring a minimum of theoretical machinery in psychology as well as physics, this must not be confused with dogmatically insisting on a minimum regardless of what the evidence might be.

Sometimes it is objected that linguists are trying to figure out the grammar consciously, but children do it unconsciously, so they are more intuitive, less hampered by preconceptions. To me this misses the point: we need a way for children to be *less* imaginative than linguists and *more* hampered by preconceptions—in fact hampered by the very same preconceptions as every other child. It is just that children's preconceptions happen to give them the right solutions.

A way for children to be more hampered than adults has been proposed by Newport (1990) and Elman (1993). They suggest, for different reasons, that children may be able to learn language only because their minds are less developed than those of adults. Linguists and adult language learners have developed many more conscious and unconscious strategies for learning, which only serve to lead them astray when they are faced with a foreign language. By contrast, according to this story, children are constrained to a more limited number of possible choices, so they get language right with less trouble. This suggestion may well have some truth to it. However, as pointed out a moment ago, the limited number of possible choices to which children are constrained had better be the *right* ones, otherwise they won't learn. That is, these constraints on their choices amount precisely to Universal Grammar (or, possibly, more general cognitive constraints that children grow out of).

¹¹ The following two objections are voiced explicitly in Hilferly et al. (1998). But they have been raised in many other places in the literature as well and they come up frequently in conversation.

In claiming that children learn grammar so easily by doing it intuitively rather than consciously, critics also miss another important point: in an information-theoretic sense, children and linguists are facing the *very same* problem. I suspect that this is the part of the argument that people find most difficult. So let me try to make it more concrete, to show how it works out in practice.

Suppose we find that in the (hypothetical) language Urtish, certain direct objects are marked dative case and others are marked accusative. The problem for the child as well as for the linguist is to find out which are which. Of course children don't know anything explicit about case marking: "dative," "accusative," and even "case" are just *our* labels. Children only have to put the right affix in the right place, without knowing why, and perhaps even without knowing *that* they have done so. But they do have to get it right.

So we linguists can ask how children manage to do this. One way might be that they simply memorize the right case for each verb, and this may indeed be the right solution; in a theoretical grammar we would call this "lexically marked quirky case." But suppose when we introduce children to new verbs, they automatically use the right case:

(2) *Experimenter*: Look! This is blizzing. Tell me what I just did.

Child: You blizzed the cake-os. [where *-os* is the case-marking in question]

Now we have to discover what clues the child used. Are they semantic, phonological, or pragmatic? If semantic, are they based on the nature of the action, the object acted on, the agent's intention, etc., etc.? The theoretical grammar will have to be formulated to capture this distinction. Or suppose children invariably use accusative case on the first try, even when the correct case is dative. Then the linguist might say accusative is the "default" case, and look for the systematicity or lack thereof behind deviation from the default. Having figured this out, then we can ask: Why did the child look for *that* clue rather than something else? Such questions are familiar to every practicing linguist. When we find these questions arising in every corner of every language (as we do), the Paradox of Language Acquisition becomes almost second nature, a fact of daily research life.

One of my favorite illustrations of the Paradox is the following set of sentences, first pointed out by Jeffrey Gruber (1965).

- (3) a. Every acorn grew into an oak.
 b. Every oak grew out of an acorn.
 c. An oak grew out of every acorn.
 d. *An acorn grew into every oak.

What is striking about this paradigm is its asymmetry. In (3a-c), we understand each acorn to correspond to an oak, that is, *every* quantifies over *an*. But (3d) has

par-
gate,
from
, and
ds to
There
nt.

lan-
prob-
might
me: it
trying
e with
tue in
ysics,
ardless

mmar
e, less
ay for
recon-
other
e right

used by
is, that
, devel-
oped
ly serve
ontrast,
nber of
gestion
the lim-
etter be
on their
ral cog-

have been
ersation.

only the absurd interpretation that a single acorn grew into multiple trees. There would be no problem if (3c, d) were both good: we could say that *every* can quantify over *an* under any circumstance. Again, if (3c, d) were both bad, the story would be simple: we could say that *every* can quantify over *an* only if it precedes it (or is structurally dominant over it). Or, if (3b, d) were both bad, we could say that *every* can quantify over *an* only if *every* is connected to a temporally prior stage (*acorn*), not a temporally subsequent stage (*oak*). But the actual data show some weird interaction of linear order in the sentence and temporal order in the semantics, a totally unexpected result. So far as I know there is no "natural" solution to this in the literature, one that does not in essence simply stipulate this asymmetry.

I also know of no data on the acquisition of this paradigm. But all English speakers (so far as I know) have a grammar that produces these judgments. How did they acquire it? One might propose that sentences like (3a, b, c) are present in the child's primary linguistic input, but sentences like (3d) are not; therefore (3d) is never around to be imitated and hence never comes to be uttered. However, such an approach will not do. We routinely assume that children generalize from the input, so if they hear, say, (4a, b, c), they will be willing to produce (4d).

- (4) a. The cow bit the horse.
 b. The horse bit the cow.
 c. The horse was bitten by the cow.
 d. The cow was bitten by the horse.

So something is preventing children from making the parallel generalization from (3a, b, c) to (3d)—otherwise (3d) would become grammatical within one generation. This something is what linguists so far have failed to discover but children intuitively *f*-know.

Linguists are looking for a "natural" or "elegant" solution to (3) rather than a purely stipulative one precisely so they can ascribe a "natural" or "elegant" structure to the child's intuitions—intuitions that result in our all having the judgments shown in (3). If the eventual "natural" or "elegant" solution turned out to rely in part on non-grammatical aspects of cognition, we would be happier; but we cannot rule out the possibility that it is a peculiarity of the language faculty, hence part of Universal Grammar. We cannot prejudge where to look, and grammatical and non-linguistic solutions call for equally rigorous argumentation.

It is important, however, to handle the poverty of the stimulus argument with care. For instance, a crucial issue in syntax is what permits noun phrases (NPs) to appear in the syntactic positions they do. One popular approach (Chomsky 1981) takes its cue from languages like Latin, German, and Russian, where all NPs are marked for case (nominative, accusative, dative, etc.). The proposal is

2 trees. There
 3ry can quan-
 4ad, the story
 5 it precedes it
 6 could say that
 7 ly prior stage
 8 a show some
 9 in the seman-
 10 l" solution to
 11 s asymmetry.
 12 it all English
 13 3ments. How
 14 are present in
 15 herefore (3d)
 16 ed. However,
 17 neralize from
 18 ice (4d).

2 neralization
 3 al within one
 4 discover but

5) rather than
 6 or "elegant"
 7 ll having the
 8 ution turned
 9 ould be hap-
 10 the language
 11 here to look,
 12 3orous argu-

3 argument with
 4 hrases (NPs)
 5 th (Chomsky
 6 an, where all
 7 e proposal is

that, universally, NPs are permitted in positions where they can be case marked. In order for this proposal to be carried through consistently, it is necessary to claim that English too is relentlessly case marked, even though it makes no overt case distinctions except on pronouns (*Ilme, shelber*, etc.). And even pronouns fail to show an accusative-dative case distinction, needed to distinguish direct from indirect objects. Thus, goes the argument, children learning English acquire the case system even though there is virtually nothing in the linguistic input that tells them about it. Therefore, it is concluded, case marking must be part of UG.

One might find it objectionable that this solution attributes a great deal of invisible structure to English, yet gets this structure in there via innate f-knowledge. In order to answer the argument, though, it is not enough to pronounce it absurd. Two tasks are incumbent on the critic. First, an alternative account of the syntactic positioning of NPs must be offered, one that is either learnable or else based on other plausible principles of UG. Preferably, this account should also be shown to be more empirically adequate in dealing with the distribution of NPs cross-linguistically. Second, an account must be offered of case marking that shows how *it* is learnable in the languages that have it. My impression is that case systems (Blake 1994) cross-linguistically show patterns reminiscent of the cross-linguistic distribution of color names (Berlin and Kay 1969). This suggests that there are at least some innate biases regarding case systems that make them come out as they do. Thus at best the critic might be able to conclude (a) that English speakers do not have an invisible case system, but that the distribution of NPs is determined by some partly innate linguistic principles more "natural" than case; and (b) that the acquisition of case in those languages that have it is guided in part by innate case principles that help the learner structure the primary linguistic input. In other words, perhaps one could arrive at a more input-driven model of the language learner with respect to these phenomena, but it is unlikely that the contribution of UG will go away altogether.

4.7 Poverty of the stimulus in word learning

The discussion so far has been couched largely in terms of the acquisition of grammar, following Chomsky's emphasis. But learning the lexicon poses if anything a far vaster problem. Suppose a grammar has as many as a few hundred rules in it, and compare this to the estimate that an average speaker knows on the order of several tens of thousands of words (this includes passive as well as active vocabulary, since after all one must understand as well as speak the language). By a simple arithmetic calculation, Carey (1978) observes that a child

must learn on the order of five words a day in order to achieve the estimated 8,000-word vocabulary of a six-year-old. Of course this learning is not for the most part deliberate; it is not like learning foreign vocabulary from flash cards. Rather, at any given time the child is probably "working on" various stages of dozens or hundreds of words. The learning of grammar pales by comparison.

It is often taken for granted that word learning is straightforward. Parents point to an object and say the word (*Doggie! See the doggie?*), and the child automatically makes the association. But behind that apparent effortless lies a great deal of complexity, once we think to look for it. Quine's (1960) doctrine of the "indeterminacy of radical translation" was early applied to the problem of word learning: does *doggie* refer to that particular dog (like *Rover* or *Snoopy*), to dogs in general, to (say) poodles in general, to pets, animals, furry things, animate things? Worse, does it refer to the dog's tail, its overall shape, its color, the substance of which it is made, what the dog is doing right now (sitting, panting, slobbering), or (less plausible but still logically possible) the collection of the dog's legs or the combination of the dog and the carpet it is sitting on?

And *dog* is just a concrete count noun. Consider the problems faced when the word is something the parents can't point to, such as *see*, *think*, *hungry*, *ask*, *from*, *any*, *when*, *but*, and *were*, to pick only a few of the many hundreds of non-concrete words a six-year-old knows. How does the child figure out what these words mean? Jerry Fodor (1975; 1998) proposes to solve the learning problem by making all word meanings innate, an extravagance with which few have concurred (see Chapter 11). But even supposing that he were correct, the child would still face the problem of figuring out which of the tens (or hundreds) of thousands of innate meanings is the right one for each of these words. That is, Fodor's move does not evade the poverty of the stimulus argument.

The semantic/conceptual sophistication of the very early word learner is examined at length in John Macnamara's 1982 *Names for Things*, and a vigorous experimental tradition has ensued, attempting to test exactly what assumptions a child makes about the meanings of newly encountered words (Katz et al. 1974; Keil 1989; Bloom 1994b; 1999; 2000; Carey 1994; Markman 1989; Hall 1999; Landau 1994, to list only a few parochially selected choices). This has led to (or has been connected with) fascinating research on how infants conceptualize the physical world and how this changes over the first couple of years of life (Baillargeon 1986; Spelke et al. 1994; Carey and Xu 1999), which in turn has led to parallel experimentation on non-human primates (some examples are cited in Hauser 2000). What is clearly emerging is that the world of the baby is far from William James's "blooming, buzzing confusion" or Quine's undifferentiated quality space. Children apparently come to the task of learning the

the estimated
; is not for the
m flash cards.
ious stages of
comparison.
ward. Parents
and the child
effortlessness
uine's (1960)
applied to the
og (like Rover
pets, animals,
ail, its overall
is doing right
ally possible)
d the carpet it

aced when the
, *hungry*, *ask*,
ndreds of non-
out what these
ning problem
few have con-
rect, the child
- hundreds) of
ords. That is,
nt.

ord learner is
s, and a vigor-
what assump-
rds (Katz et al.
ian 1989; Hall
s). This has led
nts conceptu-
of years of life
ch in turn has
examples are
of the baby is
ine's undiffer-
if learning the

world—and the words used to describe it—with a host of built-in biases that constrain their *f*-hypotheses about what a word can pertain to.

Most of the experimental work cited above pertains to words for objects, with subsidiary emphases on words for substances (*milk*), properties (*heavy*), and locations (*on*, *under*). A whole other area of research, impinging more closely on the learning of grammar, concerns the learning of verbs, which express actions and states. Different kinds of action involve different numbers and types of character. For instance, sleeping and sneezing involve only one character; eating requires an eater and something eaten; giving requires a giver, a recipient, and something given. This has implications for the corresponding verb's syntactic behavior. For instance, *sleep* and *sneeze* are intransitive, their single character being expressed as the verb's subject. *Eat* is transitive, the thing eaten being expressed as direct object—or omitted. *Give* is ditransitive: the recipient and thing given are expressed as indirect object and direct object respectively—or as object of *to* and direct object respectively. (The principles behind these correspondences will be discussed in sections 5.8 and 5.9.) So the question arises: To what extent do children figure out a verb's meaning and then its syntactic possibilities, and to what extent do they have to hear its syntactic possibilities—including various alternatives as in the cases above—in order to decide what it means? A fascinating and intricate discussion in the literature (e.g. Landau and Gleitman 1985; Pinker 1989; 1994a; Fisher et al. 1994; Grimshaw 1994; Brent 1994; Steedman 1994; Tomasello and Merriman 1995; Gillette et al. 1999) detects influences in both directions. However the answer works out, one has to assume that the child has certain biases as to what to look for.

This is only the semantic side of word learning. As Macnamara (1982) points out, the child must also *f*-identify the spoken word as a significant perceptual object, in order to have something to link a meaning to. Here too there is a significant body of research (e.g. Cutler 1994; Jusczyk 1997). Statistical regularities in syllabic structure (Saffran et al. 1996; Kuhl 2000) undoubtedly play a role in the child's determining which strings of sounds are words at all. More generally, the connectionist tradition (Rumelhart and McClelland 1986a, b; Elman et al. 1996) has shown that rather elementary statistical procedures can lead to much more sophisticated behavior and learning than could have been imagined in 1965. As pointed out in Chapter 3, such procedures do not come to terms with the all-important combinatoriality of language. However, we will return in Chapter 6 to points where this sort of learning proves useful.

The discussion in this section can hardly do justice to what has by now become a vast tradition of research. I allude to this tradition only to make clear how rich, varied, and difficult the questions of word learning are. Again we meet

the Paradox of Language Acquisition at every turn. We should be delighted if some aspects of this problem can be reduced to the child doing some relatively simple statistical analysis of the input; but on the other hand, we should not assume that this approach scales up to a solution of the entire problem.

4.8 How Universal Grammar can be related to genetics

In order for Universal Grammar to be an innate cognitive specialization, it must be transmitted genetically, just like anything else innate. But what does it mean for it to be transmitted genetically?

It is certain that the genes cannot directly code a set of functional principles. All they can do is code the synthesis of proteins under particular environmental circumstances, which in turn guide certain unknown parameters of brain growth. Edelman (1992), Elman et al. (1996), and Deacon (1997) provide fascinating discussions of how brain architecture develops and differentiates. But I think it fair to say that the manner in which this process is guided by genetics or anything else is pretty much a mystery at the moment—and this is only at the level of turning genetic instructions into neural architecture. On top of this lies the mystery pointed out in Chapter 2—how neural instantiation supports functional organization, especially at a level as complex as language. So at the moment I think there is really no hope of understanding in any detail the wonderfully indirect mechanisms for genetic transmission of Universal Grammar. In fifty years, perhaps. . . .

Jeffrey Elman et al. (1996) and, following them, Terrence Deacon (1997) mount a series of important arguments against a detailed innate language-learning capacity, based on this problem of genetic transmission. It is worth addressing these arguments in some detail, as they provide a concise distillation of many arguments in opposition to Universal Grammar over the years.

First, they argue that the only way to control behavior of an organism is through adjustment of synaptic weights. But genes cannot adjust synaptic weights—they can code only general guidelines to brain growth. At the same time, however, these authors explicitly say they are nativists, and they are ready to grant complex inborn instincts to animals. It is only UG that they object to. But then, we might ask, how are animal instincts coded on the genes? There is just as much mystery here. As Chomsky says (1965: 206), “Every known species has highly specialized cognitive capacities”; and all the evidence of the past thirty-five years has amplified this statement substantially.

I suggest that the proper questions to ask on this score are the following:

- How can *any* sort of animal behavior—spatial orientation and navigation, bird songs and nest-building, primate call systems, sexual selection, child-rearing,

delighted if
ne relatively
: should not
lem.

tion, it must
does it mean

inciples. All
nmental cir-
rain growth.
scinating dis-
I think it fair
anything else
el of turning
ystery point-
organization,
there is real-
chanisms for

acon (1997)
guage-learn-
orth address-
tion of many

organism is
just synaptic
At the same
hey are ready
ey object to.
nes? There is
nown species
e of the past

ollowing:

d navigation,
child-rearing,

understanding and producing facial expressions, conducting exchanges of reciprocal altruism (e.g. Darwin 1872; Tinbergen 1951/89; Dawkins 1989; Gallistel 1990)—be composed of innate and learned components? There are many differences among species in exactly what is learned and what is innate, as has been established particularly for bird song (Marler 1984).

- How can the genome code any such innate component of animal behavior so as to guide brain development appropriately? Consider something as simple as sneezing, which I assume we can agree is innate—or the ability of newborn horses to get up and walk immediately. How does the brain code these action patterns, and how do the genes make the brain develop so as to code them?
- To what extent is the human ability to learn language guided by such an innate component? Of these three questions, this is the only one at issue at the moment. Putting the question in terms of “to what extent” permits a wide spectrum of possible answers, rather than a binary decision for or against Universal Grammar.

Sometimes it is argued that innate capacities in animals turn out to be “cheap tricks.” For instance, Konrad Lorenz (1952; 1966) discovered that male cichlid fish attack not just other males, but anything that happens to have the right kind of red spot; and that newborn geese imprint not just on their mother, but anything that happens to move the right way. These tricks happen to work well enough in the normal environment of these animals. It is only in the context of perverse experimenters that nature’s short cuts are uncovered. The critics of Universal Grammar sometimes suggest that it too will turn out to be a collection of “cheap tricks,” rather than a detailed specification of tree structures, constraints on rules, and the like. Still, even a cheap trick is a cognitive specialization—and the genome has to code it, whether in cichlids, geese, or humans.

Another argument offered by Elman et al. and Deacon concerns individual differences. How can Universal Grammar be uniform among humans, if everyone’s brain is a little different—for instance, if localization of the language areas is not entirely identical? There are two lines of answer. First, as acknowledged in Chapter 2, not everyone’s language is entirely identical—just good enough for both parties in a conversation to be convinced that communication is taking place. These differences might be a consequence of brain differences in the initial state, or of contingencies of learning, or both—no one really knows. Second, this same question can obviously be addressed to animals’ cognitive specializations, but we are not going to argue that these are not innate. So why should language be singled out? I would imagine that differences among individuals’ brains are more or less like differences among their faces: a bit more here, a bit less there, slightly different placement here, slightly different proportions there,

but everything works essentially the same. We don't know how the genes code individual differences in human faces, much less the basic similarities that distinguish human faces from gorilla faces. Yet we have no problem agreeing that faces are partly determined by the genes.

Elman et al. and Deacon also mount an argument against Universal Grammar based on brain plasticity: young children often can recover fairly good language function (low end of normal range) in the face of early damage to the language areas of the left hemisphere (Milner 1974; Vargha-Khadem et al. 1991). This rules out a direct genetic coding of a "language box" fated to be situated in Broca's and Wernicke's areas—or for that matter a "language learning box" situated there.

Again, there are several lines of reply. First, it is not necessary to conceive of Universal Grammar as a "language learning box," isolated physically and computationally from everything else. Section 4.5 discussed the question of balance between Universal Grammar and more general capacities (or even other specializations). Second, as we will gradually see in the course of succeeding chapters, Universal Grammar need not be a single monolithic faculty. Rather, we will come to see it as a collection of smaller components, some of which may be more vulnerable to impairment than others. Third, considering again nonhuman analogies, it might just be important to look at the plasticity of cognitive specializations of other animals, especially primates. Elman et al. do compare the relative plasticity of language to the relative nonplasticity of spatial cognition. But perhaps more data points are called for. We might find similar patterns of recovery in other capacities and we might not, and it might depend heavily on the capacity in question and the species. Given the number of variables involved and the basic mysteries of the gene-to-cognitive capacity connection, I am not prepared to make any prognostications, much less speculate on their significance.

A final argument in this suite concerns the "modularity" of language. Jerry Fodor (1983) argues that many functions of the mind/brain can be treated as "mental organs" or "processing modules." According to Fodor, several factors serve together as criteria for modularity of a mental function: specialized content, automaticity, susceptibility to focal brain damage, and evidence for innateness. We will examine Fodor's notion of modularity more closely in Chapter 7. For the moment, the argument is that language is a module because it exhibits all the other symptoms: it has the specialized content of syntax and phonology; it is automatic in that one cannot help hearing language as language; it is susceptible to aphasias that affect language but not other aspects of mental functioning. Therefore, goes the argument, it is likely to be innate.

Elman et al. and Deacon quite reasonably ask about other exclusively human activities. Reading and driving cars are automatized and have rather specialized

subject matter; at least the former is subject to focal brain damage and childhood impairment. Experts at chess and (in my own case) playing a musical instrument exhibit a great deal of structured and automatized behavior. But we do not believe there are innate specializations for reading, driving, chess, and clarinet-playing. So why single out language as innate?

I think this argument has to be answered first of all by rejecting Fodor's classification of all these characteristics as together symptomatic of a processing module. Any well-practiced and overlearned ability seems to be automatic and to have specialized structure to some degree. So we may grant that reading, driving, chess, and clarinet-playing can all behave online like Fodorian modules. However, the issue of innateness of language does not concern how language is processed online: it concerns how language is acquired. Hence the comparable question to ask about these other abilities is: What must we give learners in advance in order for them to be able to overlearn these abilities, and how much of this learning follows automatically from cognitive capacities they would have anyway? In order to answer this question, we must determine exactly what they have overlearned. We have some idea of how language is structured, how hard it would be to learn from scratch, and how children do actually learn it. We have no comparable analysis for these other activities. And, as stressed in section 4.5, we have no comparable analysis for any of the perceptual, motor, or cognitive capacities in which these could be embedded.

In short, Elman et al. are right to ask why language should be different from other overlearned abilities, many of which are unlikely to have a direct narrow innate basis. However, I think the question is not just rhetorical: before a proper comparison can be made, there is much empirical work to be done on other abilities as well as on language.

A final step is required in the genetic grounding of Universal Grammar. If there is a genetic basis for language learning, not present in apes, where did it come from? The only reasonable possibility is through evolution. Chapter 8 will take up the question of possible evolutionary routes to modern language. For the moment, let me just note that Chomsky points out the balance between learning and evolution (1965: 59): "There is surely no reason today for taking seriously a position that attributes a complex human achievement entirely to months (or at most years) of experience, rather than to millions of years of evolution. . . ." That is, the more properties of language we can attribute to evolution, the easier language acquisition is for the child.

But Chomsky immediately hedges his bets on evolutionary justification of Universal Grammar, and continues: "or to principles of neural organization that may be even more deeply grounded in physical law." Though logically possible, this alternative declines to follow the argument through to its inexorable