Adaptive Behavior

What Are the Unique Design Features of Language? Formal Tools for Comparative Claims

Willem Zuidema[1,2], Arie Verhagen[3]
[1] *Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands*
[2] *Behavioural Biology, Leiden University, The Netherlands*
[3] *Leiden University Centre for Linguistics, The Netherlands*

What are the "design features" of human language that need to be explained? Starting from R. Jackendoff's scenario for the evolution of language, we argue that it is the transitions between stages that pose the crucial challenges for accounts of the evolution of language. We review a number of formalisms for conceptualizations, sound, and the mapping between them, and describe and evaluate the differences between each of Jackendoff's stages in terms of these formalisms. We conclude from this discussion that the transitions to combinatorial phonology, compositional semantics and hierarchical phrase structure can be formally characterized. Modeling these transitions is a major challenge for language evolution research.

Keywords language evolution · design features · animal communication · semantics · phonetics · syntax

1 Introduction

Human languages are unique communication systems in nature because of their enormous expressiveness and flexibility. They accomplish this by using combinatorial principles in phonology, morphology, syntax, and semantics, which impose important requirements on the cognitive abilities of language users. Explaining the origins of the structure of language and the human capacity for learning and using it, are challenging and controversial problems for linguistics, cognitive science, and evolutionary biology. Much discussion has concentrated on the following questions: whether or not this capacity has been subject to natural selection; whether it evolved in a single, in few, or in many steps; and whether articulation, perception, or cognitive processing formed the crucial bottleneck (see Christiansen & Kirby, 2003, for an overview of positions).

Surprisingly little attention, however, has been paid to a precise, quantitative investigation of the range of differences between human languages and animal communication, and the many methodological challenges that arise when trying to make such comparisons. We argue that there is a need to develop formal tools for substantiating comparative claims. Moreover, in efforts to model the gradual evolution of language from a primate-like, ancestral state, it is necessary to model intermediate stages, for which again appropriate formalisms are often lacking. In this article, we review formalisms from various branches of linguistics and evaluate whether, and how, they can be used in such comparative and evolutionary research.

Correspondence to: Willem Zuidema, Institute for Logic, Language and Computation, University of Amsterdam, P.O. Box 94242, 1090 GE Amsterdam, The Netherlands. *E-mail*: zuidema@uva.nl.
Tel.: +31 20 525 6340, *Fax*: +31 20 5255206

(2010), Vol 18(1): 48–65.
DOI: 10.1177/1059712309350973

Hockett (1960) referred to the unique features of human language as its "design features." Hockett's list has been updated by several authors, and many of the design features have made their way into introductory linguistics textbooks (as reviewed in Smith, 2003, chapter 1). Jackendoff (2002) has integrated a similar list of design features into a scenario for the various stages in the evolution of human language from primate-like communication. Unlike many other theories, Jackendoff's scenario assumes several such intermediate stages. Jackendoff's proposal is useful for structuring the discussion in this article for a number of reasons:

- This scenario is a gradualist account, with many intermediate steps. Scenarios proposed by other scholars can be seen as variants of Jackendoff's, where two or more of the stages Jackendoff proposes are collapsed into one. Jackendoff's scenario can thus be seen as a generalization of many other scenarios.
- It grounds the scenario for the evolution of language in a theory of how modern humans acquire, represent, and process language, which is in principle testable.

Although Jackendoff's account is not very formal, his partitioning of the problem is sufficiently explicit to identify potentially relevant formalisms and apply them to issues in language evolution, in order to get a better understanding of the precise nature of the specific issues to be explained. By the same token, it allows us to identify relevant empirical phenomena that may be brought to bear on these issues.

In the following sections we will briefly sketch each of the stages that Jackendoff proposes, and relate his proposal to those of some other researchers. We will then introduce some formalisms for conceptualization, sound, and the mapping between the two (including formalisms for syntax). The goal of this article is to find out how to describe the similarities and differences between Jackendoff's stages, and to identify the transitions that can be characterized formally. Such a characterization is a first step; in this article we will not yet study the next step: the mechanisms that could lead to the transition from one stage to the next (see Nowak & Krakauer, 1999; Oliphant, 1999; Oudeyer, 2006; Zuidema, 2003; Zuidema & de Boer, 2009 for a number of proposals).

2 Stages in Jackendoff's Scenario

The starting point of Jackendoff's scenario is preexisting primate conceptual structure—that is, the kind of cognitive abilities that modern primatology has found in other great apes. The first innovation is the use of symbols in a non-situation specific fashion. Jackendoff recognizes that for instance chimpanzees have a sophisticated conceptual apparatus that is adequate to deal with navigation, problem solving, and complex social interaction. But he believes that primates are incapable of symbolic vocalization, that is, of referring to objects, properties, or events independent from the present situation. Deacon (1997), Donald (1991), and others have argued that the use of symbols is the crucial biological innovation that has made modern human language possible.

A second innovation is the ability to use and acquire an open, unlimited class of symbols. Whereas primate call systems contain a few dozen different calls at most (as far as we know) and language-trained apes can be taught at most several hundred symbols, humans know tens of thousands of different words. An open class of symbols must be learnt rather than be innate; Oliphant (1999) and others have argued that a learnt vocabulary was a crucial step in the evolution of language.

To keep such large numbers of symbols distinct in perception, memory, and articulation, a third innovation has been crucial: a generative, combinatorial phonological system. The sound systems of all human languages are combinatorial, in that the basic meaningful units (words, morphemes) are built up from a relatively small repertoire of basic speech sounds (phonemes, syllables), each of which is not a meaning bearing unit itself; this combinatorial character of human phonology is independent of that of sentences being composed out of words (this is the so-called design feature of "duality of patterning," to which we will return below). Jackendoff endorses the view that the syllable is the basic unit of combination. The evolution of combinatorial phonology is seen by a number of researchers as the crucial innovation in the evolution of language (Carstairs-McCarthy, 1999; Studdert-Kennedy, 1998), because it seems to be both a necessary and a sufficient condition for open-endedness in a vocal communication system (in principle allowing for an infinite number of distinct signals based on finite means).

Jackendoff's fourth innovation is the concatenation of symbols to build larger utterances. He imagines concatenations of symbols analogous to *Fred apple*, which might refer to any of a number of connections between Fred and apples. Although simple concatenation does not fully specify the intended interpretation, it is nevertheless, Jackendoff argues, more useful than single symbols in isolation.

The fifth innovation, however, using linear position to signal semantic relations, does introduce a systematic compositionality. In this stage of the scenario, simple principles such as "agent first," "focus last," and "grouping" could structure utterances analogous to *dog brown eat mouse*, such that it is clear that the brownness applies to the dog, the dog eats, and the mouse is being eaten. In the terminology of Hurford (2000), the route from holistic to compositional language in this scenario is "synthetic," because compounds are synthesized from preexisting meaningful signals (rather than that preexisting holistic signals are reanalyzed as being built-up from component parts – the "analytical route"; cf., Arbib, 2003; Wray, 2002a, 2002b).

Jackendoff sees the fourth and fifth innovations as independent from the second and third, and he does not decide which should come first. Together, they constitute something similar to the (single) intermediate stage of "protolanguage" in the scenario of Bickerton (1990) and others, and to pidgin (the limited language negotiated between adults with different native languages, Bickerton, 1990) and to "Basic Variety" (the limited language acquired by adult second language learners, Klein & Perdue, 1997).

The sixth innovation is the invention of hierarchical phrase structure. Phrase structure has been recognized since Chomsky (1957) as a crucial design feature of human language. Jackendoff argues that phrase structure allows the principles of word order, as emerged in stage 5, to be elaborated into principles of phrase order. Hence, from stage 6 a systematic relation has existed between sentences like *dog chase mouse*, where *dog* and *mouse* are single word noun phrases, and the similarly structured but more elaborate *big dog with floppy ears and long scraggly tail chase little frightened mouse*.

The seventh innovation is a vocabulary for relational concepts, introducing words analogous to present-day English *up*, *on*, *behind*, *before*, *after*, *often*, *because*, *and*, *also*, *only*, *of*, and so forth. These words all describe relations between different phrases in a sentence, and thus require phrase structure, Jackendoff argues, but not yet syntactic categories. Jackendoff imagines that the phrase order and use of relational words are still completely guided by semantically defined notions. That is, there are no subjects, objects, nouns, verbs, or mechanisms for case, agreement, or constructions such as *if … then*: two connected elements between which a string of arbitrary length can be inserted (so-called long-distance dependencies). There are just semantic categories such as agent, patient, objects, and actions.

Grammatical categories are the eighth innovation, creating syntactic notions such as "subject" that are correlated with but not equal to the semantic notion of agent (as, for instance, in the passive construction), or even a syntactic notion such as "sentence" which makes that *a storm last night* cannot stand on its own, whereas *There was a storm last night*, with dummy subject *There*, can. The final two innovations, inflectional morphology and grammatical functions (in no particular order) complete the extensive toolkit that modern languages make use of. This list of gradual innovations is consistent with the gradualist approach championed by Pinker and Bloom (1990) and others.

In summary, Jackendoff breaks down linguistic competence into a number of different skills, and proposes a gradual scenario in which new skills are added to the existing system, each step increasing the expressivity of the language used. These steps fall into three types: innovations in the domain of conceptualization, in the domain of form, and in the domain of the mappings between the two. The first innovation, the use of symbols referring to objects and situations independent from the communication event, is about the sort of concepts early hominids had available for communication. The third, about combinatorial phonology, is about the kind of sounds they could produce and perceive. All the other innovations, from an open, learnt vocabulary and the concatenation of symbols to inflectional morphology, are about the way messages are mapped onto sound and vice versa. In the remaining sections of this article we will discuss some empirical observations about, and formalisms for, modeling conceptualization, sound, and meaning—sound mappings in animal and human communication. Where possible we will evaluate—in terms of the formalisms at hand—whether Jackendoff's stages indeed capture the relevant innovations in the evolution of language. As we will see, many of the currently available formalisms are ill-suited for the task of describing the differences between modern human communication and that of other species, but it is often not straightforward to adapt them.

Note that although Jackendoff's scenario is quite unique in its details and therefore useful for structuring the discussion in this article, it is not uncontroversial. It is very much focused on the representational abilities of language users, and Jackendoff has followed a "subtractive" approach, starting with a suite of linguistic abilities assumed by a particular linguistic theory and removing them one by one to arrive at the state of a hypothetical primate ancestor. As a result, the scenario makes little reference to findings in animal cognition research, reflects the biases of a particular linguistic tradition (e.g., the assumed centrality of word order and phrase structure), and ignores the social/pragmatic context in which language evolved and the actual evolutionary mechanisms of transitioning from one stage to the next. Other scenarios of the evolution of language, for example that of Tomasello (2008), see the representational constraints as far less important than the social functions of communication, and ignore some features (such as the transition to combinatorial phonology). Our aim in this article is to investigate the claims of Jackendoff and others about representational abilities by first evaluating how they can be formalized; this should then provide the groundwork for more detailed investigations—which we will not carry out in this article—of every individual transition, where all of these other aspects must come into play and the roles of representational and social constraints should become clearer.

3 Modeling Meaning

Animals and humans categorize their environment, and use calls, words, or grammatical sentences to express aspects of that environment. Typically, the same utterances are used on many different occasions to express common features. It is therefore reasonable to postulate an "internal state," a representation in the brain that mediates between perceptual and motor abilities, memory, and linguistic forms. We will call these representations "conceptualizations." Modeling the conceptualizations expressed in natural language utterances is difficult because we have only very indirect access to the representations in the brain and, crucially, much of that indirect access is modulated through language (Hurford, 2003). The common formal framework for modeling conceptualization is that of symbolic logic. Many different logics exists, with different levels of expressive power as well as different computational properties.

According to Jackendoff and others, the kind of conceptualizations available to modern humans for communication are qualitatively different from those available to other primates, including our prelinguistic ancestors. Jackendoff believes that the "use of symbols" was the first major innovation; other researchers have argued that a "theory of mind" was a crucial innovation (Dunbar, 1998). It would be very useful if we could characterize such conceptual differences in formal terms, using the apparatus of formal semantics, including a well-known hierarchy of logics (e.g., Gamut, 1991): propositional, predicate, and modal logic.

This is not the place to review the properties of these logics (see Table 1 for some examples of different logics), but it is important to stress that they differ with respect to the type of generalizations they allow. Thus, while we can define propositions such as "Socrates is a man" and "Socrates is mortal" in propositional logic, we cannot describe the more general inference rule "all men are mortal." For such inferences, we need the predicate–argument distinction and quantifiers from predicate logic. Similarly, in first order predicate logic we cannot adequately deal with a sentence like *Peter believes that Gavin knows that he hates Heather* (as used in the evolutionary simulation of Kirby, 2002). The reason is that predicate logic must treat *believesthatgavinknowsthatpeterhatesheather*(x) as a single predicate that might be true for Peter. But we cannot do justice to the intended structure of the expression, that is, it would not capture the relation between the statements *Peter hates Heather* and *Gavin knows that Peter hates Heather*. For such constructions, modal logic provides a more satisfactory framework.

Let us now return to the issue of giving a more precise characterization of transitions in the evolution of language. It would be attractive if we could relate these different logics to the assumed differences between the conceptual structures available for communication to modern humans, and those available to their prelinguistic ancestors and modern higher primates. For instance, Jackendoff joins other cognitive scientists in claiming that symbol use is the first major

Table 1 Examples of different logics.

Logic	(Additional) operators	Example
Propositional	negation '¬'	$\neg R$ *"it doesn't rain"*
	AND '∧'	$\neg R \wedge \neg W$ *"it doesn't rain and the streets aren't wet*
	OR '∨'	$R \vee \neg W$ *"it either rains, or the streets aren't wet"*
	implication '→'	$R \rightarrow W$ *"if it rains, the streets are wet"*
	bi-implication '↔'	
1st order predicate	universal quantifier '∀'	$\forall x\ (H(x) \rightarrow M(x))$ *"All men are mortal"*
	existential quantifier '∃'	$\exists x \neg M(x)$ *"There exists someone immortal"*
Modal	necessity '□'	$\Box(R \rightarrow W)$ *"if it rains, the streets are always wet"*
	possibility '◊'	

innovation. He does, however, make it quite clear that he believes that apes do have a human-like system of thought:

> I take it as established by decades of primate research [references omitted] that chimpanzees have a combinatorial system of conceptual structure in place (Jackendoff, 2002, p. 238).

The crucial difference, for Jackendoff, is that between human and primate *use of symbolic vocalizations*:

> [Even] 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 a 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 [references omitted]. (Jackendoff, 2002, p. 239)

Can we express this intuitive difference between humans and other primates as a difference in representational capacity similar to the difference between propositional and predicate logic? We can, of course, conjecture that humans have words for predicates and other words for objects (the arguments of those predicates), which can thus be used in all situations where the conceptual system uses that predicate or that object. Primates, on the other hand, can only vocalize complete propositions, even if they, as Jackendoff states, do have a "combinatorial system of conceptual structure."

The problem with such a proposal is that it is not clear a priori what constitutes a predicate or an object, and thus what constitutes situation-specificity in terms of the formal tools presently at our disposal. How can we be sure that the word *kitty* in an infant's one-word stage does not mean a complete proposition such as "There is a kitty involved"? This point is all the more relevant since research into first language acquisition in recent years suggests that young children's early utterances are often of a "holistic" nature. How do we know the child does not simply categorize situations as those that involve kitties, and those that do not, much like a monkey that categorizes situations as those that require running into a tree and those that do not? If so, the difference is categorization, not representational ability. The fact that two different animal species—with different anatomy and evolved for different habitats—categorize the world differently is no surprise, of course.

Conversely, how can we be sure that the meaning of a primate alarm call for leopards is not the predicate "being a leopard"? The point here is that with regard to "meanings available for communication," the difference between propositional and predicate logic only shows itself through the rules of combination, that is, through the generalizations they allow. Of course, it is likely that there is something special about the way humans categorize their environment which was crucial for the evolution of language. But the tools of formal semantics do not appear to be useful for characterizing that difference. Put yet another way: it only makes sense to talk about conceptualizations of an agent (adult, child, or animal) in terms of a particular logic if there is behavioral evidence about the types of generalizations that agent makes, which correspond to generalizations that the particular logic allows. In the one-word stage no such evidence can exist, it seems, and the issue of the best logic becomes meaningless.

That leaves us with the conclusion that in terms of this formal system for characterizing conceptualizations, the use of symbols (Jackendoff's first innovation) cannot be recognized (or perhaps even exist) independently from the fourth innovation (concatenation of symbols). Perhaps the distinction between predicate and modal logic will prove more useful for characterizing the difference between human and nonhuman thought. A debate exists about whether, and if so to what extent, great apes have thoughts about the thoughts of others, that is, have a theory of mind (Heyes, 1998; Tomasello, 2008).[1] Such embedded conceptualizations cannot be modeled with predicate logic. An interesting question is whether the ability for embedded conceptualizations (*I think that she heard that he said …*) is a prerequisite for hierarchical phrase structure (Dunbar, 1998; Worden, 1998), and to what extent it is the other way around (de Villiers & de Villiers, 2003; Lohmann & Tomasello, 2003). In view of such recent debates, it may very well be that the evolution of this aspect of human thought and its relation to language is an important factor to take into account (cf., Hinzen & van Lambalgen, 2008; Verhagen, 2008). However, because it plays no role in Jackendoff's scenario, this issue will not be further elaborated in this article.

4 Modeling Sound

The mechanisms of sound production and perception in primate and human communication are fortunately more amenable to empirical observation, and there is therefore more of a consensus about the fundamental principles. Sounds are often analyzed by decomposing them into (infinitely) many sine waves of different frequencies, each with a particular amplitude and phase, such that when all these sine waves are added together the original signal is recovered. This is *Fourier analysis*; a graph showing, for a range of frequencies, the amplitude of the corresponding sine waves is called the frequency spectrum.

For both the production and the perception of sounds, the frequency spectrum has a natural interpretation. Sound production, both in humans (as worked out by Johannes Mueller in the 19th century; see Coren, Ward, & Enns, 1994) and many other mammals (Hauser & Fitch, 2003) can be seen as a two-stage process with a vibration source and subsequent filtering (the source–filter model, Fant, 1960). The vibrations are produced by the air flow from the lungs passing the larynx. This sound then propagates through the throat, mouth, and nose (the vocal tract), where specific frequencies are reinforced through resonance.

The frequency spectrum also maps in an important way onto sound perception. When a sound wave reaches the ear, it sets in motion a cascade of vibrations of the eardrum, hammer, anvil and stirrup, oval window, and finally the endolymph fluid in the cochlea. These vibrations cause mechanical waves to travel through the cochlea's fluid. Because of the special shape of the cochlea, the traveling waves reach their maxima at different places along the cochlea's membrane (the "basilar membrane") for each different sound frequency (Coren, Ward, & Enns, 1994; von Bekesy, 1960). These differences in wave form are then translated into different neural activation patterns in the organ of Corti. In this way, the mammalian auditory system decomposes an incoming sound wave into its component frequencies, not unlike (although there are important differences) the Fourier analysis performed by phoneticians.

The frequency spectrum is thus a representation of speech sounds that is meaningful for analyzing both production and perception. However, the frequency spectrum representation abstracts out the time dimension. Temporal changes in the frequency distribution are crucial for encoding and decoding information into sound in both human and animal communication. On the articulatory side, changes in the frequency distribution correspond to movements of articulators in the vocal tract; such movements are crucial for producing consonants and diphthongs.

From a comparative perspective, the basic principles of sound perception and production (at least at the level of physiology of articulators) appear to be very similar across humans and other mammals. In contradiction of the "speech is special" hypothesis (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), recent evidence points to the conclusion that human speech perception is not fundamentally different from non-speech and nonhuman perception (Hauser, 2001; Hauser & Fitch, 2003). The fact that humans are extraordinarily good at perceiving speech sounds appears to be better explained by the observation that, unlike many animal communication systems, human language phonology is learnt and imitated (Nottebohm, 1976; Studdert-Kennedy, 1983); in the cultural transmission from one generation to the next, languages themselves have evolved to exploit the peaks in performance of the human auditory (and articulatory) systems (Zuidema, 2005).[2]

However, when analyzing the temporal structure of a repertoire of signals, a crucial difference between human and nonhuman primate communication is noted: human speech is combinatorial, that is, the basic meaningful units in human language (words, morphemes) can be analyzed as combinations of segments from a small set of basic speech sounds (which, by definition, are not meaningful themselves). Semantic vocal animal communication (in the sense of Seyfarth, Cheney, & Marler, 1980), in contrast, seems to be acoustically holistic; that is, the basic meaningful units (calls) cannot generally be decomposed in segments that are reused in other calls. There is some controversy about what the basic segments of human speech are (phonemes, syllables, or articulatory gestures), and there are many examples of combinatorial songs in primates, birds, and cetaceans (that are used as sexual display, or for identification; e.g., see Bradbury & Vehrencamp, 1998). To our knowledge, no quantitative comparison of the degree of reuse in human and nonhuman vocal communication systems exists. Nevertheless, the intuition that human language exploits this mechanism to an unparalleled degree is widely shared and uncontroversial. It is the third innovation in Jackendoff's list.

At this point it is important to make a distinction between "E-language," a collection of externally observable utterances and any regularities in it (which can be registered in a public grammar of the language), and "I-grammar," the system internal to a language user that underlies his competence to produce and interpret instances of the E-language.[3] Combinatoriality in the I-grammar can be characterized by defining the basic units and the rules of combination; combinatoriality in the E-language is easily characterized by specifying a combinatorial I-grammar that could underlie it. However, the fact that an outside observer can analyze a set of signals as combinatorial, does not necessarily imply that language users actually exploit that combinatorial potential. For example, a repertoire of sounds might superficially look combinatorial, but in fact not be *productively* combinatorial.

The same uncertainty holds, of course, for other subsystems of language, although it is not always widely recognized. Jackendoff (2002), like many other linguists, makes the same type of distinction between productive and semi-productive morphology, but he does not generalize this distinction to the other combinatorial systems in language, nor does he discuss its relevance for evolution. To a large extent the methodological part of controversies over the combinatorial nature of children's competence among investigators of language acquisition boils down to this issue: whereas some take an observable regularity in the linguistic output of a child as evidence for combinatoriality in its underlying I-grammar, others are only prepared to draw such a conclusion when the child is using a pattern productively itself (Tomasello, 2000).

Returning to the sound system of human language, if we accept that the syllable, and not the phoneme, is the unit of productive combination in human speech, then the I-grammar consists of a set of syllables and the rules of combining them. Phonemes, in such a view, are patterns in the E-language that look *as if* there is a combinatorial system underlying it; they are only superficially combinatorial. This distinction is relevant for the evolution of language, because a superficially combinatorial stage, as an effect of one or more other processes and constraints of production and perception, might precede and facilitate the evolution of *productive* combinatoriality in the I-grammar of individuals developing a representation of the E-language in their environment (see Zuidema & de Boer, 2009).

Thus, the evolutionary origins of combinatorial phonology are still a largely open question. A widely shared intuition is that the way to encode a maximum of information in a given time frame such that it can be reliably recovered under noisy conditions, is by means of a digital code. Hence, phonemic coding could be the result of selection for perceptual distinctiveness. However, this argument has, to our knowledge, never been worked out decisively for human speech (see Zuidema & de Boer, 2009, for a critique of existing formal models, such as the one of Nowak & Krakauer, 1999, and for an alternative proposal).

Alternatively, combinatorial coding could be the result of articulatory constraints. Studdert-Kennedy (1998, 2000) has argued that the inherent difficulty of producing complex sounds like *through* makes the reuse of motor programs unavoidable. Hence, the combinatorial nature of speech follows from the difficulty of production and the large repertoire of words in human languages. Consistent with this view is Deacon's (2000) claim that humans have unique forebrain-based control over articulators that shows evidence of intense selection for precisely timed phonation. If these arguments are correct, Jackendoff's third innovation is characterized by radical changes in articulatory motor control. But notice that this innovation is still seen as being driven by the need for a large repertoire of perceptually distinct signals, albeit under stringent articulatory constraints. It therefore seems safe to conclude that this pressure has at least been an important factor in the evolution of combinatorial phonology.

5 Modeling Simple Sound—Meaning Mappings

The other transitions in Jackendoff's scenario (numbers 2 and 4–10) all concern the way messages are mapped onto signals. Most existing formalisms in linguistics already assume the innovations that Jackendoff proposes: word order, compositionality, phrase structure, grammatical categories. They are therefore not of much use in characterizing the early transitions. Here we will first develop a simple formalism that describes meaning to form mappings without any assumptions about learning or combination; from that basis we will try to characterize the innovations proposed.

Given a set of relevant messages to express and a set of distinctive signals (i.e., sounds, or "forms") that can be produced and perceived, we can describe a communication system as a (probabilistic) mapping from messages to signals (in production), and from signals to messages (in interpretation). These mappings can be represented with matrices. Hence, we have a production matrix **S** and an interpretation matrix **R**. **S** gives for every message m and every signal f, the probability that the individual chooses f to convey m. Conversely, **R** gives for every signal f and message m, the probability that f will be interpreted as m. If there are M different messages and F different signals, then **S** is an M × F matrix, and **R** an F × M matrix. Variants of this notation are used by Hurford (1989), Oliphant & Batali (1996), and other researchers.

Many different **S** and **R** matrices are possible. How can we measure the quality of specific combinations? Or, in biological terms, how can we calculate the payoff (a fitness contribution) of specific **S** and **R** matrices? An important component of such a payoff function is whether speakers and hearers agree on which signals are systematically associated with which (recurrent features of) messages. However, in many cases the similarities between signals also need to be taken into account (because more similar signals are more easily confused), as well as the similarities between messages (because slightly wrong interpretations are often better than totally wrong ones).

The consequences of such similarities can be modeled with a confusion matrix **U** (of dimension F × F), which gives for each possible signal the probability that it is perceived correctly or as any of the other signals, and with a value matrix **V** (of dimension M × M), which gives for every intended message the payoff of each of the possible interpretations. Typically, **U** and **V** will have relatively high values on the diagonal (the correct signals and interpretations).

Together, these four matrices can describe the most important aspects of a communication system: which signals are used for which messages by hearers and by speakers, how likely it is that signals get confused in the transmission, and what the consequences of a particular successful or unsuccessful interpretation are. This notation is a generalization of the notation in Nowak and Krakauer (1999), and was introduced in Zuidema and Westermann (2003).

A hypothetical example, loosely based on the celebrated study of vervet monkey alarm calls (Seyfarth et al., 1980; Seyfarth & Cheney, 1997), will make the use of this formalism clear.[4] Imagine an alarm call system of a monkey species for three different types of predators: from the air (eagles), from the ground (leopards) and from the trees (snakes). Imagine further that

$$U = \left(\begin{array}{c|ccccc} & \multicolumn{5}{c}{\text{received signal}} \\ \text{sent signal} \downarrow & 16kHz & 8kHz & 4kHz & 2kHz & 1kHz \\ \hline 16kHz & 0.7 & 0.2 & 0.1 & 0.0 & 0.0 \\ 8kHz & 0.2 & 0.6 & 0.2 & 0.0 & 0.0 \\ 4kHz & 0.0 & 0.2 & 0.6 & 0.2 & 0.0 \\ 2kHz & 0.0 & 0.0 & 0.2 & 0.6 & 0.2 \\ 1kHz & 0.0 & 0.0 & 0.1 & 0.2 & 0.7 \end{array}\right) \qquad V = \left(\begin{array}{c|ccc} & \multicolumn{3}{c}{\text{interpretations}} \\ \text{intentions} \downarrow & \textit{eagle} & \textit{snake} & \textit{leopard} \\ \hline \textit{eagle} & 0.9 & 0.2 & 0.1 \\ \textit{snake} & 0.5 & 0.9 & 0.5 \\ \textit{leopard} & 0.1 & 0.2 & 0.9 \end{array}\right)$$

Figure 1 Confusion and value matrices for the monkeys in the example, describing the noise in signaling and the value of intention–interpretation pairs in their environment.

the monkeys are capable of producing a number of different signals (say 5) that range on one axis (e.g., pitch, from high to low) and that these are more easily confused if they are closer together. Thus, the confusion matrix **U** might look like the left matrix in Figure 1.

Further, although it is obviously best to interpret a signal correctly, if one makes a mistake, typically not every mistake is equally bad. For example, if a leopard alarm is given, the leopard response (run into a tree) is best, but a snake response (search surrounding area) is better than an eagle response (run into a bush, where leopards typically hide; Seyfarth & Cheney, 1997). Thus the value matrix **V** might look somewhat like the right matrix in Figure 1.

Now, assume a speaker i with her S^i as the left matrix in Figure 2, and a hearer j with his R^j as the right matrix in that figure. What will happen in the communication between i and j? One possibility is that (a) the speaker sees an eagle, (b) she sends out the 16 kHz signal, (c) the hearer indeed perceives this as a 16 kHz signal, and (d) he correctly interprets this signal as “eagle.” Given that the observation is an eagle, the contribution to the expected payoff is the probability that the remaining steps happen (that is, the product of the conditional probabilities of individual events) times the reward in such a situation (the “value” from matrix **V**):

$$P_S(\text{16 kHz sent} \mid \text{observation=eagle})$$

$$\times\, P_U(\text{16 kHz perceived} \mid \text{16 kHz sent})$$

$$\times\, P_R(\text{eagle interpreted} \mid \text{16 kHz perceived})$$

$$\times\, V(\text{eagle interpreted, eagle observed})$$

$$= 1 \times 1 \times .71 \times .9 = 0.63.$$

$$S = \left(\begin{array}{c|ccccc} & \multicolumn{5}{c}{\text{sent signal}} \\ \text{intention} \downarrow & 16kHz & 8kHz & 4kHz & 2kHz & 1kHz \\ \hline \textit{eagle} & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ \textit{snake} & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ \textit{leopard} & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{array}\right) \qquad R = \left(\begin{array}{c|ccc} & \multicolumn{3}{c}{\text{interpretation}} \\ \text{received signal} \downarrow & \textit{eagle} & \textit{snake} & \textit{leopard} \\ \hline 16kHz & 1.0 & 0.0 & 0.0 \\ 8kHz & 1.0 & 0.0 & 0.0 \\ 4kHz & 0.0 & 1.0 & 0.0 \\ 2kHz & 0.0 & 0.0 & 1.0 \\ 1kHz & 0.0 & 0.0 & 1.0 \end{array}\right)$$

Figure 2 Production and interpretation matrices for the monkeys in the example, describing which signals they use for which messages.

Another possibility, with probability 0.2, is that the hearer misperceives the signal as an 8 kHz signal, but with probability 1 still interprets it correctly. We can thus work out all possible scenarios and find that the expected payoff w_{ij} of the interaction between *i* and *j*, given the constraints on communications as in **U** and **V** in Figure 1, is:

$$w_{ij} = .7\ .9 + .2\ .9 + .1\ .2 + .2\ .5 + .6\ .9$$

$$+ .2\ .5 + .1\ .2 + .2\ .9 + .7\ .9 = 2.4.$$

More generally, such a calculation can be expressed by one simple expression for the expected payoff w_{ij} of communication between a speaker *i* with production matrix S^i and a hearer *j* with interpretation matrix R^j (Zuidema & Westermann, 2003):

$$w_{ij} = \mathbf{V} \bullet (\mathbf{S}^i \times (\mathbf{U} \times \mathbf{R}^j)).$$

In this formula, "×" represents the usual matrix multiplication and "•" represents dot-multiplication (the sum of all multiplications of corresponding elements in both matrices; the result of dot-multiplication is not a matrix, but a scalar). In this simple example, the matrices **U** and **V** are very small, and reflect only a 1-dimensional topology in both signal and conceptual space. The matrices **S** and **R** are set by hand to arbitrarily chosen values.

Note that the **S** and **R** matrix describe the production and interpretation behavior of an individual (the E-language it has access to), but they do not necessarily model the mechanism that the individual uses to map messages onto signals and vice versa (its I-grammar). In the present approach, the values can even be chosen in such a way that an individual's **S** matrix is incompatible with her own **R** matrix, that is, that she cannot understand her own utterances. It is more realistic, perhaps, to assume an underlying lexicon of (bidirectional) associations between meanings and signals (Komarova & Niyogi, 2004; Steels, 1995). Such associations can be modeled with an association matrix **A**. **S** and **R** are then functions of **A** such that, for instance, an element in **S** is 1 if the corresponding element in **A** is the highest in its row, and 0 otherwise. Similarly, an element in **R** can be set to 1 only if the corresponding element in **A** is the highest in its column.

Jackendoff's second innovation—an open, unlimited class of symbols—can now be viewed as the evolution of a learning procedure to set the values of the elements in **S** and **R** or in **A**. Assume that the set of possible relevant messages and the set of possible signals are determined by an individual's habitat and anatomy and can be defined a priori. An innate, closed call system then corresponds to settings of the elements of the matrices that are not dependent on input and show no variation; conversely, a learnt, open call system corresponds to settings that do depend on environmental input and that vary with varying inputs. Human language clearly is an open system, where the meanings of words are not naturally given, but rather emerge as conventions in a population of language users (Gamut, 1991; Keller, 1998; Lewis, 1969). Conventions are "negotiated" in a population, as is studied in models by Hurford (1989) and others. The main results from these studies is that (a) a stable communication system can emerge in a population where everybody learns from everybody else, without a need for central control (Steels, 1995); (b) the best "response language" is not necessarily the same as the current language in the population (Hurford, 1989; Komarova & Niyogi, 2004); and (c) Saussurean learners (where **S** matrices are modeled after **R** matrices) and synonymy and homonymy avoiders outcompete other learning strategies (Hurford, 1989; Oliphant & Batali, 1996; Smith, 2004).

These studies are interesting, but do not really address Jackendoff's transition from a closed, innate vocabulary to an open, learnt vocabulary. The selective advantages of such a transition—to what biologists call "phenotypic plasticity"—depend on the constraints on the innate system, the properties of the environment and the accuracy of the learning mechanism. If a learnt vocabulary can contain more signals than an innate vocabulary—as Jackendoff asserts—that must be because of biological constraints preventing the innate system being as large. Moreover, a learnt vocabulary can be easily extended to include a word for a new concept, but whether or not this confers an advantage depends on how often such new relevant concepts appear. These are interesting issues, but it is difficult to tell what reasonable assumptions are. Oliphant (1999) argues quite convincingly that it is unlikely that computational demands of learning have been the limiting factor in this transition; rather, he argues, the difficulty of identifying what meaning a signal is meant to convey explains why learnt communication systems are so rare in nature.

In conclusion, we agree with Jackendoff (2002), Oliphant (1999), and many others that the emergence of an open class of symbols is an important transition in the evolution of language. Moreover, we believe it can be modeled formally using the matrix notation introduced above. Many models that use such a formulation in one form or another have already been studied, showing that specific settings of such models have specific consequences for the evolution of combinatoriality, both in phonology and in semantics (cf., Zuidema, 2005).

6 Modeling Compositionality

The matrices discussed above can describe, for each particular meaning, which signals are associated with it or vice versa. However, they cannot make explicit any regularity in the mapping from meanings to signals. In both nonhuman and human communication such systematic relations between meanings and signals exist. For instance, in most species high pitch sounds are associated with danger and low pitch sounds with aggression. In vervet monkey calls, there are clear similarities between the various calls used in social interactions, which are all very different from the alarm calls (Seyfarth & Cheney, 1997). In human language, on the level of words, there is some evidence—albeit controversial—that similar words tend to refer to similar objects, actions, or situations, and that humans generalize such patterns to nonsense words (Hinton, Nichols, & Ohala, 1995). For the level of morphosyntax it is in fact uncontroversial that similar phrases and similar sentences generally mean similar things,[5] that is, the mapping from messages to signals is compositional.

In Section 4, we introduced the distinction between E-language, the collection of externally observable utterances, and I-grammar, the individual linguistic system internal to a language user. We observed that an E-language may exhibit regularities that need not be determined by (productive) rules of an I-grammar (the fact that an outside observer can analyze a set of signals as combinatorial, does not necessarily imply that language users actually exploit that combinatorial potential). An important reason why this distinction is relevant in the context of evolution is that a superficially combinatorial stage of E-language may precede and facilitate the evolution of *productive* combinatoriality in the I-grammars of individuals. Obviously, this possibility is not only relevant in the domain of combinatorial phonology, but also in the present domain of semantic compositionality.

This should be kept in mind when we proceed to sketch a formalism for modeling compositionality. What this will amount to is essentially a way to characterize regularities that can be observed in an E-language. If this can be done, then the question of whether and how this may have led to the evolution of productive compositionality in the I-grammars of language users can at least be addressed.

We can describe the systematicities in the meaning–signal mappings as the preservation of topology between meaning space and signal space, that is, meanings that are close are expressed with signals that are close. In the **S**, **R**, and **A** matrices, such "topology preservation" might be noticeable as a nonrandom pattern if both the meaning and signals axes are ordered. We can be more precise, however, by systematically comparing each pair of associations. Brighton (2002) proposes using the correlation ("Pearson's r") between the distance between each pair of meanings and the distance between the corresponding signals:

$$\mathrm{r} = \mathrm{correlation}(\mathrm{D}(m, m'), \mathrm{D}(\mathrm{S}[m], \mathrm{S}[m'])),$$

$$\text{calculated over all } m, m' \neq m \in M,$$

where $\mathrm{S}[m]$ gives the most likely signal used to express m according to S, $\mathrm{D}(m, m')$ gives the distance between two meanings m and m', and $\mathrm{D}(f, f')$ between two signals f and f'. Although only a correlate of compositionality, such a measure can reveal a tendency for related meanings to be expressed with related signals. Hence, expressed in this formalism, Jackendoff's fourth and fifth innovation (concatenation and compositionality) correspond to high values of r in this equation.

We can go further, however, and explicitly model the way in which combinations of signs form more complex signs. The common way to deal formally with the meanings of combinations of lexical entries, is Church's lambda calculus (see e.g., Gamut, 1991, for a discussion). Semantic descriptions, such as discussed in Section 3, should be extended with the possibility to include lambda (λ) terms. Lambda terms can be seen as listing the variables that still need to be substituted; they disappear when a complete semantic description is reached. Formally, a lambda term in front of an expression turns that expression into a function that maps an argument onto a new expression where that argument has found its proper place. For instance, we can model the semantics of the verb walks as λx W(x) and apply it to an argument j (for *John*) to yield W(j) (for *John walks*).

The lambda calculus gives a mechanical procedure to derive the semantic expression that results from applying a function to an argument. A word (or phrase) corresponding to the function is said to dominate a word corresponding to the argument. Hence, if we model the compositional semantics of *John walks* with a function λx W(x) and an argument j, then we have assumed that *walks* dominates *John*.

In modern languages, this dominance structure is, to a large extent, determined by principles of word order and morphological marking. Thus, if we model the meaning of the transitive verb *hates* in *George hates broccoli* as λy λx H(x,y) (i.e., as a function with two arguments), the principles of word order need to guarantee that *hates* dominates *broccoli*, and *hates broccoli* dominates *George*. In the fourth and fifth stage of Jackendoff's scenario there is no phrase structure or morphological marking, so the dominance structure is largely underdetermined. The word order principles of "agent first," "focus last," and "grouping" that Jackendoff proposes, constrain the structural ambiguity that arises from this underdeterminacy.

In conclusion, correlation r gives us a provisional measure of compositionality in the E-language. Moreover, we can characterize compositionality in the I-grammar by identifying the units and rules of combination. Zuidema (2003) studies the transition to compositional semantics using the former, and argues that the compositionality in I-grammars can more easily evolve if some form of compositionality in the E-language has already been established for other reasons.

7 Modeling Hierarchical Phrase Structure

One of the defining characteristics of human language is that sentences exhibit phrase structure, and the ability to represent phrase structure has, since Chomsky (1957), been the one of the most important criteria in judging the adequacy of linguistic formalisms. Theories that do not precisely adhere to Chomsky's specific view of phrase structure or even regard it as derivative (including, for example, Langacker, 1997), still all recognize the hierarchical character of structure in one way or another (with phrases consisting of identifiable smaller parts, and also a possible part of larger phrases). Rewriting grammars, such as proposed by Chomsky, remain the archetype formalism for describing syntax. We will first introduce this formalism in some detail, then define phrase structure in terms of it, and then evaluate its usefulness for characterizing the transition to hierarchical grammatical structure in the evolution of language.

Rewriting grammars are specified by four sets of elements: the terminal symbols, the nonterminal symbols, the production rules, and the start symbols. Terminal symbols (V_t) are in a sense the atoms of a language; they do not exhibit further syntactic structure (e.g., the words or morphemes of a language, but possibly also complete idioms or frozen expressions). Nonterminal symbols (V_{nt}) are variables that stand for more categories that define the constituents of a sentence; they can correspond to anything from the syntactic category of a word (or morpheme) such as *N*(*oun*) of *V*(*erb*), a phrasal category such as *PP* ("Prepositional Phrase"), to a whole sentence (S). Production rules (R) specify which nonterminal symbols can be replaced by which terminal or nonterminal symbols in the process of deriving a sentence, starting with a start symbol (S). If the production rules are of the form $A \rightarrow w$ where the left-hand side is a nonterminal symbol (V_{nt}), and the right-hand side is any (non-null) string of terminals and nonterminals, the grammar is said to be context free (because the context in which A occurs is not relevant for the applicability of the rule). Figure 3 gives an example context-free rewriting grammar for a fragment of English.

Chomsky (1957) showed that more restricted versions of this formalism such as finite-state grammars or their probabilistic version, Markov processes, are unable to describe the long-range dependencies that are observable in natural languages. He further argued (for the wrong reasons, it emerged later, cf., Pullum & Gazdar, 1982) that even context-free grammars are not powerful enough to model certain phenomena in language, for which he proposed using *transformations.* Through this analysis, Chomsky discovered a fundamental hierarchy of formal grammars that is now termed the Chomsky hierarchy (see Table 2). This prompted a long debate on where to locate human lan-

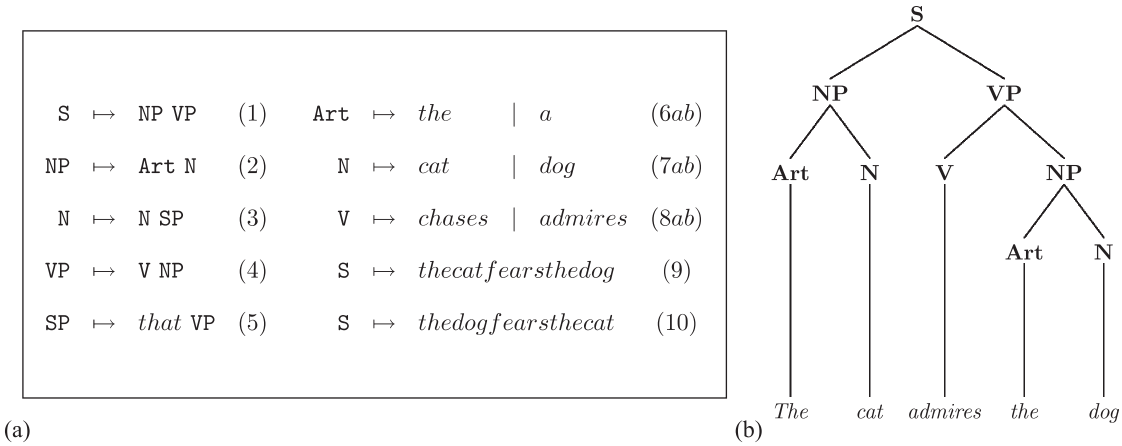
S ↦ NP VP (1)
NP ↦ Art N (2)
N ↦ N SP (3)
VP ↦ V NP (4)
SP ↦ *that* VP (5)
Art ↦ *the* | *a* (6*ab*)
N ↦ *cat* | *dog* (7*ab*)
V ↦ *chases* | *admires* (8*ab*)
S ↦ *thecatfearsthedog* (9)
S ↦ *thedogfearsthecat* (10)
(a)
S
NP
VP
Art
N
V
NP
Art
N
The
cat
admires
the
dog
(b)

Figure 3 (a) An example context-free grammar for a fragment of English, with a terminal alphabet V_t = {the, a, dog, chases, admires, that} and a nonterminal alphabet V_{nt} = {S, NP, VP, Art, N, V}. Production of a sentence ("derivation") always starts with the symbol S and proceeds by replacing symbols matching the left-hand side of some rule with the string at the right-hand side of that rule. "Parsing" means searching the sequence of rewriting steps that would produce the sentence with a given grammar. Rules 1–4 are "combinatorial," rule 5 is "recursive." The grammar can generate infinitely many sentences such as "the cat chases the dog" or "a dog admires a cat that chases a dog that admires a cat," and so forth. Rules 6–8 constitute what is traditionally described as the lexicon, and can be represented in the same formalism. Rules 9 and 10 illustrate a "lexical," noncombinatorial and much less efficient strategy for generating sentences. Context-free grammars are considered to be not quite sufficiently powerful to describe natural languages. The formalism can be extended in several ways. For instance, it can be extended to attribute in a systematic way meanings to words and sentences; the resulting system is "compositional." (b) A parse tree that can be generated by the given context-free grammar.

Table 2 Rewrite grammars.

Definition (Chomsky hierarchy) A grammar G = < P, S, V_{nt}, V_{te} > is classified according to the following restrictions on the form of rewriting rules of P:

- A grammar is of type 3 (the "right-linear" or "regular grammars") if every rule is of the form A → b C or A → b, where A, C ∈ V_{nt}, and b ∈ V_{te}^* or b = λ (the "empty" character).
- A grammar is of type 2 (the "context-free grammars") if every rule is of the form A → w, where A ∈ V_{nt}, and w ∈ $(V_{nt}\ V_{te})^*$.
- A grammar is of type 1 (the "context-sensitive grammars") if every rule is of the form vAw → vzw, where z is any combination of terminal or nonterminal symbols: z ∈ $(V_{nt}\ V_{te})^*$ and z ≠ λ. In addition, a single rule S→λ is allowed if S does not appear at any right-hand side of the rules.
- Any rewriting grammar, without restrictions, is of type 0.

This classification constitutes a strict hierarchy of languages. Hence, if L_i is the set of all languages of type i, then the following is true: $L_3 \subset L_2 \subset L_1 \subset L_0$.

guage on the hierarchy. Already in the 1960s it was realized that the original transformational grammars were too powerful, because they made necessary a long list of rather ad hoc constraints and exceptions to exclude obviously ungrammatical sentences.

There seems to be a reasonable consensus now that the necessary level of generative power is slightly more than context free, a level now termed "mildly context sensitive" (Joshi, Vijay-Shanker, & Weir, 1991). The additional power over context-free grammars is needed for constructions such as the crossed dependencies in the Dutch example (1) below. Examples (2) and (3) show the translation in English and German. Different fonts are used to show the different types of dependencies: crossing dependencies in Dutch, local dependencies in English, center-embedding in German.

1. Gilligan beweert dat **Kelly** *Campbell* <u>Blair</u> het publiek **zag** *helpen* <u>bedriegen</u>.
2. Gilligan claims that **Kelly saw** *Campbell help* <u>Blair deceive</u> the public.
3. Gilligan behaupte dass **Kelly** *Campbell* <u>Blair</u> das Publikum <u>belügen</u> *helfen* **sah**.

From an empirical perspective, it should perhaps be noted that such phenomena are relatively rare. There are many languages that do not at all exhibit crossing dependencies of this kind (which are beyond the scope of context-free grammars), and even in a language like Dutch that does seem to exemplify them, the phenomenon is actually rare, both in terms of the number of rules in the grammar (this type of embedding of nonfinite verbs is the only one), as in terms of its occurrence in actual language use—to the extent that many native speakers find examples such as (1) unacceptable.

However, there are more principled reasons to raise doubts about the relevance of the boundary between context-free and context-sensitive grammars for understanding the evolution of human language. Given a formal definition of complexity in terms of the Chomsky hierarchy, and a consensus about where modern human language should be situated, it is perhaps natural to try to describe the evolution of language as a climb of that hierarchy. In such a scenario, selection for increased computational power one by one removed the computational constraints for dealing with the full complexity of modern language. An explicit example of such a scenario is Hashimoto and Ikegami (1996), but it is implicit in many other accounts (e.g., Fitch & Hauser, 2004). However, there are a number of problems with such attempts.

First, we have to be very careful with what we mean by phrases such as "at least context-free power," "human language syntax is mildly context sensitive," or "where human language is situated on the Chomsky hierarchy." The classes of formal languages on the Chomsky hierarchy are subsets of each other. Chomsky's (1957) analysis that finite state grammars are insufficient, and subsequent analysis that context-free grammars are also insufficient, suggests that natural languages are in that subset of the context-sensitive languages that cannot be modeled with a finite-state grammar or context-free grammars (that is, in the complement of the context-free languages within the set of context-sensitive languages). Most context-sensitive languages, however, such as the standard example $a^n b^n c^n$ (a language consisting of strings like aabbcc and aaaabbbbcccc, i.e., strings with an arbitrary but equal number of a's, b's, and c's) have very little in common with natural languages; natural languages are thus constrained in many ways (e.g., semantics, learnability) that have nothing to do with the Chomsky hierarchy.

Second, it would be wrong to assume that complexity in terms of the Chomsky hierarchy is actually hard to get. Just a few neurons connected in a specific way can generate temporal patterns that are of type 1 or 0 in the Chomsky hierarchy (i.e., that can only be described with context-sensitive or Turing complete grammars; continuously valued activation levels then implement the unbounded memory typical for type 0 grammars). Like natural languages, such patterns would justify the label "at least context sensitive," even though they are not likely to be interesting from the point of view of encoding information. In short, the classes of the Chomsky hierarchy divide up the space of formal grammars in a way that is not particularly relevant for the evolution of language. That is, it is possible that most of the evolutionary developments of natural language grammar occurred *within* one and the same class of the Chomsky hierarchy. Moreover, even if a class boundary was crossed, formalization in terms of the Chomsky hierarchy and architectural constraints offer no insights about the causes for crossing it.

The question is therefore: Are there ways to divide up the space of formal grammars that do suggest an incremental, evolutionary path to the complexity of modern language? A starting point for answering that difficult question should be, we suggest, a precise model of how natural language is learnt. Language learning is a peculiar learning problem. Languages are population level phenomena, and the mechanisms of their reproduction crucially include cultural transmission, which can lead to a process of cultural evolution. It is plausible that the incremental evolution of the human capacity for language can only be understood as a co-evolution of languages and the brain (see also Deacon, 1997), with humans adapting to communicative pressures as much as languages adapting to human learners and in the process acquiring certain structural characteristics (allowing a feature to pass through the learning bottleneck better than another one).

8 Conclusions

Humans show in their communication system a number of "design features" that distinguish us from nonhuman primates and, by assumption, from our prelinguistic ancestors. Jackendoff's list of innovations in the evolution of languages provides a useful framework to address the origins of these design features. Jackendoff's account can and should, however, be improved in a number of ways.

First, we have argued that although a scenario with successive stages is an important ingredient of an evolutionary explanation, Jackendoff does not address the important other ingredient: How did the transitions happen? Evolutionary explanations require a plausible account of how innovations spread in the population. Alternative scenario's emphasizing the social/pragmatic nature of human communication such as the one elaborated in Tomasello (2008), may be especially relevant here (even if these in turn lack mechanisms for some specific features, such as learnt combinatorial phonology).

Second, although Jackendoff makes liberal use of diagrams, trees, and logical formulae, his account is not precise enough to be implemented in formal models. In this article we have tried to sketch the formal tools available to describe evolutionary innovations in conceptualizations, sounds, and the mapping between them. From that discussion it has become clear that Jackendoff's first innovation, the use of symbols, cannot be precisely defined. In contrast, combinatorial phonology, compositional semantics, and hierarchical phrase structure can be precisely characterized. Modeling studies of these innovations may thus be expected to be able to advance our understanding of the transitions involved.

Third, Jackendoff's evolutionary scenario does not make a distinction between the structure of the language as observed from "outside" (E-language), a population level phenomenon, and the structure of the representations used in an individual's brain (I-grammar). As we have seen in this article, it is possible for a language, that is, an observable set of utterances, to show the hallmarks of combinatorial phonology, compositional syntax, and perhaps phrase structure, without the language user necessarily being able to actively exploit them. We therefore propose to extend the evolutionary scenario with (at least) this hypothetical transitional step. That is, we propose to strictly distinguish between E-language and I-grammar in the following way (cf., Croft, 2000):

1. An E-language is a "population" of observable utterances, distributed over several individuals (as a consequence of communication processes between members of a human community).
2. An I-grammar is an individual's representation of linguistic units and rules that results from learning in communication and that in turn underlies the individual's contribution of new utterances to the E-language.

As argued by Croft (2000), this specific view of the difference and the relation between the two levels allows for an integrated model of language change (cf., Landsbergen, 2009). What we claim is that it will also make it possible to model the co-evolution of E-languages with their properties and the capacity to infer I-grammars, the former not only being the product of the latter, but also subject to other constraints.

Notes

[1] Not only our close relatives may have evolved these sophisticated skills of social cognition; compare Emery and Clayton, 2004.

[2] This is a point of general importance: characteristics of human languages can be the result of processes of biological evolution, of cultural evolution, or both (gene-culture-coevolution), and it is not necessarily clear a priori what the balance is.

[3] In the literature, this distinction is usually referred to as the one between E-language and I-language. The difference is not only one of external versus internal, however, and we therefore prefer our terminology: E-language concerns observable utterances, but I-language/I-grammar refers to an individual system of units and rules that in a sense *defines* a language.

[4] The actual alarm calls of vervet monkeys are very different from the ones we use in this example. For instance, eagle alarm calls are low-pitched rather than high-pitched, and all three types of alarm calls have a temporal structure. The example here is just to illustrate the use of the formalism.

[5] In functionally and cognitively oriented approaches, this is completely uncontroversial, and especially manifested in the so-called principle of iconicity. But also in more formally oriented approaches, even though it may be denied

that the same structure always conveys exactly the same meaning, some (more loosely conceived, but still systematic) connection between a structure and one or more prototypical functions is a standard assumption. Moreover, any differences of opinion are in practice confined to syntax; as far as morphology is concerned, there is general consensus.

Acknowledgments

This article is based on Chapter 3 of Zuidema (2005). We have profited greatly from comments on earlier versions by Simon Kirby, Jim Hurford, Tim O'Donnell, Bart de Boer, and two anonymous reviewers, and we are grateful to Bart de Boer and Paul Vogt for organizing the workshop that has led to our collaborative work, as partly reflected in this article.

References

Arbib, M. A. (2003). The evolving mirror system: a neural basis for language readiness. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 182–200). Oxford, UK: Oxford University Press.

Bickerton, D. (1990). *Language and species*. Chicago, IL: University of Chicago Press.

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication*. Sunderland, MA: Sinauer Associates.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*(1), 25–54.

Carstairs-McCarthy, A. (1999). *The origins of complex language: An inquiry in the evolutionary beginnings of sentences, syllables, and truth*. Oxford: Oxford University Press.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Christiansen M. H., & Kirby S. (Eds.) (2003). *Language evolution*. Oxford, UK: Oxford University Press.

Coren, S., Ward, L. M., & Enns, J. T. (1994). *Sensation and perception*. Fort Worth, TX: Harcourt Brace.

Croft, W. (2000). *Explaining language change. An evolutionary approach*. London: Longman.

Deacon, T. (1997). *Symbolic species, the co-evolution of language and the human brain*. New York, NY: The Penguin Press.

Deacon, T. W. (2000). Evolutionary perspectives on language and brain plasticity. *Journal of Communication Disorders*, *33*(4), 273–290.

de Villiers, J. G., & de Villiers, P. A. (2003). Language for thought: Coming to understand false beliefs. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind* (pp. 335–384). Cambridge, MA: MIT Press.

Donald, M. (1991).*Origins of the modern mind*. Cambridge, MA: Harvard University Press.

Dunbar, R. (1998). Theory of mind and the evolution of language. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.). *Approaches to the evolution of language: social and cognitive bases*. Cambridge, UK: Cambridge University Press.

Emery, N. J., & Clayton, N. S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science*, *306*, 1903–1907.

Fant. G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *303*(5656), 377–380.

Gamut, L. T. F. (1991). *Logic, language and meaning* (Vol. 2). Chicago, IL: The University of Chicago Press.

Hashimoto, T., & Ikegami, T. (1996). The emergence of a net-grammar in communicating agents. *BioSystems*, *38*, 1–14.

Hauser, M. D. (2001). What's so special about speech? In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler*. Cambridge, MA: MIT Press.

Hauser, M. D., & Fitch, W. T. (2003). What are the uniquely human components of the language faculty? In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 317–337). Oxford, UK: Oxford University Press.

Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, *21*, 101–134.

Hinton, L., Nichols, J., & Ohala, J. J. (Eds.) (1995). *Sound symbolism*. Cambridge, UK: Cambridge University Press.

Hinzen, W., & van Lambalgen, M. (2008). Explaining intersubjectivity. A comment on Arie Verhagen, Constructions of Intersubjectivity. *Cognitive Linguistics*, *19*, 107–123

Hockett, C. F. (1960), The origin of speech. *Scientific American*, *203*, 88–111.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, *77*(2), 187–222.

Hurford, J. R. (2000). Social transmission favours linguistic generalization. In C. Knight, J. R. Hurford, & M. Studdert-Kennedy (Eds.). *The evolutionary emergence of language: Social function and the origins of linguistic form*. Cambridge, UK: Cambridge University Press.

Hurford, J. R. (2003). The neural basis of predicate-argument structure. *Behavioral and Brain Sciences*, *26*(3), 261–283.

Jackendoff, R. (2002). *Foundations of language*. Oxford, UK: Oxford University Press.

Joshi, A., Vijay-Shanker, K., & Weir, D. (1991). The convergence of mildly context-sensitive grammar formalisms. In P. Sells, S. Shieber, & T. Wasow (Eds.), *Foundational* *issues in natural language processing* (pp. 21–82). Cambridge, MA: MIT Press.

Keller, R. (1998). *A theory of linguistic signs*. Oxford, UK: Oxford University Press.

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.) *Linguistic evolution through language acquisition: formal and computational models*. Cambridge, UK: Cambridge University Press.

Klein, W., & Perdue, C. (1997). The basic variety, or: Couldn't language be much simpler? *Second Language Research*, *13*, 301–347.

Komarova, N. L., & Niyogi, P. (2004). Optimizing the mutual intelligibility of linguistic agents in a shared world. *Artificial Intelligence*, *154*(1–2), 1–42. URL http://www.math.ias.edu/natalia/preprints/agents.ps.

Landsbergen, F. (2009). *Cultural evolutionary modeling of patterns in language change*. Utrecht: LOT Publications.

Langacker, R. W. (1997). Constituency, dependency, and conceptual grouping. *Cognitive Linguistics*, *8*, 1–32.

Lewis, D. K. (1969). *Convention: a philosophical study*. Cambridge, MA: Harvard University Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*, 431–461.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: A training study. *Child Development*, *74*, 1130–1144.

Nottebohm, F. (1976).Vocal tract and brain: A search for evolutionary bottlenecks. *Annals of the New York Academy of Sciences*, *280*, 643–649.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences of the USA*, *96*, 8028–8033.

Oliphant, M. (1999). The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior*, *7* (3/4), 371–383.

Oliphant, M., & Batali, J. (1996). Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, *11*(1), 1–46.

Oudeyer, P.-Y. (2006). *Self-organization in the evolution of speech*. Oxford, UK: Oxford University Press.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707–784.

Pullum, G., & Gazdar, G. (1982). Natural languages and context-free languages. *Linguistics and Philosophy*, *4*(4), 471–504.

Seyfarth, R. M., & Cheney, D. L. (1997). Some general features of vocal development in nonhuman primates. In C. T. Snowdon & M. Hausberger (Eds.), *Social influences on vocal development* (pp. 249–273). Cambridge, UK: Cambridge University Press.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, *210*, 801–803.

Smith, K. (2003). *The transmission of language: models of biological and cultural evolution*. Unpublished doctoral dissertation, Theoretical and Applied Linguistics, University of Edinburgh.

Smith, K. (2004). The evolution of vocabulary. *Journal of Theoretical Biology*, *228*(1), 127–142.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, *2*(3), 319–332.

Studdert-Kennedy, M. (1983). On learning to speak. *Human Neurobiology*, *2*, 191–195.

Studdert-Kennedy, M. (1998). The particulate origins of language generativity: From syllable to gesture. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: social and cognitive bases*. Cambridge, UK: Cambridge University Press.

Studdert-Kennedy, M. (2000). Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function. In C. Knight, J. R. Hurford, & M. Studdert-Kennedy (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form*. Cambridge, UK: Cambridge University Press.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209–253.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Verhagen, A. (2008). Intersubjectivity and explanation in linguistics – a reply to Hinzen & Van Lambalgen. *Cognitive Linguistics*, *19*, 125–143.

von Bekesy, G. (1960). *Experiments in hearing*. New York: McGraw-Hill.

Worden, R. (1998). The evolution of language from social intelligence. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: social and cognitive bases*. Cambridge, UK: Cambridge University Press.

Wray, A. (2002a). *Formulaic language and the lexicon*. Cambridge, UK: Cambridge University Press.

Wray, A. (2002b). Dual processing in protolanguage: performance without competence. In A. Wray (Ed.), *The transition to language*, (pp. 113–137). Oxford, UK: Oxford University Press.

Zuidema, W. (2003), Optimal communication in a noisy and heterogeneous environment. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, & J. Ziegler (Eds.), *Advances in Artificial Life – Proceedings of the 7th European Conference on Artificial Life (ECAL)*, Lecture Notes in Artificial Intelligence (Vol. 2801, pp. 553–563). Berlin: Springer Verlag.

Zuidema, W. (2005). *The major transitions in the evolution of language*. Unpublished doctoral dissertation, Theoretical and Applied Linguistics, University of Edinburgh, UK.

Zuidema, W., & de Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics 37*, 125–144.

Zuidema, W. & Westermann, G. (2003). Evolution of an optimal lexicon under constraints from embodiment. *Artificial Life*, *9*(4), 387–402.

About the Authors

Willem Zuidema received his PhD in linguistics from the University of Edinburgh (2005). Since 2004 he has worked as a postdoctoral researcher at the Institute for Logic, Language and Computation at the University of Amsterdam, primarily on models of language learning and evolution. From 2007 until 2009 he also worked at the department of Behavioural Biology, Leiden University. He is a member of the Cognitive Science Center Amsterdam and teaches in the MSc program *Brain & Cognitive Science*. In 2007 he received a 3-year NWO-Veni fellowship for the project *Discovering Grammar* to study the cognitive mechanisms underlying sequence learning in humans and other species.

Arie Verhagen obtained his PhD in 1986 at the Free University in Amsterdam, where he also became assistant professor. In 1991 he was appointed as associate professor of text linguistics at Utrecht University, and in 1998 as chair of Dutch Linguistics at Leiden University. He has been editor of the Dutch linguistics journal *Nederlandse taalkunde*, and editor-in-chief of *Cognitive Linguistics* (1996–2004). His research focuses on relationships between language use and language structure, in a cultural evolutionary approach to meaning and grammar (broadly conceived). His most recent book is *Constructions of Intersubjectivity. Discourse, Syntax, and Cognition* (OUP, 2005). *Address*: Leiden University Centre for Linguistics, Leiden University, The Netherlands.
E-mail: arie@arieverhagen.nl.