FOUNDATIONS OF LANGUAGE

Brain, Meaning, Grammar, Evolution

Ray Jackendoff

OXFORD
UNIVERSITY PRESS

CHAPTER 8

An Evolutionary Perspective on the Architecture

8.1 The dialectic

We now return to an important aspect of the hypothesis that the ability to acquire a language is a human cognitive specialization. As observed in section 4.8, such cognitive specialization must be coded somehow in the genes, which determine, very indirectly, how the brain is built. In order for these genes to come into existence, some evolutionary step is required at some time since humans diverged from our nearest relatives, the chimpanzees, about five million years ago.

If we are to take full responsibility for the hypothesis, then, it is incumbent on us to address the evolution of the language capacity. This chapter suggests some elements of a possible scenario, in part because of its intrinsic interest, in part towards justification of the UG hypothesis, but also in part as a way to further investigate and refine the architecture of the language faculty proposed in Chapters 5 and 6.

A number of factors stand in the way of developing evolutionary arguments concerning language. First of all, it is a running joke that in 1866 the Linguistic Society of Paris expressly prohibited papers on the origins of language. Evidently too many people had made fools of themselves; today's linguists don't want to fall into the same trap. On the other hand, 1866 was only nine years after the publication of *The Origin of Species*, and in recent years our understanding of evolutionary principles in general and of human origins in particular has expanded vastly. Now that evolutionary talk is rampant, plenty of other people are happy to speculate about evolution of language (e.g. Calvin 1990; Corballis 1991; Deacon 1997; Dennett 1991; Donald 1991)—without taking into account much of what is really known about language. So, as Derek Bickerton suggests (Calvin and Bickerton 2000), it is important for linguists to

participate in the conversation, if only to maintain a position in this intellectual niche that is of such commanding interest to the larger scientific public. Some linguists have indeed risen to the challenge, as will be seen below. The present chapter too is offered in this spirit: I am not sure how seriously I want to take it, but as long as there is a debate, it is worth taking part.

Beyond the sociological issues, proposals about language evolution face two major difficulties. One is a question of data. There is no direct evidence for early forms of language until the advent of writing about 5,000 years ago, and by then we are dealing with fully modern language. Languages may change and "evolve" in the sense of *cultural* evolution, but as far as can be determined, this is in the context of a fully *biologically* evolved language capacity. For the prior five million years, we can make only very indirect inferences based on the nature of artifacts such as tools and pictures, and on equivocal hints about the structure of the brain and the vocal tract.

Indeed, the latter have over time proven less telling than originally thought. For instance, one of the early pieces of evidence (Lieberman 1984) concerned the fact that the Neanderthal larynx, like that of apes, is situated much higher in the vocal tract than that of modern humans, a position not conducive to producing the modern human variety of speech sounds. (Darwin pointed out this difference between humans and other primates, but Lieberman actually worked out the acoustics.) On the other hand, Fitch (2000) shows that, although the larynx of monkeys and goats is positioned much like that of Neanderthals, it descends substantially during the animal's vocalizations—to something much closer to the modern human position. There is no reason not to assume the same was true of Neanderthals, in which case certain aspects of Neanderthal acoustics would have more closely approached the modern standard than Lieberman claimed.

Fitch also reviews fossil evidence from brain endocasts (which can reveal hemispheric differences), from fossil hyoid bones (the attachment point for many vocal tract muscles), and from the size of the canal in the base of the skull for the hypoglossal nerve that controls the tongue. Though each of these has been offered as evidence for or against speech in hominids, Fitch concludes that recent results have rendered all this evidence rather equivocal. And even though Lieberman's arguments about the more limited acoustic possibilities of the upper Neanderthal vocal tract (tongue position and so forth) have not been challenged, this still tells us little about whether Neanderthals *spoke*, and, more important, about what they had to say. In short, there is virtually nothing in the paleontological record that can yield strong evidence about when and in which stages the language capacity evolved.

Moreover, although there are numerous systems of animal communication,

none of them has anything like the expressive capacity of human language (Hauser 1996). They consist either of small collections of discrete messages (such as vervet monkey call systems), messages that vary along a very limited number of dimensions (such as honeybee communication about location of food sources), or messages in which greater elaboration of the signal conveys greater intensity or charisma but not a concomitant elaboration of the message (as in bird songs). As observed already by Darwin (1872), most aspects of primate communication have good human analogues in our systems of facial expression, tone of voice, and "body language." Thus there is no comparative basis in other species for most of the distinctive characteristics of language, and in particular no evidence for significant precursors of language in the apes.

Accordingly, the main evidence I will adduce here comes from the structure of language as we see it today; I will look within modern language for traces of its past. This is to some extent a justifiable methodology in evolutionary theory. For instance, there is virtually no fossil evidence for the evolution of the structure of eyes, as soft tissue is only rarely left behind. Therefore the main evidence for evolution is comparative study of the eyes of modern organisms. Of course we do not have comparative studies of language in other species; but in partial compensation we have comparative linguistic typology as a source of hints.

A second major difficulty in thinking about the evolution of the language capacity is internal to linguistic theory. The common view of Universal Grammar treats it as an undecomposable "grammar box," no part of which would be of any use to hominids without all the rest. The syntactocentric perspective in particular presents serious conceptual difficulties to an evolutionary story. Syntax is useless without phonology and semantics, since it generates structures that alone can play no role in communication or thought; so syntax could not have evolved first. But phonology and semantics could not have evolved first, because (in this architecture) they are simply passive handmaidens of syntax.[1]

There therefore has arisen a characteristic dialectic which if anything has hardened over time, as evolutionary arguments about cognition have gained in ascendancy and at the same time generative grammar has retreated from direct connections with performance and brain instantiation. Opponents of UG argue that there couldn't be such a thing as UG, because there is no evolutionary route to arrive at it. Chomsky, in reply, has tended to deny the value of evolutionary argumentation. For instance, section 4.8 cited an allusion in *Aspects* (59) to a

[1] Chomsky (1981) subdivides GB syntax into a number of components or "modules," such as case theory, binding theory, theta-theory, and so on. But these are not candidates for independent evolution either; each is useless in the absence of the others.

possible alternative to natural selection, "principles of neural organization that may be even more deeply grounded in physical law." Perhaps Chomsky's most famous quote about evolutionary argumentation is this one (among several cited in Newmeyer 1998a):

> We know very little about what happens when 10^{10} neurons are crammed into something the size of a basketball, with further conditions imposed by the specific manner in which this system developed over time. It would be a serious error to suppose that all properties, or the interesting properties of the structures that evolved, can be 'explained' in terms of natural selection. (Chomsky 1975: 59)

As Toulmin (1972), Newmeyer (1998a), and Dennett (1995) point out, this is virtually a retreat to mysticism, appealing to the simple increase in brain size plus the convergence of unknown physical principles. We must not discount the possibility that Chomsky is right; but surely it is worth attempting to make use of the tools at our disposal before throwing them away.

Piattelli-Palmerini (1989) argues, along more evolutionarily defensible lines, that language is nothing but a "spandrel" in the sense of Gould and Lewontin (1979).[2] In his scenario, a number of unrelated developments motivated by natural selection coincidentally converged on a brain structure that happened to instantiate UG, which itself was not selected for. A similar hypothesis appears in Toulmin (1972: 459): "the physiological prerequisites of language developed, in proto-human populations, in a manner having nothing whatever to do with their subsequent 'linguistic' expression." Toulmin ends up hoping that "language might then turn out to be the behavioural end-product, not of a unitary and specific 'native capacity' precisely isomorphic with our actual linguistic behaviour, but rather of more generalized capacities" (465). That is, he specifically wishes to deny the UG hypothesis. As Newmeyer (1998a) points out, one cannot both have a specialized eccentric UG, as Piattelli-Palmerini would like, and claim that it is merely a consequence of general capacities, as Toulmin would like.

Chomsky, Piattelli-Palmerini, and Toulmin all are in effect taking the position that UG was not something that natural selection directly shaped—that it is in some way just a fortunate accident. The former two are using this position to answer the critics of UG; Toulmin is using a similar position to deny a special UG. Without further evidence, then, this argument is a standoff.

Pinker and Bloom (1990) argue for a different position: that the communica-

[2] Dennett (1995) observes that Gould and Lewontin's use of the term "spandrel" is not analogous to the architectural sense of the term on which they claim to draw. However, the term has taken on its own life in evolutionary theory, like "Universal Grammar" in linguistics, so I suppose we have to live with it.

tive advantages of human language are just the kind of cognitive phenomenon that natural selection is sensitive to. Therefore our best hypothesis is that language has evolved incrementally in response to natural selection. This is the position I will take here; I will therefore attempt to provide some parts of a plausible evolutionary scenario. Two basic insights contribute to breaking the log-jam posed by Chomsky's version of UG. First, Chapters 5 and 6 have begun to decompose the language capacity into many semi-independent parts. It thus becomes possible to ask to what extent they could have emerged either independently or in sequence. In this respect I will concur with Toulmin's incremental story—with the major difference that the pieces added incrementally are specifically linguistic rather than general-purpose.

The second insight that contributes to breaking the evolutionary impasse posed by a "grammar box" is Derek Bickerton's proposal for two incremental steps in the evolution of language (a similar and somewhat more highly structured proposal is offered by Givón 1995). Although the proposal to be developed here differs from Bickerton's in many respects, he provides an excellent starting point for substantive discussion.

8.2 Bickerton's proposal and auxiliary assumptions

In his book *Language and Species* (1990), Bickerton proposes that the human capacity for language evolved in two stages. His second stage is language as we know it—let's call it "modern language." He calls the first stage "protolanguage"; for a first approximation one can think of it as modern language minus syntax. His hypothesis is that for several million years, hominids spoke only in a protolanguage, and that the development of modern language is perhaps as recent as 50,000 years ago, with the appearance of *Homo sapiens*.

What elevates Bickerton's story above mere speculation is his claim that protolanguage is still present in the modern human brain. It surfaces when modern language is disrupted; examples are pidgin languages (Bickerton 1981), "Genie," the woman isolated from human contact from age three to thirteen (Curtiss 1977), and possibly agrammatic aphasics. It also surfaces in situations when modern language has not developed: on one hand in early child language, and the other hand in the experiments in teaching language to apes (Linden 1974; Savage-Rumbaugh et al. 1998). Thus evolution did not throw a Good Idea away; rather it built on it. (This story is reminiscent of Rod Brooks's (1986) notion of a "subsumption architecture," in which new, more refined systems are added on top of less articulated existing ones.)

Bickerton (1990) still views the development from protolanguage to modern language as a single rather miraculous leap (in Calvin and Bickerton 2000 he

takes a more gradualist position). Nevertheless, his insight is the opening wedge in conceiving of a more graceful incremental evolution of the language capacity. Lewontin (1990: 741), in a reply to Pinker and Bloom, presents the challenge:

> The explanatory reconstruction of the origin of the camera eye by natural selection requires a particular ordering of light receptor and ennervation first, followed by lens, followed by focusing distortion of the lens and iris diaphragm. The reverse order would not work, if every stage was to be an improvement in vision. Is there an unambiguous ordering for the elements of natural language? Did we have to have them all at once, in which case the selective theory is in deep trouble?

I will argue that one actually can reconstruct from modern human language a sequence of distinct innovations over primate calls, some prior to Bickerton's protolanguage, and some later, each of which is an improvement in communicative expressiveness and precision. Like Bickerton, I will look for traces of these stages in degraded forms of modern language, and relate these stages to what apes have been trained to do. But in addition—and I take this to be an important innovation—in some instances I will be able to show, not just that these earlier stages are still present in the brain, but that *their "fossils" are present in the grammar of modern language itself*, offering a new source of evidence on the issue.

The consequence is that it will no longer be meaningful to ask the divisive question, "Does primate P (e.g. Sarah, Washoe, Koko, Kanzi) and did hominid H have language?" We can only ask "What *elements* of a language capacity might primate P have, and what elements might hominid H have had?" If nothing else, opening this room for a middle ground should be a useful contribution to discourse.

I will make a number of assumptions without justification. All are arguable, but they either make little difference or would take us too far afield here.

- I will not be concerned with the question of "what makes humans unique." There seems often to be an impulse to find the single innovation from which flowed everything distinguishing humans from apes, whether it is walking upright, having opposable thumbs, eating more meat, females having continuous sexual receptivity, or something else. All kinds of things make humans unique, just as all kinds of things make every species unique.
- I assume that language arose primarily in the interests of enhancing communication, and only secondarily in the interests of enhancing thought. (See Chapters 9 and 10 and Jackendoff 1996b; 1997a: ch. 8 for my position on the relation of language and thought.)
- I assume that language arose in the vocal-auditory modality, not in the gestural-visual modality, as has been proposed by Corballis (1991) and Givón (1995), among others. This is just a matter of convenience in exposition; a gestural-visual origin would not materially change my story.

- Along with Pinker and Bloom (1990), I assume that the complexity and specialization of language precludes it being simply a natural development from (or spandrel of) increased memory, planning abilities, motor functions, or other more general functions.
- Most importantly, I assume that any increase in expressive power and precision of the communicative system is adaptive, whether for cooperation in hunting, gathering, defense (Pinker and Bloom 1990), gossip, "social grooming," or deception (Dunbar 1998; Power 1998; Worden 1998). I see no reason to champion any particular one of these as *the* driving force behind language; they would all have benefited from increased expression.
- I will not inquire as to the details of how increased expressive power came to spread through a population (I agree with practically everyone that the "Baldwin effect" had something to do with it),[3] nor how the genome and the morphogenesis of the brain accomplished these changes. Accepted practice in evolutionary psychology (e.g. Dawkins 1989; Barkow et al. 1992) generally finds it convenient to ignore these problems; I see no need at the moment to hold myself to a higher standard than the rest of the field.
- I will not be concerned with establishing the absolute timing of the successive innovations in the language capacity. What concerns me here is the logical progression of stages, hence only their relative timing.

Following the lead of Bickerton and many others, I will draw on evidence from child language, late second language acquisition, aphasia, pidgin languages, and ape language experiments. It is of course never clear how relevant such evidence is for evolutionary concerns—in particular, to what degree ontogeny really does recapitulate phylogeny. Nevertheless, this is all the evidence we've got, so we must make the most of it, while recognizing that it should be taken with a grain of salt.

Finally, I take my cue from an important observation of Wolfgang Köhler (1927) in connection with his studies of animal problem-solving: cognitive steps which appear to us altogether natural may decompose into some parts that are natural for another organism and some parts that are very difficult. The evolutionary counterpart of this observation is that no matter how natural and adaptive some aspect of cognition might appear, it is by no means inevitable that evolution should immediately chance upon it. Thus, for instance, I cannot concur with Corballis's (1991) assumption that an organism with hierarchically

[3] The "Baldwin effect": If organisms are capable of learning some task that is important in their environment, natural selection may favor those individuals who, by virtue of genetic variation, happen to have an innate "leg up" on learning the task. The effect is that innate knowledge will gradually develop over generations, making learning easier. (See Dennett 1995, among many others.)

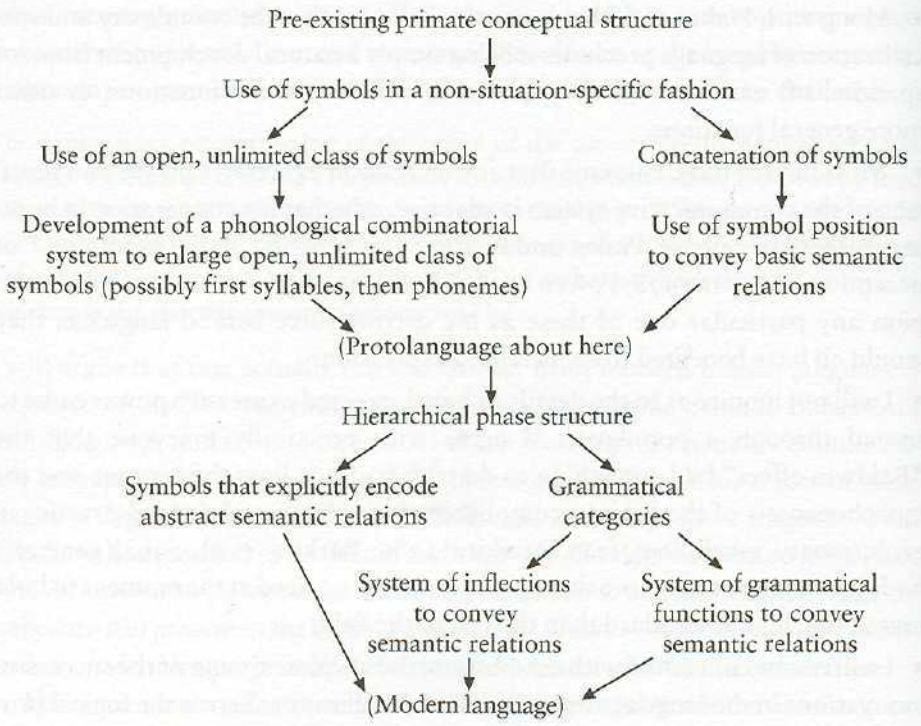

Fig. 8.1. Summary of incremental evolutionary steps

organized behavior is therefore poised to invent syntax (see discussion by Bloom 1994a). Rather, each expansion in the range of the organism's behavior must potentially be regarded as an independent evolutionary step.

The steps I propose are summarized in Fig. 8.1. Logically sequential steps are ordered top to bottom; logically independent steps are presented side by side.

8.3 The use of symbols

The most important preconditions for language are already demonstrably present in primates: there must be a community of individuals who have thoughts worth communicating to each other. I take it as established by decades of primate research (Köhler 1927; Cheney and Seyfarth 1990; de Waal 1996; Tomasello 2000b, among many many others) that chimpanzees have a combinatorial system of conceptual structure in place, adequate to deal with physical problem-solving, with navigation, and above all with rich and subtle social interaction incorporating some sense of the behaviour patterns and perhaps intentions of others.

However, primates are strictly limited in the sorts of information they can communicate.[4] The most important step in getting human language off the ground is the voluntary use of discrete symbolic vocalizations (or other signals such as gestures). Achieving this stage is a major evolutionary step: Deacon (1997), Donald (1991), and Aitchison (1998) are correct in seeing symbol use as *the* most fundamental factor in language evolution. I will not join them in speculating how this ability arose in the hominid line, nor on what precursors had to be present for this ability to evolve. Instead I will concentrate on what had to happen next—on what many researchers shortsightedly view as a straightforward and inevitable development of language from such humble beginnings.

Deacon in particular seems to think that symbols require grammatical combination; he therefore attempts to vault immediately into grammar without any intervening step. However, a single vocalization, as in a one-year-old's single-word utterance, can clearly serve symbolically. I therefore concur with most speculation on the subject in thinking that this initial stage consisted of single-symbol utterances, lacking combinatoriality.

Single-symbol utterances in young children go beyond primate calls in important respects that are crucial in the evolution of language. Perhaps the most important difference is the non-situation-specificity of human words. The word *kitty* may be uttered by a baby to draw attention to a cat, to inquire about the whereabouts of the cat, to summon the cat, to remark that something resembles a cat, and so forth. Other primates' calls do not have this property. A food call is used when food is discovered (or imminently anticipated) but not to suggest that food be sought. A leopard alarm call can report the sighting of a leopard, but cannot be used to ask if anyone has seen a leopard lately (Cheney and Seyfarth 1990; Hauser 1996).

In addition, the child's one-word stage shows considerable conceptual subtlety. For instance, as demonstrated by Macnamara (1982), very young children already appreciate the logical distinction between proper nouns (symbols for tokens—mostly token humans, pets, and places) and common nouns (symbols for types or kinds of any sort). Considerable inquiry has been focused on how children may acquire (or innately have) this aspect of semantics (e.g. Bloom 1999; 2000; Carey and Xu 1999; Hall 1999). Notably, all the famous ape language training experiments of the past three decades seem to have achieved this stage (at least on the more enthusiastic assessments such as Savage-Rumbaugh

[4] Sometimes when I claim that primate thought is in many respects like ours, people ask, "But if apes can think, why can't they talk?" The answer is that they don't have a capacity to acquire phonological and syntactic structures that map thought into linguistic expression. That's exactly what it means to say language is a cognitive specialization, separate from thought.

et al. 1998): non-situation-specific use of a repertoire of single symbols, including both symbols for individuals (proper names) and symbols for categories (common nouns).[5]

However, we can potentially go back further in evolution than the one-word stage: certain little-remarked aspects of modern language are if anything *more* primitive than the child's one-word utterances. Consider the "defective" lexical items mentioned in Chapter 5. These items have no syntax and therefore cannot be integrated into larger syntactic constructions (other than direct quotes and the like). One group of them is associated with sudden high affect, for instance *ouch!*, *dammit!*, *wow!* and *oboy!* These can remain in the repertoire of the deepest aphasics, apparently coming from the right hemisphere (Jackson 1874). Another group includes situation-specific utterances such as *shh*, *psst*, and some uses of *hey* that have almost the flavor of primate alarm calls. Though the *ouch* type and the *shh* type both lack syntax, they have different properties. *Ouch* is often used non-communicatively, but *shh* calls for a hearer; and the *ouch* type are more likely to be uttered involuntarily than the *shh* type, which are usually under conscious control. Also among single-word utterances are the situation-specific greetings *hello* and *goodbye* and the answers *yes* and *no*. The latter are not completely situation-specific: in addition to answering questions, one can use *yes!* to encourage or congratulate the addressee and *no!* as a proto-command for the addressee to cease what (s)he is doing. It is important to notice that no animal call system includes a signal of generalized negation like *no*, which as all parents know is one of the earliest words in child vocabulary.

I would like to think of such words as these as "fossils" of the one-word stage of language evolution—single-word utterances that for some reason are not integrated into the larger combinatorial system. English probably has a few dozen of these—let us not forget exotica such as *abracadabra* and *gadzooks*—and I imagine every language has parallels. Their semantic and pragmatic diversity suggests that they are island remnants of a larger system, superseded by true grammar.[6]

---

[5] It is an interesting question why apes have this capacity, if in fact they do. Why should they be able to acquire and use symbols, despite the fact that it is not something they ever encounter in the wild? I don't know. What I do know is that they don't spontaneously invent symbols, the way deaf children do in inventing "home sign" (section 4.9.4). My inclination is to think symbol use is a "spandrel" for apes, a consequence of other capacities, but that it has been ramped up into a robust specialization in humans, perhaps by the "Baldwin effect."

[6] Just to be very clear: I am not suggesting that the actual "defective" lexical items of English are historical holdovers from this stage of evolution. Rather, what is a holdover is the possibility for a language to contain such "defective items"; those of English are realizations of this possibility.

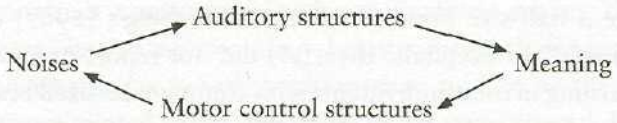

Fig 8.2. Architecture of early single-symbol stage

At this point, then, the system has an architecture like Fig. 8.2. The symbols of the system are long-term memory associations between meanings and auditory/motor codes; we might call them "paleo-lexical items." The arrows in this diagram stand for interfaces of the sort described in Chapter 5.

8.4 Open class of symbols

To go beyond single symbols toward modern language, we need two major innovations. The first is to permit an unlimitedly large class of symbols in the system (a large lexicon); the second is the concatenation of symbols into larger utterances (the beginning of syntax). These two are logically independent: one could have a communicative system involving only one or the other. This is reflected in Fig. 8.1 in the two parallel tracks descending from "use of symbols."[7]

Let's think first about the open vocabulary, the repertoire of meaningful linguistic units stored in long-term memory. By contrast with primate call repertoires (the closest appropriate comparison), which number roughly in the dozens at most, the vocabulary of an average speaker is estimated to run into the tens of thousands. Beginning around the age of two, children learn these in droves, and we keep picking up new words all our lives (Carey 1978). As stressed by Donald (1998), such a large vocabulary places significant demands on long-term memory and rapid retrieval.

The language-trained apes, by contrast, acquire at most several hundred symbols, mostly through extensive training, but in some instances (e.g. Savage-Rumbaugh's bonobo Kanzi) appearing to "just pick them up." At present it is unknown what accounts for the hundredfold difference in vocabulary size. It might be a consequence of the larger human brain, or alternatively of some special human "tuning" that makes vocabulary learning vast and effortless. Some indirect evidence suggests the latter. Reports of children subjected to early hemispherectomy (Vargha-Khadem et al. 1991) do *not* observe massive vocabulary

7 Intuitively, it makes sense that development of an open vocabulary probably went on concurrently with the advent of combinatoriality. I keep them separate so as to make clear the logical independence of the two developments.

deficits despite a half-size brain. Similarly, Lenneberg's (1967) discussion of Seckel Syndrome ("nanocephalic dwarfs") did not report massively impaired vocabulary learning in these individuals with chimpanzee-sized brains. This evidence hints that the effortlessness of vocabulary acquisition is indeed a human cognitive adaptation. Though apes can learn vocabulary, I suspect that this learning is qualitatively different, rather like children's learning of reading: a largely effortful undertaking requiring much motivation and instruction, quite unlike children's spoken (and signed) word learning, which is rapid and spontaneous.

Late second-language learners can be counted on to acquire substantial vocabulary, even when their grammar and especially pronunciation is far from fluent. In the famous case of "Genie" (Curtiss 1977), vocabulary acquisition began immediately upon her discovery, and her rate of vocabulary acquisition approximated that of young children. Yet after years of training, her grammar remained exceedingly rudimentary. Parallel results obtained in a similar case are reported in Curtiss (1994). These well-known facts suggest that the capacity for an open vocabulary is independent of that for grammatical elaboration.

At some point, then, the hominid line had to adapt to learning this vast number of symbols. As Donald (1991) observes, the uniquely human ability to imitate—yet another important cognitive adaptation—obviously plays a role here in the acquisition of the sounds of words. And given the importance of pointing (both by parents and children) in early language acquisition, it is surely significant that apes do not appear to appreciate pointing in the way that human infants do from before one year of age (Povinelli et al. 2000). Imitation and pointing are both important preadaptations for the acquisition of an open vocabulary.

In turn, in order for there to be this vast number of symbols to learn, hominids had to be adapted to be able occasionally to invent *new* symbols, yet another adaptation. It is not clear to me how much metasymbolic capability this would require; the issue requires more investigation, perhaps by looking at the metasymbolic abilities of very young children.

8.5 A generative system for single symbols: proto-phonology

As the class of symbols becomes larger, the perceptual problem arises of making all the utterances discriminable and memorable. If the symbols were holistic vocalizations like primate calls, even a thousand symbols would be impossible to keep distinct in perception and memory (Nowak et al. 1999). Modern language deals with this problem by building words up combinatorially from a repertoire of a few dozen smaller meaningless speech sounds. Using concatenated speech

sounds to construct conventionalized vocalizations makes the distinction among vocalizations a categorical/digital matter rather than a graded one.

Recall Lieberman's (1984) observation that, as late as the Neanderthals (however they are related to us), the shape of the vocal tract did not allow the multitude of easily perceptible distinctions among speech sounds found in modern language (this point is disputed by Aiello 1998). Lieberman, however, points out that an open vocabulary is still possible with a less highly differentiated phonological system. For example, with a repertoire of ten distinct phonemes, one could still construct thousands of words of reasonable length; after all the modern language Hawaiian makes do with only thirteen phonemes. The evolution of the vocal tract—and of the brain machinery that operates it—can be seen as driven by the adaptivity of a larger vocabulary, made possible in part by more rapid articulation and enhanced comprehensibility.

An intermediate stage in evolving a phoneme-based vocabulary might have been based on the syllable rather than the phoneme as the generative unit. The syllable is basically a unit of articulatory gesture, and, as we saw in Chapter 5, the rhythmic organization of language (stress and timing) revolves around the syllable rather than the individual phoneme. Its basic organization is a move from some relatively closed position of the mouth, through a relatively sonorous segment (usually a vowel but occasionally a "syllabic consonant," as in the final syllable of *syllable*), to relative closure again (either the close of the syllable or the beginning of the next). MacNeilage (1998) proposes that this basic form is an adaptation from the mouth's basic close–open–close cycle used for chewing.

The syllable demonstrably plays an important role in speech perception and production. Levelt and Wheeldon (1994) offer psycholinguistic evidence that the repertoire of most frequently used syllables (generally numbering around a few hundred) is stored in what they call a "syllabary"; among other things the syllabary includes a repertoire of motor scripts that aid in rapid articulation.

Clara Levelt (1994) suggests that children around one year of age organize their phonetic articulation in terms of syllables that are not entirely decomposed into independent phonemes. The earliest words tend to have a uniform place of articulation: the mouth opens and shuts but tongue and lip position are held constant. Thus child prefers words such as /tɯ/, in which the tongue tip is near the teeth throughout, or /pom/, where the lips form the major vocal tract constriction throughout. Then the child begins to vary the place of articulation within the syllable, producing things like /pɯ/, where the closure starts at the lips and ends with the tongue on the teeth. It is at this point that we can begin to speak of the child having a real differentiation into phonemes.

In terms of conscious control of speech, children can be taught to count syllables (in my experience) quite readily at three. They cannot be taught to individuate speech sounds until five or six, the age when most children are ready to learn to read; reading alphabetic orthography depends on decomposing words into speech sounds. Even very young children, of course, intuitively appreciate rhyme, which depends on everything in the syllable from the vowel onward. And many cultures have developed syllabic scripts (one character per syllable), whereas by contrast alphabetic script seems to have been invented only once. All these bits of circumstantial evidence point to a certain cognitive primacy to the syllable, despite its being phonetically composite.

Thus we might speculate that the earliest open-ended class of protowords in hominids was composed not from individual speech sounds but, as suggested by MacNeilage, (proto)syllables, each of which was a holistic vocal gesture. A repertoire of ten such gestures could be used to build 100 two-protosyllable vocalizations and 1,000 three-protosyllable vocalizations—well on the way to being open-ended. I imagine that a system of this sort would be possible with the Neanderthal vocal tract. The differentiation of protosyllables into modern syllables analytically composed of phonemes could then be seen as a further step in language evolution; this would make possible a larger and more systematically discriminable class of syllables, in the interests of adding an order of magnitude to the size of the vocabulary. At the same time, the syllable retains some primacy as a phonological unit owing to its longer evolutionary pedigree.

As many linguists (e.g. Hockett 1960; Lieberman 1984; Studdert-Kennedy 1998)—but not many non-linguists—have recognized, the innovation of phonological structure is a major cognitive advance. It requires us to think of the system of vocalizations as *combinatorial*, in that the concatenation of inherently meaningless phonological units leads to an intrinsically unlimited class of words. This is not the fancy recursive generativity of syntax, but, as observed in Chapter 5, it is generativity nonetheless: it is a way of systematizing existing vocabulary items and being able to create new ones, based on the principle of concatenating syllables fairly freely. In turn, syllables are built up from concatenated speech sounds, following fairly strict universal principles of sonority plus arbitrary restrictions and elaborations that differ from language to language.[8] A generative phonological system is thus a crucial step in the evolution of language, necessary for the vocabulary to achieve its presently massive size. (I

[8] It is interesting that the constraints on English syllable structure are violated by some of the single-word English utterances mentioned in the previous section, for instance *shh*, *psst*, *ʔm-hm* ('yes'), *ʔm-ʔm* ('no'), and the apical click of disapproval usually spelled *tsk-tsk*. Perhaps this attests to their primitivity in the linguistic system, "fossils" of the protosyllabic stage.

have not touched at all upon the evolution of other aspects of the phonological system: tone in tone languages, stress, speech rhythm and intonation.)

As mentioned above, child language develops phonological organization very early. By contrast, to my knowledge none of the ape experiments has achieved this step (or even tested it). In the cases where the "language" being taught is visual symbols (lexigrams), each symbol seems to be an unanalyzed visual form. In the cases where sign language was taught, I am not familiar with any evidence that the apes learned the signs in terms of the analytic features of handshape, position, and movement that (as argued by Wilbur 1990 among others) constitute the parallel to syllabic structure in spoken languages.

It should be mentioned, however, that creative concatenation of meaningless elements does appear in the songs of certain bird species, whose repertoire is enlarged by recombination of discriminable song fragments (Hauser 1996; Hultsch et al. 1999; Slater 2000). At the moment the consensus seems to be that no meaning differences accrue from the newly created songs. Rather, larger song repertoires appear to be associated with relative social dominance. Given that the function of this recombination is so different, and given the phylogenetic distance between humans and songbirds, I see no reason to believe there is any inherent link between birdsong and phonology. This is just one of those cases where evolution happened to come up with the same trick on different occasions. Similar concatenative procedures appear to exist in cetacean songs and in possibly some primate "long calls" (Marler 1998; Ujhelyi 1998; Payne 2000); only in the last case is there justification for a possible evolutionary link with human phonology.

8.6 Concatenation of symbols to build larger utterances

A baby's use of single-word utterances is highly context-dependent and must be interpreted in any given situation with a liberal dose of pragmatics. Still, communication *does* take place—a baby's needs are much easier to understand when (s)he has a few dozen words than when there are no words at all. I therefore take it that a communicative system entirely of this sort—where all words behaved grammatically like *hello*—would still be useful to hominids, especially if it had a sizable vocabulary.

One virtue of Bickerton's proposed two-stage evolution of language is in pointing out how one can go beyond single-word utterances without having modern syntax. Much of the rest of this chapter will involve pulling syntax apart, seeking plausible evolutionary steps to the modern state of affairs.

The first essential innovation would be the ability simply to concatenate two or more symbols into a single utterance, with the connection among them dictated

purely by context. For example, *Fred apple* (imagine this uttered by an eighteen-month-old or a signing chimp) might express any number of connections between Fred and apples, expressible in modern language in sentences such as *That's Fred's apple*, *Fred is eating an apple*, *Fred likes apples*, *Take the apple from Fred*, *Give the apple to Fred*, or even *An apple fell on Fred*. Though still vague, then, *Fred apple* is far more precise than just *Fred* or *apple* in isolation. Moreover, it isn't totally vague: it probably wouldn't be used to express *Fred has incorrect beliefs about the color of apples* or *Apples frighten Fred's sister*. That is, although there are many possible connections, the pragmatics are not unlimited.

Concatenating more than two symbols multiplies the number of pragmatic possibilities. Much depends on the symbols in question. *Bread cheese beer* might well express *I want bread, cheese, and beer*. *Bread cheese Fred* is less obvious, *Bread Fred cheese* even less so.

This is clearly a different kind of combination than that discussed in the previous section. Phonological generativity is a way of analyzing meaningful symbols and producing new ones in terms of a repertoire of smaller meaningless units. The present sort of combination puts together meaningful symbols to form larger utterances whose meanings are a function of the meanings of the constituent symbols. The two kinds of combination could have evolved simultaneously or in either order.

This sort of combination has not been attested in the ethological literature. As mentioned above, the units of bird songs, cetacean songs, and primate "long calls" are not meaningful on their own, and/or different combinations are not distinctively meaningful. (As Hauser 1996 points out, however, this may be for lack of means to assess such combinations.) On the other hand, the language-trained apes do show this capability, at least to some degree, on some assessments.

To see if this is where apes' capability stops, it is most revealing to look at the less controlled cases, in which free utterances were possible: the experiments with sign. Terrace (1979) claims that his chimp Nim reached this stage and this stage only, producing large numbers of concatenated (and repeated) signs in an utterance, but without any further organization. He claims that a careful look at the full data from the other signing experiments reveals similar results (see also Seidenberg and Petitto 1978; Ristau and Robbins 1982; Kako 1999), though more enthusiastic researchers have claimed greater organization.

8.7 Using linear position to signal semantic relations

Concatenating symbols opens up many opportunities for enhancing expressive power and precision. Two important classes of innovations are orthogonal:

using the linear order of concatenated symbols to express relations between them, and introducing new sorts of vocabulary item that convey relations explicitly. We take these up in turn.

With just symbol concatenation, *eat apple Fred* and *eat Fred apple* might be used to convey exactly the same message. In this particular case there would be no problem, because of the pragmatics of the words involved. But in *hit tree Fred*, did Fred hit the tree or did the tree hit Fred? Though the larger context might tell us, the pragmatics of the words alone do not. Pinker and Bloom (1990) point out this problem and argue that using principles of word order would be communicatively adaptive.

However, one needn't advance to a full generative syntax, replete with recursive trees, in order to improve the situation. Modern languages display some robust principles that are in some sense prior to syntax, and that reveal themselves more clearly in less fully developed situations. An important piece of evidence comes from Wolfgang Klein and Clive Perdue's (1997) massive longitudinal study of adult second-language learners with various native languages and target languages. The subjects, immigrant workers who "picked up" the target language without explicit instruction, uniformly achieved a stage of linguistic competence that Klein and Perdue call "The Basic Variety" (BV); some, but not all, went beyond this stage in their competence at the new language.

The relevant features of BV are (a) lexical competence; (b) absence of inflectional morphology, e.g. verbs always appear in a fixed form rather than undergoing tense and agreement inflection; (c) omission of contextually supplied arguments, i.e. no obligatory subjects or objects; (d) absence of sentential subordination (no relative clauses, indirect quotes, etc.); (e) simple, largely semantically based principles of word order. The most prominent of these principles are Agent First and Focus Last.[9] So BV is quite far from full linguistic competence. (And, given that many subjects in the study never went beyond this stage, it belies occasional claims that adult second language learning, aside from pronunciation, is usually pretty close to complete).

Agent First and Focus Last are of interest here. A speaker employing Agent First would use *hit tree Fred* to mean only that the tree hit Fred and not that Fred hit the tree; this principle enables one to disambiguate a large proportion of utterances involving two characters. It remains quite powerful in structuring word order in modern language: it appears as the default principle "Agent is expressed

[9] In addition, if there is more than one noun argument in a sentence, the verb regularly falls after the first argument, i.e. there is SVO order. However, this may be an artifact of the target languages in the study, all of which were verb-second in main clauses.

in subject position," which can of course be mitigated by constructions such as the passive (Givón 1995; Van Valin and LaPolla 1997). We saw this principle in section 5.9 as the most robust component of the Argument Linking Hierarchy, and also as one of the principles of constructional meaning in section 6.7.

Agent First seems to be observed as well in pidgin languages (Givón 1995). Piñango (1999; 2000) argues that agrammatic aphasics also fall back on this principle to some degree; this explains some of their errors on reversible passives (*The boy was hit by the girl*), object relatives (*The boy who the girl kissed is tall*), and (in a previously unattested class of errors) certain *because*-clauses (*The girl that drowned because of the boy is tall*). To my knowledge, no one has tried to train an ape in a language that violates this principle, so we don't know whether apes spontaneously observe it or not. (The "home signs" invented by deaf children of non-signing parents (Goldin-Meadow and Mylander 1990) appear to use instead the converse, Agent Last.)

Agent First concerns an element in the system of thematic roles, the specification of who did what to whom (sections 5.8, 5.9, 11.8). By contrast, Focus Last concerns an element in the discourse coding of given and new information—the "information structure" tier (section 12.5). English shows some reflections of Focus Last, for instance in the construction *In the room sat a bear*, where the subject appears at the end for focal effect.

The two principal designated roles in the information structure tier are Focus and Topic. Thus a natural mirror image of Focus Last is Topic First. This is observed in pidgin (Bickerton 1981) and is prominent in the grammatical structure of Japanese. More generally, in many languages of the world, discourse coding plays a far greater role than it does in English; Japanese, Hungarian, and Tagalog are prominent examples (Lambrecht 1994; Van Valin and La Polla 1997). To my knowledge, no one has investigated discourse coding in language-trained apes; I also know of no results from home sign.

Next consider an utterance like *dog brown eat mouse*. Assume this obeys the Agent First principle, so that the dog is doing the eating. There still remains the question of what is brown. It is natural to assume that it's the dog—but notice that this judgment relies on a principle of "Grouping": modifiers tend to be adjacent to what they modify. Although such a principle might follow from general properties of cognition, it is by no means inevitable. Indeed, it can be violated in modern language in constructions like *Bill ate the hot dog naked*. Still, like Agent First, it is a default principle in modern language (Givón 1995; Newmeyer 1998b) and appears in pidgins (Givón 1995) and BV (Wolfgang Klein, p.c.) And like Agent First and Focus Last, Grouping is a purely semantically based principle that maps into linear adjacency without using anything syntactic like a Noun Phrase.

BV is fairly close to what Bickerton (1990) describes as "protolanguage," under which he lumps the organization of pidgins, the grammatical competence attained by Genie, and the achievements of the language-trained apes. His characterization in particular agrees with features (a)–(e) of BV. However, Bickerton attributes to protolanguage a less stable word order than that of BV; this may be partly because his evidence comes from pidgins, which are heavily influenced by the native languages of their speakers.

I suggest, then, that Agent First, Focus Last, and Grouping are "fossil principles" from protolanguage, which modern languages often observe and frequently elaborate. Like the features Bickerton discusses, they often survive in degraded forms of language, which may serve as evidence for their evolutionarily more primitive character. Crucially, these principles correlate linear order with semantic roles. They do not require syntactic structure: the linear order of words can be determined directly in terms of phonological concatenation.

Another possible protolinguistic "fossil" in English is the formation of compound nouns such as *snowman* and *blackboard*. About the only solid principle of meaning in English compounds is that the second word is the "head," the word that denotes the larger category into which the compound noun falls. For instance a snowman is basically a kind of man, not a kind of snow. (And even this is violated in cases like *pickpocket*, which is not a kind of pocket, and *bonehead*, which is not a kind of head but a kind of person.) Within the constraints of this "Head Principle," a wide variety of semantic relations is possible between the nouns, in large part mediated by their meanings. (1) offers a sample (the presence or absence of a space between the nouns is purely an accident of spelling).

(1) a. Locative relations:
doghouse = house for a dog to live in
housedog = dog that lives in a house.
b. Part–whole relations:
wheelchair = a chair with wheels as parts
chairleg = leg that serves as part of a chair
snowman = man made of snow
cake flour = flour that cakes are made of
c. Resemblance relations:
zebra fish = fish that resembles a zebra
d. Actions performed by or on objects:
garbage man = man who carries away garbage
fruit man = man who sells fruit
sun hat = hat that protects against the sun
butter knife = knife used for spreading butter

However, the relation between the nouns is not totally free: while *snowman* might have meant a man who shovels away snow or who makes snow at a ski area, it is not likely to have meant a man whose sister once fell in the snow. Thus the situation resembles the possible meaning relations conveyed by raw concatenation. Jackendoff (in preparation), based on earlier work such as Lees (1960), Levi (1978), Gleitman and Gleitman (1970), and Downing (1977), finds a repertoire of perhaps twenty relations that can be conveyed in a compound through pragmatics alone—though the reason for this particular set of relations remains for the moment unclear. These relations are enriched and constrained by the meanings of the particular words in the compound, along lines suggested by Pustejovsky (1995). It is this reliance on pragmatics and the details of word meaning that has made the analysis of compounds resistant to standard analytic techniques of generative grammar.

Thousands of compounds with partially idiosyncratic meanings are stored in long-term memory. But in addition, one constantly encounters novel examples such as *health management cost containment services* and *two-axle diesel electric engine dump truck* (examples from the Boston *Globe*), whose meanings can be computed on the spot. Thus this is a productive concatenative system involving only words. As observed by, for instance, Sadock (1998), this system is an entirely different sort of combination than other forms of morphology. Klein and Perdue report that noun compounding is the only kind of morphology found in the Basic Variety; and children improvise compounds very early (Clark et al. 1985).

The facts of compounding thus seem symptomatic of protolinguistic "fossils": the grammatical principle involved is simply one of concatenating two nouns into a bigger noun, and the semantic relation between them is determined by a combination of pragmatics and memorization. Determining the meaning of a newly encountered compound is hence much like determining the meaning of *hit tree Fred* discussed above—one uses the Head Principle, plus the repertoire of possible semantic relations, plus a dose of pragmatics, to put together a meaning that makes sense in context.

Whatever the particular details of these sorts of principle that map between semantic roles and pure linear order, they sharpen communication. They are therefore a plausible step between unregulated concatenation and full syntax. In fact, unregulated concatenation need not necessarily have preceded the appearance of these principles: the evidence in modern language is scant, and only possibly the case of Nim shows us raw concatenation without semantically based ordering principles. Notably, the free utterances of the bonobo Kanzi seem to show some limited use of semantically based word order (Savage-Rumbaugh et al. 1998; but see Kako 1999 for a less positive assessment).

The architecture at this point thus looks like Fig. 8.3, in which the new generative system of phonology has been interposed into the mapping between meaning and the auditory/motor levels. This is a parallel generative system without a level of syntax.

The interface between phonology and meaning includes on one hand lexical items and on the other hand principles that map phonological linear order to semantic relations. At the same time, the older interfaces straight from meaning to the auditory/motor systems need not have gone away. In fact they are still used in modern language in tone of voice and perhaps (modulated by phonology) in onomatopoeia.[10]

A final note: Bickerton insists that protolanguage is *not* language, while Klein and Perdue claim that BV *is* language. Yet the two phenomena are incredibly similar. How do they arrive at these opposed positions? Bickerton wants to stress the difference between protolanguage and modern language (in particular creoles), and hence wants to distance protogrammar from Universal Grammar, the essential part of modern language. Klein and Perdue, in contrast, want to stress the role of UG in late second-language acquisition, so they want UG to be involved in BV. Therefore, they tentatively assert that BV represents the default settings of all the parameters in a Principles-and-Parameters type of Universal Grammar. But, as Bierwisch (1997) and Meisel (1997) point out, this either attributes too much sophistication to BV or not enough to UG.

I suggest that Bickerton and Klein and Perdue are each forced into their position because they assume a discrete "grammar box" with a syntactocentric architecture. UG for both of them *is* syntax, and you either have it or you don't. Moreover, they forget that phonology is part of UG too. The present approach allows us to make the appropriate compromise: we can say that protolanguage and BV both have *part* of UG—and approximately the same part.

At the same time, protolanguage/BV is still a long way from the expressive possibilities of modern language. We now progress through some further steps.

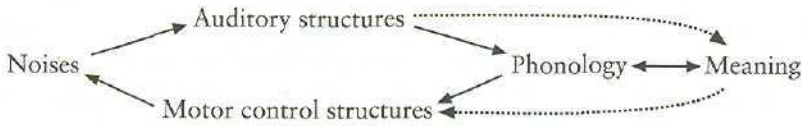

Fig. 8.3. Architecture of protolanguage/Basic Variety

[10] The standard line on onomatopoeia is that it is illusory. After all, dogs go *bow-wow* in English but *gnaf-gnaf* in French, and roosters go *cockadoodledoo* in English but *kikiriki* in German. Still, there is some degree of sound symbolism here. I doubt there's a language where dogs go *kikiriki* and roosters go *thud*.

8.8 Phrase structure

All the phenomena discussed so far use word order to signal semantic relations among *words*; but this is not sufficient for modern language. For example, in the sentence *The little star smashed into a big star*, the entire phrase *the little star* enters into a semantic relation with the verb *smash*. This collection of words functions as an elaborated version of the single word *star*, the head of the phrase. More generally, for purposes of both syntax and semantics, a noun phrase counts as sort of a fancy version of Noun, an adjective phrase counts as a fancy version of Adjective, and so forth. As stressed in Chapters 1 and 5, this is a crucial design feature of modern language, called in the generative tradition "X-Bar theory."

The provision of headed phrases in grammar allows principles of *word order* to be elaborated into principles of *phrase order*. For example, Agent First now applies not to the word that denotes the Agent, but to the *phrase* that denotes the Agent, yielding a major increase in the complexity of conveyable messages: not just *dog chase mouse* but *[big dog with floppy ears and long scraggly tail] chase [little frightened mouse]*. In particular, phrase structure makes possible expressions with hierarchical embedding such as *[the dog [that bit the cat [that chased the rat]]]*—which expresses an equally hierarchical conceptual structure. Such hierarchical embedding in syntax, one of the hallmarks of modern language, is not so simple or inevitable. It does not occur so relentlessly in phonological structure, for example.

Most of the discussion of ape syntax has concerned word order. However, it is not so clear that apes have hierarchical phrase structure. Similarly, discussion of pidgin languages has not made a clear distinction between word order and phrase structure. This distinction thus deserves closer examination.

The potential complexity offered by phrase structure raises new problems of communicability. When there are only three words or so in a sentence, the semantic relations among them can be conveyed by simple word order plus pragmatics. But when sentences get longer and are grouped into phrases, it becomes a pressing issue for the speaker to make the phrase boundaries and the semantic relations among the words more explicit to the hearer. Since the only perceptible signals of abstract phrase structure are linear order and to some extent intonation, language needs further devices in order to make semantic relations explicit.[11]

[11] Kayne (1994) and (following him) the Minimalist Program have taken linear order to be absent from syntax. In this approach, all phrase structure is strictly binary branching, so that the phonological component can read linear order directly off the branching in an inherently unordered tree. Bickerton's most recent work (Calvin and Bickerton 2000) adopts this position as

8.9 Vocabulary for relational concepts

One possible way of encoding semantic relations among words and phrases is to invent words that express them. At the one-word stage, relational words are pointless. But once multiple-symbol utterances are possible, many classes of "utility" vocabulary items offer themselves as design possibilities. In modern language, some are words, some are morphological affixes, and some are realized as variants of word order ("constructions" in the sense of Chapter 6). Here are a few types.

- *Spatial relation terms.* To give someone directions to some spatial location, we don't do a dance like the honeybees. We say "Go up the stream to a tree next to a big rock. Behind the tree and a little to the side you'll see a bush that has great fruit on it." Such description is impossible without all the words that indicate spatial relations: *up*, *to*, *next to*, *behind*, *to the side*, and *on*.[12]
- *Time terms.* These include explicit time terms such as *now*, *yesterday*, and *Tuesday*; temporal relational terms such as *before*, *after*, and *until*; and (once inflection develops) tense and aspect inflection.
- *Marks of illocutionary force and modality.* These differentiate declaratives from questions, commands, and exclamations. They appear in modern language sometimes as variations in word order, sometimes as verbal inflection, sometimes as differences in intonation, and sometimes as a particular word that marks the

well. I take this approach to be profoundly anti-evolutionary. Given that linear order is already present before the advent of phrase structure (and is in any event necessary for discourse!), there is no point in throwing it out of syntactic theory. Rather, syntactic theory should make as much use as possible of linear order, which is after all present in the overt signal. See Culicover (2000) for more extended discussion of this issue.

Carstairs-McCarthy (1999) asks the intriguing question of why so many languages show a major syntactic split between subject and predicate (VP) constituents, where the latter includes (at least) the verb and the direct object. Such a split is not so natural from a logical point of view: after all, first-order logic has no constituent containing the predicate and all but one privileged argument—and neither do computer languages. Carstairs-McCarthy proposes that this asymmetry of subject and predicate—[N [V N]] rather than just [N V N]—is exapted from the asymmetry of the syllable, which (as seen in Chs. 1 and 5) has the structure [C [V C]] rather than [C V C]. Hence, he says, the asymmetry of syntactic structure arose not from the logic of what sentences mean, but rather from the accidental availability of a structure elsewhere in cognition. Although this asymmetry had good acoustic or articulatory reasons in phonology, in syntactic structure it is just one of those accidents of evolution. Whether or not one endorses this argument, I find it does have the right sort of flavor. Another possibility, however, is that subject–predicate structure arose from Topic–Comment organization in information structure. It subsequently became grammaticalized and overlaid with lots of other grammatical phenomena, so that some languages (including English) came to reinvent another pre-subject topic position.

[12] Some of these, for example *up*, are already present in children's vocabulary at the one-word stage. At this point the child probably uses *up* to denote upwardly directed motion, so it is verb-like rather than relational in its semantics.

force of the utterance. A familar case of the last of these is the use in French of *est-ce que* as a fixed formula that converts a declarative sentence into a yes–no question. Perhaps also in this class goes sentential negation, which often seems to get tied up in the tense and question systems, *doesn't it*? We might also include expressions of conditionality such as *if*, *may*, and *can*; these meanings also appear in the tense system, as in the subjunctive and conditional of French.

- *Markers of discourse status*. These include at least the determiners *a* and *the*, which serve to inform the hearer whether the item being mentioned is new to the discourse or identifiable by the hearer from context. At least in English, these are also reliable markers for the beginning of an NP, so they give the hearer help with parsing as well as with keeping characters in a discourse straight.
- *Quantification*. These include the standard logical quantifiers *some*, *all*, and *every*, as well as numerals, expressions like *a lot of* and *oodles of*, and temporal quantifiers like *often* and *always*. A notable case is *more*, which cuts across noun, verb, and adjective modification (*more pudding*, *run more*, *more beautiful*), and which is often acquired by children even at the one-word stage, where what it quantifies must be inferred pragmatically.
- *Purposes, reasons, and intermediate causes*. Compare *You live in this house* and *This house is for you to live in*. The latter can be expressed only if one has a vocabulary item with the meaning of "purpose," here the word *for*. Or compare *I threw the spear and it hit the pig* to *I hit the pig with the spear*. The latter makes explicit my ultimate agency in the pig's fate (while making implicit exactly how I did it). Similarly, compare *He ate the apple and he died* with *He died because he ate the apple*. Only the latter is explicit about the nature of the connection between the two events: one is the reason for the other. Without explicit expressions of reason, one cannot ask *Why?* and therefore seek explanation.[13]
- *More general discourse connectors*. These include words such as *but*, *however*, *therefore*, *moreover*, *what's more*, and *and so forth*.

Each of these classes presents a different challenge to the evolution of the language capacity. Having symbolic utterances or primitive word order or hierarchical structure does not automatically provide any of these classes; nor would organisms that had one class necessarily discover any of the others automatically. The evolution of these possibilities in the language capacity can be speculated about only through the sorts of evidence we have been considering so far: child and adult language acquisition, aphasia, ape experiments, and so on. (Pidgins would be less telling because they draw upon the vocabulary of their source languages.)

[13] Does the famous explosion of *why*s in young children represent their discovery of reasons, as suggested by Kelemen (1999)?

Relational vocabulary plays an important role in thought. It has been argued (Dennett 1991; Jackendoff 1996b, among many others) that language enhances thoughts by making them available as perceptual objects (namely sentences), so that they can be attended to, focused on, modified, and remembered. Upon the invention of this "utility vocabulary," it would all of a sudden be possible explicitly to wonder if *p* and suppose that *p*, and to give reasons and purposes for actions, with a tremendous effect on the power of individual and communal reason and planning. ("What *should* I say to so-and-so? *If* he says this, *then maybe* I'll do that; *but if* . . . ." Try to perform this reasoning without the italicized words.)

Suppose we add phrase structure and all this utility vocabulary to a protolanguage. We still don't yet have modern language. In particular, there is no notion of subject and object of a sentence—only semantically defined notions like Agent and Patient. There is no grammatical differentiation of parts of speech, only Object words versus Action words. There is no inflection for case and agreement, and no use of pronouns and other proforms. There is no regimented way of constructing long-distance dependencies such as the relation of a *wh*-phrase at the front of a clause to the "gap" or "trace" elsewhere within the clause, as in *[Which bananas]$_i$ do you think Fred advised us to buy t_i for our soup?*

However, we are moving towards something that begins to be recognizable as modern language. In particular, we find "fossils" of this stage in the very productive system of "sentential adverbials" of various syntactic categories that appear freely at the beginning of the sentence, after the subject, or at the end. Consider the examples in (2).

(2) a. Obviously,
In my opinion,
With a sigh,
Susan having gone,
Having nothing better to do,
Sick at heart,
Though basically a happy guy,
} Fred left town.

b. Fred, { obviously
in my opinion,
with a sigh,
Susan having gone,
having nothing better to do,
sick at heart,
though basically a happy guy, } left town.

c. Fred left town, { obviously. / in my opinion. / with a sigh. / Susan having gone. / having nothing better to do. / sick at heart. / though basically a happy guy. }

The use of these expressions is governed only by rudimentary syntactic principles. As long as the semantics is all right, a phrase of any syntactic category can go in any of the major breakpoints of the sentence: the front, the end, or the break between the subject and the predicate.

Another such subsystem comprises the prepositional phrases and "adverbials" denoting time, place, instrument, accompaniment and so forth. These are freely ordered at the end of the verb phrase; syntax apparently just lets one lump them there any old way:

(3) a. Sam struck gold last night in Alaska with his trusty pick.
Sam struck gold in Alaska last night with his trusty pick.
Sam struck gold with his trusty pick last night in Alaska.
b. Beth bought a book yesterday for her sister for $10.
Beth bought a book for her sister yesterday for $10.
Beth bought a book for $10 for her sister yesterday.

Again, this freedom bespeaks a somewhat protosyntactic phenomenon. The relation of each phrase to the sentence is determined more or less pragmatically, using the meaning of the noun as a guide: *last night* is a time because *night* is; ***for*** *Bill* is a recipient because *Bill* is a person, ***for*** *an hour* is a duration because *an hour* is a time, and ***for*** *$10* is a quantity of exchange because *$10* is an amount of money. Similarly, ***with*** *a knife* denotes an instrument, ***with*** *Bill* denotes a collaborator, and ***with*** *care* denotes a manner. *For* tends to be used for relations directly or indirectly involving purposes, but it is not entirely consistent; *with* is even less characterizable in general.

The architecture at this point is a true tripartite system, as shown in Fig. 8.4.

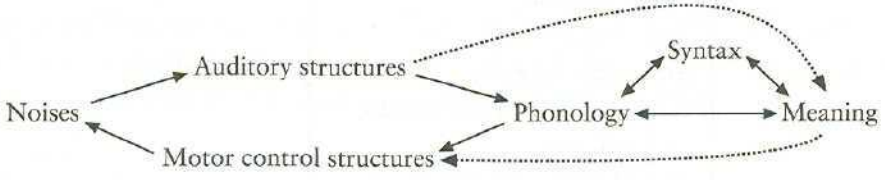

Fig. 8.4. 'Early modern' tripartite architecture

The syntax in this architecture determines a set of syntactic categories which are domains of word and phrase order regularities. For instance, the word order possibilities within a phrase are the same no matter where the phrase occurs in the sentence. In addition, syntactic phrases are domains for interface principles to semantics; for instance, Agent First is now an interface principle between noun *phrases* and thematic roles rather than between *nouns* and thematic roles.

At the same time, because evolution is incremental, the direct relation between phonology and meaning does not go away. In particular, the connection between word meanings and word pronunciations is direct, bypassing syntax. And this important characteristic remains in modern language: as has been observed several times already, syntactic features of words are far coarser than phonological and semantic features. For instance, as far as syntax is concerned, *dog*, *cat*, *worm*, *amoeba*, and *tree* are indistinguishable: they are all just singular count nouns. Similarly, verbs with identical argument structure such as *jog*, *walk*, *shuffle* and *strut* are syntactically indistinguishable. Thus the syntactic features serve only to tell the syntax how this word can fit into a phrase structure; they correspond only very coarsely to meaning. (As Levelt 1999 puts it, "Syntax is the poor man's semantics." See also Pinker 1989, where the semantic coarseness of argument structure plays an important role in the theory of children's acquisition of verbs.) The mapping between phonology and fine-scale meaning differences ought therefore to be maintained as part of a straight phonology–semantics interface without syntactic intervention. (And the "defective" words like *ouch* have *only* this connection.)

8.10 Grammatical categories and the "basic body plan" of syntax

So far we have managed to do without the distinction between nouns and verbs. Everything is done semantically. How does this grammatical distinction arise along with the further differentiation of adjectives, prepositions, and so on? I have only a very speculative story here, but I will offer it anyway.

An important asymmetry between nouns and verbs came up in section 5.5. Nouns can express any semantic category whatsoever: not just objects but situations (*situation*, *concert*, *earthquake*, *perusal*), properties (*size*, *intelligence*, *redness*), spatial concepts (*distance*, *region*, *place*), times (*Tuesday*, *millennium*), and so on. But verbs can express only situations (events, actions and states). In section 5.5 this was left as an unexplained design feature of the syntax–semantics interface.

Carstairs-McCarthy (1999) observes another significant asymmetry in language, one that is not at all necessary to its expressive power. The sentence (4a)

and the noun phrase (4b) convey the same information in their words, and yet only the sentence is acceptable as an independent non-elliptical assertion.

(4) a. Fred perused a book yesterday.
b. Fred's perusal of a book yesterday

More generally, an utterance cannot stand on its own without a verb.[14] Sometimes this is even at the price of adding dummy items. Why do we have to say the full sentence (5a) rather than the cryptic (5b)?

(5) a. There was a storm last night.
b. a storm last night

Carstairs-McCarthy observes that we could easily design a language that lacked the noun–verb distinction, in which (4a) and (4b) would translate into the same utterance, and in which something closer to (5b) than to (5a) would be grammatical. But there are no human languages like that.[15]

A telling case of this asymmetry is provided by Japanese and Korean. These languages make heavy use of the "light verb" construction mentioned in section 5.9.4. An example of this construction in English is *Sally took a walk*, where *take* is a "light verb," contributing little to the meaning, and where the nature of the action is conveyed by the noun *walk*. (6) is an example from Korean (thanks to Jong Sup Jun); note that the word translated as *study* receives an accusative case marker like other nouns.

(6) Inho-ka hakkyo-eyse yenge-lul kongpu-lul yelsimhi ha-n-ta
Inho-NOM school-at English-ACC study-ACC hard do-Pres-Declarative

'Inho studies English hard at school.'

It turns out that a sizable proportion of Korean verbs are actually such complexes of nominal plus light verb, and there is no simple verb with the same meaning. (6) is such a case; examples in English might be *take umbrage* and *make a deal*. According to a count by Jee-Sun Nam (Nam 1996), about 9000 out of 13,500 "verbs" listed in Korean dictionaries are actually light verb complexes. That is, despite so many actions being expressible only as nouns, Korean and Japanese still need a verb in the sentence.

---

[14] Some might say that what is necessary is Tense rather than a verb; and since Tense normally requires a verb, the verb comes along automatically. If anything this only exacerbates the puzzle.

[15] There is one notable exception to this generalization. Many languages, e.g. Russian and Hebrew, have no present tense form of the verb *be*, so that *Beth is hungry* comes out *Beth hungry*. There are various resolutions to this case. For the moment I will take the easy way out and say that present tense *be*, the verb with the least possible content, is expressed in these languages by a lexical item that lacks phonology but still appears in syntax.

Is there a connection between these two asymmetries between nouns and verbs? Suppose at some point in the evolution of UG, words expressing situations took on a special function in governing phrase structure and word order. These are, after all, the relational words par excellence, the ones that have the most articulated semantic argument structure. This special function might consist in (or develop into) their becoming grammatically essential to expressing an assertion. That would leave everything else as the default class. This is not too far off the noun–verb distinction: verbs are situation words and are essential to expressing an assertion. Nouns are everything else. In fact, when situation words are not used with their special grammatical function, they can easily fall into the default class, so nouns can express situations too. In other words, we get both asymmetries from the supposition that syntactic categories first emerged as a result of distinguishing verbs from everything else.

We might think of these special features of verbs versus nouns as a "basic body plan" for language. Japanese and Korean might be thought of, then, as languages whose vocabulary is not optimally suited to this plan: they could be more concise by just dispensing with the light verb most of the time. But they are stuck with it. Just as whales cannot go back to using gills and are therefore hampered by having to go to the surface to breathe, these languages cannot abandon the need for a verb in the sentence.

As we will see in a moment, once the noun–verb distinction is present, many other design features can collect around it.

8.11 Morphology and grammatical functions

To move from this point to modern language, two independent sets of machinery must be added: morphology and further aspects of syntax. Bickerton (along with many modern generative linguists) treats these as a completely integrated whole that forms the core of the "grammar box." On the other hand, as seen in section 6.2, attempts within generative theory to integrate morphology and phrasal syntax seamlessly have (to my taste) resulted in a certain artificiality.

How might the *brain* treat morphology and phrasal syntax? A good analogy elsewhere in the brain is depth perception, where we find a variety of disparate mechanisms, ranging from very sensory (lens accommodation) through perceptual (stereopsis and occlusion) through very cognitive (knowing what sizes things should be). These all converge on a single aspect of perceptual representation—the distance of visible surfaces from the viewer. Sometimes they are redundant; at some distances one or another predominates; and in illusions they may conflict.

A similar approach to phrasal syntax and morphology, advocated by Autolexical Syntax (Sadock 1991), Role and Reference Grammar (Van Valin and LaPolla 1997), and recent Lexical-Functional Grammar (Bresnan 2001), treats them as somewhat independent systems that accomplish partially overlapping functions. Section 5.6 referred to these systems as the tiers of phrasal syntax and morphosyntax, and pointed out some of the differences in the sorts of combinatoriality they offer. These systems, like those involved in depth perception, interweave with each other, sometimes redundantly, sometimes not, but both helping make explicit the semantic relations among components in an utterance. For instance, phrasal syntax may signal thematic roles (who did what to whom) through the order of phrases in relation to the verb. Inflection may do the same thing by means of verb agreement with the subject (and in some languages, with the object as well). Inflection may alternatively express semantic roles through a system of case marking, as in German, Russian, and Latin. In particular, many of the so-called "semantic cases" in these languages have much the same character as the phrasal markers *for* and *with* discussed in the previous section. And parallel to verbs in English that require "governed prepositions" (section 5.8), case-marking languages often have verbs that govern so-called "quirky case" on an argument. Languages tend to mix and match these strategies in different proportion; languages with rich inflectional systems often allow more freedom in word order for different purposes, usually for focus–topic information. On the other hand, inflection can be used (freely or redundantly with word order) to indicate focus or topic as well, for example the Japanese suffix *-wa*, which typically marks topic redundantly with initial position in the clause.

Thus we might think of phrasal syntax and morphosyntax as independently evolved systems, each built on top of the system of protolanguage, each refining communication through its own expressive techniques. In a similar vein, Casey and Kluender (1995) suggest that agreement inflection evolved as an extra system to provide redundant (and hence more reliable) information about semantic relations of arguments. I see no immediate reason to assert the temporal priority of one of these systems over the other in the course of evolution.

Notice that parts of these systems, especially inflection, depend strongly on the noun–verb distinction. Verbs are marked for agreement with nouns; nouns are marked by verbs for case.

An important innovation that remains is the system of grammatical functions: subject, object, and (in some languages) indirect object. Section 5.10 suggested that this system may well form a separate tier of syntax, "GF-structure," interfacing with both phrasal syntax and morphosyntax and relating them to semantic functions. The present evolutionary context invites the intriguing

hypothesis that grammatical functions were perhaps the latest-developing part of the architecture—an extra bit of power for efficiently regulating the syntactic expression of semantic relations. Deviating from practice in Lexical-Functional Grammar and Relational Grammar, section 5.10 proposed that this tier applies basically only to NP arguments, and that oblique arguments (i.e. arguments introduced by prepositions), adjuncts, and "adverbials" such as those illustrated in (2) and (3) above are "invisible" to it.

In section 5.10, the idea of such a limited system was justified by analogy with phonological tiers. Now a more powerful analogy suggests itself. Phrasal syntax regulates only a limited aspect of the connection between phonology and semantics, while leaving other, evolutionarily prior aspects (including the all-important word-to-meaning mapping) to the "earlier" system. We can see grammatical functions in a similar light: the syntax–semantics interface is partially regulated by the supervening, more "advanced" system of grammatical functions; but the "earlier" system is left in place, revealing itself around the corners of the "new" system.

The architecture now comes out like Fig. 8.5. It is not elegant. But then again, this is what we have come to expect of the brain. It hardly matches the connectivity of the visual system in its complexity.

8.12 Universal Grammar as a toolkit again

Suppose we take Fig. 8.5 as a sketch of "architectural universals" of language. How much of it is actually universal?

At rock bottom, the open vocabulary, phonology, and word concatenation are surely universal. But then we start running into exceptions. Some Australian languages, for example Warlpiri, Jiwarli, and Wambaya (Hale 1983; Nordlinger 1998; Austin 2001) show no consistent evidence of phrase structure. Their word order is almost entirely free, and adjectival modifiers are routinely separated

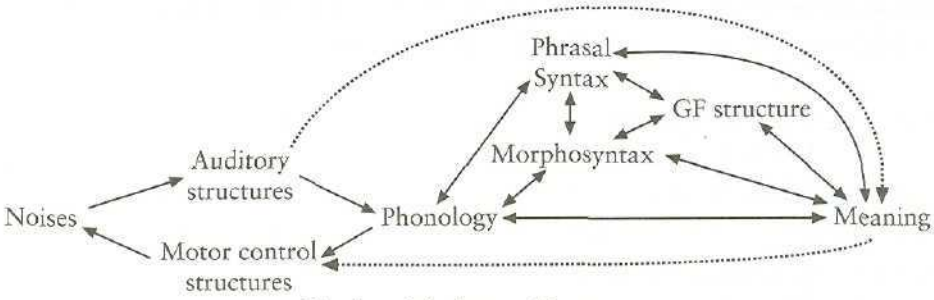

Fig. 8.5. Modern architecture

from the nouns they modify. Semantic relations are regulated by an exuberant case system; modifiers are connected with their heads by a shared case. Subordinate clauses are kept together, but within them order is also totally free. Of course, the rich morphological system requires use of the morphosyntax component. (Stephen Anderson has suggested (p.c.) that Classical Latin too may fit this characterization.)

As mentioned earlier, even in the usual cases when phrase structure is demonstrably present, there is wide variation among languages in how rich a use is made of phrase order and how rich the morphology is.

Turning to grammatical functions, Van Valin and LaPolla (1997) argue that Acehnese (an Austronesian language from Sumatra) does not make use of grammatical functions. Recall that the argument *for* grammatical functions is that there are grammatical principles that refer to subject and/or object independent of their semantic roles. Van Valin and LaPolla show that all analogous grammatical principles in Acehnese actually are dependent on the semantic roles Actor and Undergoer (Patient), so there is no justification for syntactically independent subject and object roles. They mount a similar argument for Mandarin Chinese, where the relevant semantic roles are Topic and Comment. They also argue that in intransitive sentences, Warlpiri acts like it has grammatical subjects, but in transitive sentences, the relevant notion is Actor.

In English, of course, grammatical functions play a role with respect to the arguments of verbs. But for the arguments of nouns the issue is not so clear. For instance, there is something that looks rather like a passive in noun phrases. Parallel to the active and passive sentences (7a) and (7b) are the "active" and "passive" noun phrases (7c) and (7d).

(7) a. The enemy destroyed the city.
b. The city was destroyed by the enemy.
c. the enemy's destruction of the city
d. the city's destruction by the enemy

But such "passive" noun phrases are highly quirky and subject to ill-understood semantic factors (see Grimshaw 1990). For instance, the paradigms in (8) and (9) admit a verbal passive but not a nominal "passive."

(8) a. John observed Bill.
b. Bill was observed by John.
c. John's observation of Bill
d. *Bill's observation by John

(9) a. John knew the answer.
b. The answer was known by John.

c. John's knowledge of the answer
d. *the answer's knowledge by John

This suggests that the grammatical-function tier applies to verbal structures in English, so that verbal passives are governed by grammatical functions; but that NP structures, including their "passives," are left to the more "primitive" semantically based system. Stiebels (1999) argues that in Classical Nahuatl, by contrast, nouns and verbs are totally symmetrical with respect to argument structure properties, suggesting that perhaps the grammatical-function tier in this language has a more extensive interface with syntax than in English.

My (perhaps self-centered) impression is that English rather than Acehnese or Classical Nahuatl represents the typical situation. Thus again we face an important grammatical asymmetry between verbs and nouns, with verbs typically playing a much more intricate role in determining sentence structure. This accords with the account suggested in the previous section, in which verbs were at the forefront of the evolution of syntactic categories, and therefore are syntactically more specialized.

Faced with this range of variation, what is the appropriate position to take on Universal Grammar? Van Valin and LaPolla's position is that if a characteristic is not universal, it is not part of Universal Grammar. Yet if Universal Grammar is to be the unlearned basis from which language is learned, it had better be available to help children learn case systems, agreement systems, fixed word order, and grammatical functions in case the language in the environment happens to have them. These are after all the most abstract parts of language, the ones least amenable to semantic and pragmatic support. This leads us back to the view of Universal Grammar as a "toolkit," introduced in section 4.3: beyond the absolutely universal bare minimum of concatenated words—the components of protolanguage—languages can pick and choose which tools they use, and how extensively.

This view of the evolved architecture of language has ramifications for other phenomena adduced as evidence for Universal Grammar. Section 8.7 discussed the apparent conflict between calling the Basic Variety an instance of UG but the very similar pidgin languages *not* an instance of UG. The difficulty was that UG was being thought of on both sides as an indivisible "grammar box." Here we were able to resolve the paradox by saying that both make use of part but not all of modern UG.

With this observation in hand, consider the critical period in language acquisition (section 4.9.2). As has been observed in the course of this chapter, some parts of language do not display critical period effects, in particular the acquisition of vocabulary, the concatenation of words, and the simple semantically

based principles of word order. Other parts of language, such as details of phonology, phrase structure, and in particular the inflectional system, can be severely impaired in late language learning, yielding Basic Variety and pidgins. Similarly, morphology above all is vulnerable to agrammatic aphasia and to Specific Language Impairment (at least as currently described).

We can localize these problems through viewing Universal Grammar as built out of layered subcomponents. UG is not simply on or off in these abnormal situations. Rather, some of its subcomponents are particularly impaired—significantly, the same ones in case after case. The robust remnant is protolanguage.

The overall conclusion is that grammar is not a single unified system, but a collection of simpler systems. Many of these systems are built up as refinements of pre-existing interfaces between components. Hence the evolution of the language capacity can be seen as deeply incremental, adding more and more little tricks to the cognitive repertoire available to the child faced with acquiring a language.

We should also observe that these subsystems are added specializations *for language*. For instance, a system of grammatical relations and a system of morphological agreement make a lot of sense as refinements of a syntax–semantics mapping. But they are totally useless to any other cognitive capacity; they are exquisitely specialized. This should occasion no surprise: the successive refinements of the structure of the eye—and the visual parts of the brain such as stereopsis—are useless for anything else too. In other words, there is no need to appeal to changes in general-purpose capacities, *à la* Piattelli-Palmerini and Toulmin, to explain the incremental development of the language capacity.

What is also new here is the hypothesis that certain design features of modern language resemble "fossils" of earlier evolutionary stages. To some degree, then, the examination of the structure of language can come to resemble the examination of the physical structure of present-day organisms for the traces of "archaic" features.