Possible stages in the evolution of the language capacity

Ray Jackendoff

Much current discussion of the evolution of language has concerned the emergence of a stage in which single vocal or gestural signals were used symbolically. Assuming the existence of such a stage, the present review decomposes the emergence of modern language into nine partially ordered steps, each of which contributes to precision and variety of expression. Bickerton's proposed 'protolanguage' falls somewhere in the middle of this succession. In addition to the by-now accepted evidence from language learning, language disorders, and ape language experiments, modern languages provide evidence of these stages of evolution through the presence of detectable 'fossils' in vocabulary and grammar.

Bickerton has proposed that the human capacity for language evolved in two stages[1]. His second stage is language as we know it ('modern language'). He calls the first stage 'protolanguage'; for now, one can think of it as modern language minus syntax. Bickerton's interesting claim is that protolanguage is still present in modern humans, surfacing in the course of language learning and when normal language is disrupted. Thus evolution did not throw a Good Idea away; rather it built on it.

Here, I want to elaborate on Bickerton's idea: one actually can discern in the structure of human language a substantial number of distinct innovations over primate calls, some prior to Bickerton's protolanguage, and some later. 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, in some instances I will be able to demonstrate 'fossils' of earlier stages of language in the modern language itself, offering a new source of evidence on the issue.

The consequence will be that the language capacity can be conceived of as having evolved incrementally, rather than appearing all at once in an undecomposable bloc. It will no longer be meaningful to ask 'Does primate P 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?' This helps defuse a long-running dispute. On one side are those such as Chomsky[2] who have advocated a complex, innate, and unified language capacity that would seem difficult to explain through natural selection; they have therefore been forced to devalue evolutionary argumentation[3]. On the other side are those who insist on evolutionary justification and are therefore inclined to deny or at least minimize an innate language capacity[4]. The position proposed here – a complex language capacity that evolved incrementally – helps define a middle ground that I hope will be a useful contribution to discourse.

I will assume without justification that any increase in explicit expressive power of the communicative system is adaptive, whether for cooperation in hunting, gathering, defense[5], or for social communication such as gossip[6–8]. I will also take it for granted (although it has been disputed) that linguistic adaptation arose first in the interest of enhancing communication and secondarily in enhancing or refining thought[9]. Finally, I will assume that the evolution of language proceeded through the vocal–auditory channel, though nothing in my argument precludes an initial stage of gestural (sign) language. (All these assumptions are no doubt controversial, but that is a topic for a different article.)

In this discussion I take my cue from the observation of Wolfgang Köhler[10] that cognitive steps which appear to us altogether natural might 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 it is not inevitable that evolution should immediately chance upon apparently natural and adaptive aspects of cognition. Thus we must not take it for granted, as some researchers do[11], that, for example, an organism with hierarchically organized behavior is therefore poised to invent syntax. The steps I propose are summarized in Box 1.

The use of symbols

Donald[3], Deacon[12], and Aitchison[13] have stressed that, well before fully-fledged modern language, there must have been voluntary use of symbolic vocalizations (or other signals such as gestures). Deacon seems to think that symbols

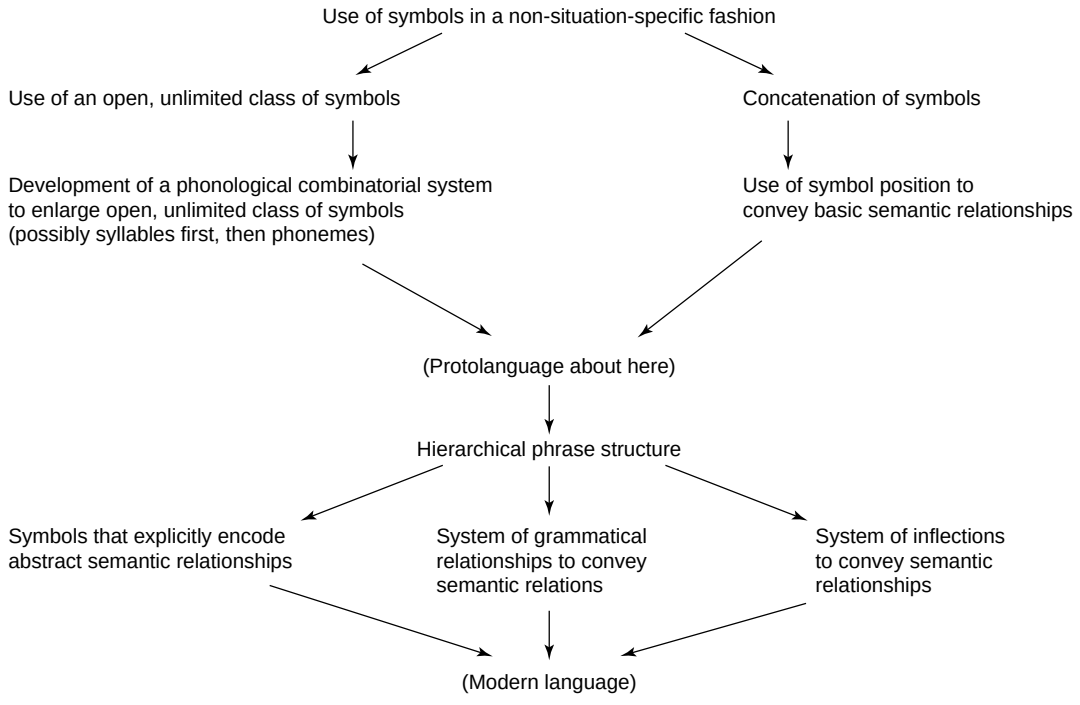
R. Jackendoff is at the Volen Center for Complex Systems, Brandeis University, Waltham, MA 02454, USA.

tel: +1 781 736 3264
fax: +1 781 736 2398
e-mail: jackendoff@brandeis.edu

1364-6613/99/$ – see front matter © 1999 Elsevier Science. All rights reserved. PII: S1364-6613(99)01333-9

Box 1. Steps in the evolution of language

Independent steps appear side by side; dependencies among steps are indicated vertically.



require grammatical combination; however, a single vocalization (as in a one-year-old's single-word utterance) can clearly serve symbolically. On the other hand, single-symbol utterances go beyond primate calls in important respects.

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 ask if anyone has seen a leopard lately[14].

Achieving this stage is a major evolutionary step: Deacon and Donald 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 see as a straightforward and inevitable development of language from such humble beginnings.

Notice that even the one-word stage shows considerable subtlety. For instance, very early in child language we already see an appreciation of 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)[15]. Considerable inquiry has been focused on how children can acquire (or innately have) this aspect of semantics[16–19]. 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[20]); that is, non-situation-specific use of a repertoire of single symbols, including both symbols for individuals (proper names) and symbols for categories (common nouns).

However, we can potentially go back further: certain little-remarked aspects of modern language are if anything more primitive than the child's one-word utterances. Consider utterances associated with sudden high affect, such as *ouch!*, *dammit!*, *wow!* and *oboy!* These exclamations have no syntax and therefore cannot be integrated into larger syntactic constructions (other than those that allow direct quotes). They can remain in the repertoire of the deepest aphasics, apparently coming from the right hemisphere[21]. There also exist 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 noncommunicatively, 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. Further single-word utterances include 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. (Note that no animal call system includes a signal of generalized negation like *no*.) 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.

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.

Consider first the open vocabulary, the repertoire of meaningful linguistic units stored in long-term memory. By contrast with animal call repertoires (the closest appropriate comparison), which number roughly in the dozens at most, estimates of the vocabulary of an average speaker run into the tens of thousands. Children learn these in droves, and we keep picking up new words all our lives[22]. Such a large vocabulary places significant demands on long-term memory and rapid retrieval[3].

The language-trained apes, by contrast, are reported to have acquired at most several hundred symbols, mostly through extensive training, but in some instances appearing to 'just pick them up'[20]. At present it is unknown what accounts for the hundredfold difference in vocabulary size. It might be a consequence of the larger brain, or alternatively of some special human 'tuning' that makes vocabulary learning vast and effortless. I suspect that vocabulary learning for apes is rather like children's learning of reading, a largely effortful undertaking requiring much motivation and instruction, and that children's *spoken* (and signed) word learning is a rather more specialized cognitive process.

Late second-language learners can be counted on to acquire substantial vocabulary, even when their grammar and especially pronunciation is far from fluent. Moreover, in the famous case of Genie, vocabulary acquisition began immediately upon her discovery, and her rate of vocabulary acquisition approximated that of young children[23]. Yet after years of training, her grammar remained exceedingly rudimentary. 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[3] has stressed, the uniquely human ability to imitate obviously plays a role here. 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. 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.

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 gestalts, like primate calls, even a thousand symbols would be impossible to keep distinct in perception and memory. Modern language deals with this problem by building words up combinatorially from a repertoire of a dozen to a few dozen smaller meaningless speech sounds. Using concatenated speech sounds to construct conventionalized vocalizations turns the distinction among vocalizations into a categorical/digital matter rather than a graded one.

Lieberman[24] observes that as late as the Neanderthals, the shape of the vocal tract did not allow the multitude of easily perceptible distinctions among speech sounds in modern language. (This conclusion is not universally accepted, however[25].) Still, as Lieberman points out, an open vocabulary is 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 (some Polynesian languages make do with only twelve phonemes). The evolution of the vocal tract can be seen as driven by the adaptivity of a larger vocabulary, through more rapid articulation and enhanced comprehensibility.

An intermediate stage in evolving a phoneme-based vocabulary might use the syllable as the generative unit. The syllable is basically a unit of articulatory gesture, and 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). Levelt and Wheeldon[27] have offered psycholinguistic evidence that the repertoire of 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.

Thus an open-ended class of single-symbol utterances could be composed from a generative system whose basic units were not individual speech sounds but rather (proto-) syllables, each of which was a holistic vocal gesture[27]. With a repertoire of 10 such gestures, one could build 100 two-proto-syllable vocalizations and 1000 three-proto-syllable 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 development of (modern) syllables analytically composed of phonemes could then be seen as a further step in language evolution, making 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.

As many linguists (but not many nonlinguists) have recognized, the innovation of phonological structure is a major cognitive advance[24,28]. It requires us to think of the system of vocalizations as 'generative', 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 generativity nonetheless: it is a way of systematizing existing vocabulary items and being able to create new ones. A generative phonological system is thus a crucial step in the evolution of language, necessary for the vocabulary to achieve its presently massive size.

To my knowledge none of the ape experiments have 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 unanalysed 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 constitute the parallel to syllabic structure in spoken languages[29].

Concatenation of symbols to build larger utterances

We have so far considered systems of symbols in which each symbol constitutes a complete utterance, analogous to a very young child's one-word stage, but with a larger vocabulary. 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.

One important virtue in Bickerton's proposal of a two-stage evolution for the language capacity is in pointing out that one can go beyond single-word utterances without achieving modern syntax. Much of the rest of this discussion 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 18-month-old or a signing chimp) might express any number of connections between Fred and apples, expressible in modern language 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, this is far better than just *Fred* or *apple* in isolation. Moreover, it isn't totally vague: it probably wouldn't be used to express '*Fred has certain 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 from that discussed in the previous section. Phonological generativity is a way of analysing 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 meaningful utterances. The two could have evolved simultaneously or in either order.

This sort of combination has not been attested in the ethological literature. We do find bird songs, cetacean songs, and possibly some primate 'long calls' built up out of smaller units; but the units are not meaningful on their own, and/or different combinations are not distinctively meaningful[30,31]. (This might, however, be for lack of means to look for such combinations[14].) On the other hand, the language-trained apes show this capability, at least on some assessments[20].

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. Herbert Terrace 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 further claims that a careful look at the full data from the other signing experiments reveals similar results[32]. Other researchers have claimed greater organization, to which we refer below.

Using linear position to signal semantic relationships

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 relationships between them, and introducing new sorts of vocabulary items that convey relationships 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 Fred tree*', did Fred hit the tree or did the tree hit Fred? Though the larger context might tell us, the pragmatics of the words alone don't tell us. Pinker and Bloom[5] point out this problem and argue that using principles of word order would be communicatively adaptive.

However, one need not 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 Klein and Perdue's massive longitudinal study of adult second-language learners with various native languages and target languages[33]. 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: (1) lexical competence; (2) absence of inflectional morphology, e.g. verbs always appear in a fixed form rather than undergoing tense and agreement inflection; (3) absence of sentential subordination (no relative clauses, indirect quotes, etc.); (4) simple, largely semantically based principles of word order. The most prominent of these principles are 'Agent First' and 'Focus Last'. So BV is quite far from full linguistic competence.

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; it enables one to disambiguate a large proportion of utterances involving two characters. It further remains quite powerful in structuring word order in modern language: it appears as the default principle 'Agent is expressed in subject position', which can of course be mitigated by constructions such as the passive[34,35].

Agent First seems to be observed as well in the 'home signs' invented by deaf children of non-signing parents[36], and in pidgin languages[37]. Agrammatic aphasics also fall back on this principle to some degree, explaining some of their errors on reversible passives ('*The boy was hit by the*

Box 2. Sample semantic relationships in English noun–noun compounds

(In these examples, the presence or absence of a space between the nouns is purely an accident of spelling.)

Locative relationships: *doghouse* = house for a dog to live in; *housedog* = dog that lives in a house.

Part–whole relationships: *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.

Resemblance relationships: *zebrafish* = fish that resembles a zebra.

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.

Note however that the relationship 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.

girl') and object relatives ('*The boy who the girl kissed is tall*')[38]. 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.

Agent First concerns an element in the system of 'thematic roles', the specification of who did what to whom. By contrast, Focus Last concerns an element in the discourse coding of given and new information. 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. In many languages of the world, discourse coding plays a much greater role than it does in English; Japanese, Hungarian and Tagalog are prominent examples[35]. To my knowledge, no one has investigated discourse coding in language-trained apes; I also know of no results from home sign.

BV is fairly close to what Bickerton[1] 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 (1)–(3) of BV. However, Bickerton attributes to protolanguage a less stable word order than that of BV; this might be partly because his evidence comes from pidgins, which are heavily influenced by the native language(s) of their speakers. Bickerton does not address Agent First and Focus Last, which may well be present in agrammatics and Genie – but perhaps not in the apes.

I suggest, then, that Agent First and Focus Last 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.

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 leaves 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*'.

Like Agent First, Grouping is a purely semantically-based principle that maps into linear adjacency without using anything syntactic like a Noun Phrase. But such linear groupings might well be the underpinnings of constituents like Noun Phrase (see the next section). This principle can thus be seen as another 'fossil' feature of modern language.

Another possible protolinguistic 'fossil' in English is the formation of compound nouns such as *snowman* and *blackboard*. A wide variety of semantic relationships is possible between the nouns, in large part mediated by their meanings (see Box 2). The situation resembles the possible meaning relationships conveyed by raw concatenation: in unpublished work, I have found a repertoire of perhaps twenty relationships that can be conveyed in compounds through pragmatics alone (though the reason for this particular set of relationships remains for the moment unclear). Klein and Perdue report that noun compounding is the only kind of morphology found in the Basic Variety; and children use compounding very early[39].

The facts of compounding hence seem symptomatic of a protolinguistic 'fossil': the grammatical principle involved is simply one of concatenating two nouns into a bigger noun, and the semantic relationship between them is determined by a combination of pragmatics and memorization. Thus determining the meaning of a newly encountered compound is much like determining the meaning of '*hit tree Fred*' discussed above.

Whatever the particular details of such principles mapping between semantic roles and linear order, they sharpen communication. They are therefore a plausible step between raw concatenation and full syntax. In fact, raw 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[20].

At the same time, protolanguage of this sort is still far from the expressive possibility of modern language. I will now discuss some further steps on the route to modern language.

Phrase structure

All the phenomena discussed so far use word order to signal semantic relationships among words; but this is not sufficient for modern language. For example, in the previous

Box 3. Families of vocabulary that express abstract semantic relationships

Spatial relationship 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 relationships: *up*, *to*, *next to*, *behind*, *to the side*, and *on*.

Time terms: This includes 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: 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 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, '*n'est-ce pas*'? 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.

Quantification: These include the standard logical quantifiers *some*, *all*, and *every*, as well as the 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 flattering*), and which is acquired very early by children.

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*. Similarly, compare *He ate the apple and he died* and *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. With only implicit expressions of reason, one cannot ask *Why?* and therefore seek explanation. (Does the famous explosion of *why*s in young children represent their discovery of reasons?; Ref. a). Finally, compare '*I threw the spear and it hit the pig*' with '*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). More generally, the prepositions *for* and *with* in English seem to serve as all-purpose connectors for a wide range of relationships of this sort among events.

More general discourse connectors: This includes words such as *but*, *however*, *therefore*, *moreover*, *what's more*, *and so forth* and so on.

Reference

a Kelemen, D. (1999) Beliefs about purpose: on the origins of teleological thought, in *The Descent of Mind* (Corballis, M.C. and Lea, S.E.G., eds), pp. 278–294, Oxford University Press

sentence, the entire phrase 'all the phenomena discussed so far' enters into a semantic relationship with the verb 'use'. This collection of words functions as an elaborated version of the single word 'phenomena', the head of the phrase.

The grouping of words into higher-order headed units is a crucial design feature of modern language. It allows principles of word order to be expanded to 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. An important part of this innovation is that a phrase can consist not just of words but also of smaller phrases. For instance, 'all the phenomena discussed so far' contains the phrase 'discussed so far', which itself contains the phrase 'so far'. This hierarchical embedding, 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, much discussion of pidgin languages has not made a clear distinction between word order and phrase structure. This distinction thus deserves closer examination.

Vocabulary for relational concepts

Another possible way of encoding semantic relationships among words and phrases is to invent words that explicitly express such relationships. 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. Box 3 lists some of these.

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 them; nor would organisms that had one class of them 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.)

Relational vocabulary plays an important role in thought. It has been argued[9,40] 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 consciously 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.)

Beyond phrase structure: inflection and further syntax

To move from here to modern language, two independent sets of machinery must be added: inflection and further aspects of syntax. Bickerton (along with many modern generative linguists) treats these as of a piece, a seamless whole that forms the core of the language module. On the other hand, attempts within generative theory to integrate the two seamlessly have (to my taste) resulted in a sense of artificiality.

An alternative approach treats phrasal syntax and inflectional morphology as somewhat independent systems that accomplish partially overlapping functions[41]. A good brain analogy is depth perception, where we find a variety of disparate mechanisms, ranging from very sensory (lens accommodation) through perceptual (stereopsis and occlusion)

Box 4. Phenomena bearing traces of 'protosyntax'

Consider the range of 'adverbial' expressions of various syntactic categories that appear freely at the beginning of the sentence, after the subject, or at the end (Fig. I).

{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.

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.

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.}

These expressions are governed only by rudimentary syntactic principles. As long as the semantics is correct, 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. Similarly, the prepositional phrases and adverbials denoting time, place, instrument, accompaniment, manner, and so forth are freely ordered at the end of the verb phrase; syntax apparently just lets one lump them there any old way:

Sam struck gold unexpectedly last night in Alaska with his trusty pick.
Sam struck gold in Alaska unexpectedly last night with his trusty pick.
Sam struck gold with his trusty pick last night unexpectedly in Alaska.

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 speaks of a somewhat more protosyntactic phenomenon.

Fig. I. See text.

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 can conflict.

Similarly, phrasal syntax and inflection both help make explicit the semantic relationships among components in an utterance. For instance, 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. 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. In addition, around the corners of language, one can still find traces of the simpler system, where semantic relationships are indicated only 'by the seat of the pants'. Two examples are briefly presented in Box 4.

There are many important substantive differences between the two systems. Inflection normally marks only relationships internal to a single clause; by contrast, phrasal syntax is replete with long-distance dependencies, where a phrase appears outside the clause in which it would 'normally' be positioned. When a word bears multiple inflections, these come in a fixed 'templatic' order, often with no hint of hierarchical stacking; by contrast, phrasal syntax is full of alternative word orders and hierarchical structure. Inflection lends itself to a great deal of idiosyncratic irregularity; phrasal syntax much less so.

Under this conception, the two systems of grammar are built independently on top of the system of protolanguage, each refining communication through its own expressive techniques. I see no immediate argument for the temporal priority of one over the other in the course of evolution.

One might justifiably ask why syntax and inflection evolved the way they did and not some other conceivable way. For instance, why do languages almost invariably show a 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.

A possible scenario, proposed by Andrew Carstairs-McCarthy[42], is 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 according to strong phonological and phonetic evidence 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. In turn, this structure might have developed in phonology for acoustic or articulatory reasons, but in syntactic structure it is just one of those accidents of evolution.

Whether this proposal is correct or not, it has the right sort of flavor. In particular, it shows how attention to the

internal details of language structure can contribute to arguments about language evolution, which is of course one of the more general points of the present article.

I also hope to have illustrated how we can go beyond the idealization, perhaps appropriate for 1965 (Ref. 43) but not for now, of the language faculty as a *sui generis* mental phenomenon, unrelated to general cognition. Abandoning this idealization does not require us to jump to the other extreme and say there is nothing at all special about language. I have tried to show that (1) there are indeed many special aspects of language, but (2) that they could have evolved incrementally, not unlike the eye and the parts of the brain that the eye serves. Having less than the whole system would still have been useful.

What is also new here is the hypothesis that certain design features of modern language might be '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.

Acknowledgements

I am grateful to Dan Dennett for giving me the impulse to write this article. Discussions with him, Terry Deacon, Merlin Donald, Marc Hauser, and Tecumseh Fitch were instrumental in bringing the article to fruition; comments on earlier versions by Andrew Carstairs-McCarthy, Ludmila Lescheva, Biljana Mišić, Marc Hauser, and Paul Bloom helped me refine the article into its present form..

References

1 Bickerton, D. (1990) *Language and Species*, University of Chicago Press

2 Chomsky, N. (1975) *Reflections on Language*, Pantheon

3 Newmeyer, F.J. (1998) On the supposed 'counterfunctionality' of Universal Grammar: some evolutionary implications, in *Approaches to the Evolution of Language* (Hurford, J., Studdert-Kennedy, M. and Knight, C., eds), pp. 305–319, Cambridge University Press

4 Donald, M. (1999) Preconditions for the evolution of protolanguages, in *The Descent of Mind* (Corballis, M.C. and Lea, S.E.G., eds), Oxford University Press

5 Pinker, S. and Bloom, P. (1990) Natural language and natural selection *Behav. Brain Sci.* 13, 707–726

6 Dunbar, R. (1998) Theory of mind and the evolution of language, in *Approaches to the Evolution of Language* (Hurford, J., Studdert-Kennedy, M. and Knight, C., eds), pp. 92–110, Cambridge University Press

7 Power, C. (1998) Old wives' tales: the gossip hypothesis and the reliability of cheap signals, in *Approaches to the Evolution of Language* (Hurford, J., Studdert-Kennedy, M. and Knight, C., eds), pp. 111–129, Cambridge University Press

8 Worden, R. (1998) The evolution of language from social intelligence, in *Approaches to the Evolution of Language* (Hurford, J., Studdert-Kennedy, M. and Knight, C., eds), pp. 148–166, Cambridge University Press

9 Jackendoff, R. (1996) How language helps us think *Pragmatics and Cognition* 4, 1–24

10 Köhler, W. (1927) *The Mentality of Apes*, Routledge Kegan Paul

11 Corballis, M.C. (1991) *The Lopsided Ape*, Oxford University Press

12 Deacon, T. (1996) *The Symbolic Species*, Norton

13 Aitchison, J. (1998) On discontinuing the continuity–discontinuity debate, in *Approaches to the Evolution of Language* (Hurford, J., Studdert-Kennedy, M. and Knight, C., eds), pp. 17–29, Cambridge University Press

14 Hauser, M. (1996) *The Evolution of Communication*, MIT Press

15 Katz, N., Baker, E. and Macnamara, J. (1974) What's in a name? A study of how children learn common and proper names *Child Dev.* 45, 469–473

16 Macnamara, J. (1982) *Names for Things*, MIT Press

17 Bloom, P. (1999) The role of semantics in solving the bootstrapping problem, in *Language, Logic, and Concepts: Essays in Memory of John Macnamara* (Jackendoff, R., Bloom, P. and Wynn, K., eds), pp. 285–310, MIT Press

18 Carey, S. and Xu, F. (1999) Sortals and kinds: an appreciation of John Macnamara, in *Language, Logic, and Concepts: Essays in Memory of John Macnamara* (Jackendoff, R., Bloom, P. and Wynn, K., eds), pp. 311–336, MIT Press

19 Hall, D.G. (1999) Semantics and the acquisition of proper names, in *Language, Logic, and Concepts: Essays in Memory of John Macnamara* (Jackendoff, R., Bloom, P. and Wynn, K., eds), pp. 337–372, MIT Press

20 Savage-Rumbaugh, S., Shanker, S., and Taylor, T. (1998) *Apes, Language, and the Human Mind*, Oxford University Press

21 Jackson, J.H. (1874) On the nature of the duality of the brain *Medical Press and Circular* 1, 19–41 [cited by Marshall, J. (1980) Biology of language acquisition, in *Biological Studies of Mental Processes* (Caplan, D., ed.), pp. 106–148, MIT Press]

22 Carey, S. (1978) The child as word learner, in *Linguistic Theory and Psychological Reality* (Halle, M., Bresnan, J. and Miller, G., eds), pp. 269–293, MIT Press

23 Curtiss, S. (1977) *Genie: A Linguistic Study of a Modern-Day 'Wild Child'*, Academic Press

24 Lieberman, P. (1991) *Uniquely Human*, Harvard University Press

25 Aiello, L. (1998) The foundations of human language, in *The Origin and Diversification of Language* (Memoirs of the California Academy of Sciences, No. 24) (Jablonski, N. and Aiello, L., eds), pp. 21–34, University of California Press, San Francisco

26 Levelt, W.J.M. and Wheeldon, L. (1994) Do speakers have access to a mental syllabary? *Cognition* 50, 239–269

27 MacNeilage, P. (1998) Evolution of the mechanisms of language output: comparative neurobiology of vocal and manual communication, in *Approaches to the Evolution of Language* (Hurford, J., Studdert-Kennedy, M. and Knight, C., eds), pp. 222–241, Cambridge University Press

28 Studdert-Kennedy, M. (1998) The particulate origins of language generativity: from syllable to gesture, in *Approaches to the Evolution of Language* (Hurford, J., Studdert-Kennedy, M. and Knight, C., eds), pp. 202–221, Cambridge University Press

29 Wilbur, R.B. (1990) Why syllables? What the notion means for ASL research, in *Theoretical Issues in Sign Language Research* (Vol. 1) (Fischer, S.D. and Siple, P., eds), pp. 81–108, University of Chicago Press

30 Marler, P. (1998) Animal communication and human language, in *The Origin and Diversification of Language* (Memoirs of the California Academy of Sciences, No. 24) (Jablonski, N. and Aiello, L., eds), pp. 1–20, University of California Press, San Francisco

31 Ujhelyi, M. (1998) Long-call structure in apes as a possible precursor for language, in *Approaches to the Evolution of Language* (Hurford, J., Studdert-Kennedy, M. and Knight, C., eds), pp. 177–189, Cambridge University Press

32 Terrace, H. (1979) *Nim*, Knopf

33 Klein, W. and Perdue, C. (1997) The Basic Variety, or: Couldn't language be much simpler? *Second Lang. Res.* 13, 301–347

34 Givón, T. (1995) *Functionalism and Grammar*, John Benjamins

35 Van Valin, R. and LaPolla, R. (1998) *Syntax: Structure, Meaning, and Function*, Cambridge University Press

36 Goldin-Meadow, S. and Mylander, C. (1990) Beyond the input given: The child's role in the acquisition of language *Language* 66, 323–355

37 Bickerton, D. (1981) *Roots of Language*, Karoma

38 Piñango, M. (1999) Syntactic displacement in Broca's aphasia comprehension, in *Grammatical Disorders in Aphasia: A Neurolinguistic Perspective* (Bastiaanse, R. and Grodzinsky, Y., eds), pp. 74–89, Whurr

39 Clark, E.V., Gelman, S.A. and Lane, N.M. (1985) Compound nouns and category structure in young children *Child Dev.* 56, 84–94

40 Dennett, D.C. (1991) *Consciousness explained*, Little, Brown & Co.

41 Jackendoff, R. (1997) *The Architecture of the Language Faculty*, MIT Press

42 Carstairs-McCarthy, A. (1999) *The Origins of Complex Language*, Oxford University Press

43 Chomsky, N. (1965) *Aspects of the Theory of Syntax*, MIT Press