Chapter 8

The Role of Sequential Learning in Language Evolution: Computational and Experimental Studies

Morten H. Christiansen, Rick A.C. Dale,
Michelle R. Ellefson and Christopher M. Conway

Introduction

After having been plagued for centuries by unfounded speculations, the study of language evolution is now emerging as an area of legitimate scientific inquiry. Early conjectures about the origin and evolution of language suffered from a severe lack of empirical evidence to help rein in proposed theories. This lead to outlandish claims such as the idea that Chinese was the original ur-language of humankind, surviving the biblical flood because of Noah and his family (Webb, 1669, cited in Aitchison, 1998). Or, the suggestion that humans have learned how to sing and speak from the birds in the same way as they would have learned how to weave from spiders (Burnett, 1773, cited in Aitchison, 1998). Given this state of the art, it was perhaps not surprising that the influential Société Linguistique de Paris in 1866 imposed a ban on papers discussing issues related to language origin and evolution, and effectively excluded such theorizing from the scientific discourse.

It took more than a century before this hiatus was overcome. Fueled by theoretical constraints derived from recent advances in the brain and cognitive sciences, the last decade of the twentieth century saw a resurgence of scientific interest in the origin and evolution of language. What has now become clear is that the study of language evolution must *necessarily* be an interdisciplinary endeavor. Only by amassing evidence from many different disciplines can theorizing about the evolution of language be sufficiently constrained to remove it from the realm of pure speculation and allow it to become an area of legitimate scientific inquiry. Nonetheless, direct experimentation is needed in order to go beyond existing data. As the current volume is a testament to, computational modeling has become the paradigm of choice for such experimentation. Computational models provide an important tool to investigate how various types of hypothesized constraints may affect the evolution of language. One of the advantages of this approach is that specific constraints and/or interactions between constraints can be studied under controlled circumstances.

In this chapter, we point to *artificial language learning* (ALL) as an additional, complementary paradigm for exploring and testing hypotheses about language evolution. ALL involves training human subjects on artificial languages with particular structural constraints, and then testing their knowledge of the language. Because ALL permits researchers to investigate the language learning abilities of infants and children in a highly controlled environment, the paradigm is becoming increasingly popular as a method for studying language acquisition (for a review, see Gomez & Gerken, 2000). We suggest that ALL can similarly be applied to the investigation of issues pertaining to the origin and evolution of language in much the same way as computational modeling is currently being used.

In the remainder of this chapter, we show how a combination of computational modeling and ALL can be used to elicit evidence relevant for the explanation of language evolution. First, we outline our theoretical perspective on language evolution, suggesting that the evolution of language is more appropriately viewed as the selection of linguistic structures rather than the adaptation of biological structure. Specifically, we argue that limitations on sequential learning have played a crucial role in shaping the evolution of linguistic structure. In support for this perspective we report on convergent evidence from aphasia studies, human and ape ALL experiments, non-human primate sequential learning studies, and computational modeling. We then present two case studies involving our own computational modeling and ALL research. The results demonstrate how constraints on basic word order and complex question formation can be seen to derive from underlying cognitive limitations on sequential learning. Finally, we discuss the current limitations and future challenges for our approach.

Language as an Organism

Languages exist only because humans can learn, produce, and process them. Without humans there would be no language (in the narrow sense of *human* language). It therefore makes sense to construe languages as organisms that have had to adapt themselves through natural selection to fit a particular ecological niche: the human brain (Christiansen, 1994; Christiansen & Chater, in preparation). In order for languages to "survive", they must adapt to the properties of the human learning and processing mechanisms. This is not to say that having a language does not confer selective advantage onto humans. It seems clear that humans with superior language abilities are likely to have a selective advantage over other humans (and other organisms) with lesser communicative powers. This is an uncontroversial point, forming the basic premise of many of the adaptationist theories of language evolution. However, what is often not appreciated is that the selection forces working on language to fit humans are significantly stronger than the selection pressure on humans to be able to use language. In the case of the former, a language can *only* survive if it is learnable and processable by humans. On the other hand, adaptation towards language use is merely *one out of many* selective pressures working on humans (such as, for example, being able to avoid predators and find food). Whereas humans can survive without language, the opposite is not the case. Thus, language is more likely to have adapted itself to its human hosts than the other way round. Languages that are hard for humans to learn simply die out, or more likely, do not come into existence at all.

The biological perspective on language as an adaptive system has a prominent historical pedigree. Indeed, nineteenth-century linguistics was dominated by an organistic view of language (for a review, see e.g., McMahon, 1994). For example, Franz Bopp, one of the founders of comparative linguistics, regarded language as an organism that could be dissected and classified (Davies, 1987). More generally, languages were viewed as having life cycles that included birth, progressive growth, procreation, and eventually decay and death. However, the notion of evolution underlying this organistic view of language was largely pre-Darwinian. This is perhaps reflected most clearly in the writings of another influential linguist, August Schleicher. Although he explicitly emphasized the relationship between linguistics and Darwinian theory (Schleicher, 1863; cited in Percival, 1987), Darwin's principles of mutation, variation, and natural selection did not enter into the theorizing about language evolution (Nerlich, 1989). Instead, the evolution of language was seen in pre-Darwinian terms as the progressive growth toward attainment of perfection, followed by decay.

More recently the biological view of language evolution was resurrected by Stevick (1963) within a modern Darwinian framework, later followed by Nerlich (1989). Christiansen (1994) proposed to view language as a kind of beneficial parasite — a nonobligate symbiant — that confers some selective advantage onto its human hosts without whom it cannot survive. Building on this work, Deacon (1997) further developed this metaphor by construing language as a virus. The asymmetry in the relationship between language and its human host is underscored by the fact that the rate of linguistic change is far greater than the rate of biological change. Whereas it takes about 10,000 years for a language to change into a completely different "species" of language (e.g., from protolanguage to present day language, Kiparsky, 1976), it took our remote ancestors approximately 100-200,000 years to evolve from the archaic form of Homo sapiens into the anatomically modern form, Homo sapiens sapiens (see, e.g., Corballis, 1992). Consequently, it seems more plausible that the languages of the world have been closely tailored through linguistic adaptation to fit human learning, rather than the other way around. The fact that children are so successful at language learning is therefore best explained as a product of natural selection of linguistic structures, and not as the adaptation of biological structures, such as an innately specified linguistic endowment in the form of. universal grammar (UG)[1].

---

[1] Many functional and cognitive linguists also suggest that the putative innate UG constraints arise from general cognitive constraints (e.g., Givón, 1998; Hawkins, 1994;

Cognitive constraints on language evolution and acquisition

From this perspective, it is clear that there exist innate constraints guiding language learning. Indeed, a recent population dynamics model by Nowak, Komarova, and Niyogi (2001) provides a mathematical setting for exploring language acquisition under constraints (such as UG), and evolutionary competition among them. This mathematical model is based on what the authors call a "coherence threshold". In order for a population to communicate successfully, all its members must acquire the same language. The coherence threshold is a property that UG or other potential constraints must meet for them to induce "coherent grammatical communication" in the linguistic community. When the authors used this mathematical framework to compare competing systems of constraints (different UGs), they found that complexity confers a fitness advantage upon them[2]. This is offered as an explanation for the emergence of complex, rule-based languages. Although UG is the purported object of study in Nowak et al., there is little to preclude extending these findings to our own perspective. The innate constraints need not be language-specific in nature for the model's assumptions to be satisfied. The important question is therefore not about the existence of innate constraints on language—we take this to be given—but rather what the nature is of such constraints.

Given our perspective on language evolution, we suggest that many of these innate constraints derive from limitations on sequential learning. By "sequential learning" we here focus on the learning of hierarchically organized structure from temporally-ordered input, in which combinations of primitive elements can themselves become primitives for further higher-level combinations. For example, consider the case of following a recipe involving mixing separately one set of ingredients in one bowl and other ingredients in another bowl before mixing the contents of the two bowls together (possibly with additional ingredients). The preparation of certain plant foods by mountain gorillas (*Gorilla g. beringei*) in Rwanda, Zaire and Uganda provides another example of complex sequential learning (Byrne & Russon, 1998). Because their favorite foods are protected by physical defenses such as spines or stings, the gorillas learn hierarchical manual sequences with repeated subpatterns in order to collect the plant material and make it edible. Although sequential learning appears to be ubiquitous across animal species (e.g., Reber, 1993), humans may be the only species with complex

---

Langacker, 1987). Our approach distinguishes itself from these linguistic perspectives in that it emphasizes the role of sequential learning in the explanation of linguistic constraints. Another difference is our general emphasis on the acquisition of language, rather than the processing of language (cf. Hawkins, 1994).

[2] Nowak et al. (2001) also noted that when they varied the number of sentences available to the learners, they found that intermediate values maximized fitness. They claim this provides an explanation for the critical language acquisition period. Though the model is touted as an evolutionary framework for illuminating a supposedly biological property of our species (UG), this explanation for the critical period relies on an unbiological basis. Hypotheses of critical periods involve maturational issues of the learning mechanism, not the number of sentences offered by the environment.

sequential learning abilities flexible enough to accommodate a communication system containing several layers of temporal hierarchical structure (at the level of phonology, morphology and syntax). Next we present converging evidence from studies of aphasia, ALL, studies of non-human primates, and computational modeling—all of which points to the importance of sequential learning in the evolution language.

Language and Sequential Learning

Several lines of evidence currently support the importance of sequential learning in language evolution. This evidence spans a number of different research areas, ranging from sequential learning abilities of aphasic patients to computational modeling of language evolution. When these sources are considered within the framework argued for here, they converge in support of a strong association between sequential learning and language evolution, acquisition, and processing.

Evidence from aphasia studies

The first line of evidence comes from the study of aphasia. If language and sequential learning are subserved by the same underlying mechanisms, as we have suggested here, then one would expect that breakdown of language in certain types of aphasia to be associated with impaired sequential learning and processing. A large number of Broca's aphasics suffer from agrammatism. Their speech lacks the hierarchical organization we associate with syntactic structure, and instead appears to be a collection of single words or simple word combinations. Importantly, Grossman (1980) found that Broca's aphasics, besides agrammatism, also had an additional deficit in sequentially reconstructing hierarchical tree structure models from memory. He took this as suggesting that Broca's area subserves not only syntactic speech production, but also functions as a locus for supramodal processing of hierarchically structured behavior. Another study has suggested a similar association between language and sequential processing. Kimura (1988) reported that sign aphasics often also suffer from apraxia; that is, they have additional problems with the production of novel sequential hand and arm movements not specific to sign language.

More recently, Christiansen, Kelly, Shillcock, and Greenfield (in preparation) provided a more direct test of the suggested link between breakdown of language and breakdown of sequential learning. They conducted an ALL study using agrammatic patients and normal controls matched for age, socio-economic status, and spatial reasoning abilities. Artificial language learning experiments typically involve training and testing subjects on strings generated from a small grammar. The vocabulary of these grammars can consist of letters, nonsense words, or non-linguistic symbols (e.g., shapes). Because of the underlying sequential structure of the stimuli, the experiments can serve as a window onto the relationship between the learning and processing of linguistic and sequential structure. The subjects in the Christiansen *et. al.* study were trained on an artificial language using a match-mismatch pairing task in which they had to decide whether two consecutively presented symbol strings were the same or different. After training, subjects were presented with novel strings, half of which were derived from the grammar and half not. Subjects were told that the training strings were generated by a complex set of rules, and asked to classify the new strings according to whether they followed these rules or not. The results showed that although both groups did very well on the pairing task, the normal controls were significantly better at classifying the new test strings in comparison with the agrammatic aphasics. Indeed, the aphasic patients were no better than chance at classifying the test items. Thus, the study indicates that agrammatic aphasic patients have problems with sequential learning in addition to their more obvious language deficits. Together, these experimentally observed sequential learning and processing deficits associated with agrammatic aphasia point to a close connection between the learning and processing of language and complex sequential structure.

Evidence from artificial language learning experiments

Our approach hypothesizes that many of the cognitive constraints that have shaped the evolution of language are still at play in our current cognitive and linguistic abilities. If this hypothesis is correct, then it should be possible to uncover the source of some of the universal linguistic constraints in human performance on sequential learning tasks. We therefore review a series of ALL studies with normal populations as a second line of evidence for the close relationship between language and sequential learning.

The acquisition and processing of language appears to be facilitated by the presence of multiple sources of probabilistic information in the input (e.g., concord morphology and prosodic information; see contributions in Morgan & Demuth, 1996). Morgan, Meyer, and Newport (1987) demonstrated that ALL is also facilitated by the existence of multiple information sources. They exposed adults to artificial languages with or without additional cue information, such as prosodic or morphological marking of phrases. Subjects provided with the additional cue information acquired more of the linguistic structure of the artificial language. More recently, Saffran (2001) studied the learning of an artificial language with or without the kind of predictive constraints found in natural language (e.g., the presence of the determiner, *the*, is a very strong predictor of an upcoming noun). She found that both adults and children acquired more of the underlying structure of the language when it incorporated the "natural" predictive constraints. Saffran (2000) has also demonstrated that the same predictive constraint is at play when subjects are exposed to an artificial language consisting of non-linguistic sounds (e.g., drum rolls, etc.), providing further support for the non-linguistic nature of the underlying constraints. In unison with our perspective, the authors of these ALL studies suggest that human languages might contain certain sequential patterns, not because of linguistic constraints, but rather because of the general learning constraints of the human brain.

The ALL studies with normal and aphasic populations together point to a strong association between language and the learning and processing of sequential structure. The close connection in terms of underlying brain mechanisms is further underscored by recent neuroimaging studies of ALL. Steinhauer, Friederici, and Pfeifer (2001) had subjects play a kind of board game in which two players were required to communicate via an artificial language. After substantial training, event-related potential (ERP) brainwave patterns were then recorded as the subjects were tested on grammatical and ungrammatical sentences from the language. The results showed the same frontal negativity pattern (P600) for syntactic violations in the artificial language as has been found for similar violations in natural language (e.g., Osterhout & Holcomb, 1992). Another study by Patel, Gibson, Ratner, Besson, and Holcomb (1998) further corroborates this pattern of results but with non-linguistic sequential stimuli: musical sequences with target chords either within the key of a major musical phrase or out of key. When they directly compared the ERP patterns elicited for syntactic incongruities in language with the ERP patterns elicited for incongruent out-of-key target chords, they found that the two types of sequential incongruities resulted in the same, statistically indistinguishable P600 components. In a more recent study, Maess, Koelsch, Gunter, and Friederici (2001) used magnetoencephalography (MEG) to localize the neural substrates that may be involved in the processing of musical sequences. They found that Broca's area in the left hemisphere (and the corresponding frontal area in the right hemisphere) produced significant activation when subjects listened to musical sequences that included an off-key chord. The ALL studies reviewed here converge on the suggestion that the same underlying brain mechanisms are used for the learning and processing of both linguistic and non-linguistic sequential structure, and that similar constraints are imposed on both language and sequential learning.

Evidence from non-human primate studies

The perspective on language evolution presented here suggests that language to a large extent "piggy-backed" on pre-existing sequential learning and processing mechanisms, and that limitations on these mechanisms in turn gave rise to many of the linguistic constraints observed across the languages of the world. If this evolutionary scenario is on the right track, one would expect to see some evidence of complex sequential learning in our closest primate relatives—and this is exactly what is suggested by the third line of evidence that we survey here.

A review of recent studies investigating sequential learning in non-human primates (Conway & Christiansen, 2001) indicates that there is considerable overlap between the sequential learning abilities of humans and non-human primates. For instance, macaque monkeys (*Macaca mulatta* and *Macaca fascicularis*) not only are competent list-learners (Swartz, Chen, & Terrace, 2000) but they appear to encode and represent sequential items by learning each item's ordinal position (Orlov, Yakovlev, Hochstein, & Zohary, 2000) rather than by a simple association mechanism. In addition, cotton-top tamarins (*Saguinus oedipus*) are able to successfully segment artificial words from an auditory speech stream by relying on statistical information in a manner similar to human infants (Hauser, Newport, & Aslin, 2001; Saffran, Aslin, & Newport, 1996). Finally, as mentioned earlier, a group of African mountain gorillas apparently observationally learn sequences of complex and hierarchically organized manual actions to bypass the natural defenses of edible plants (Byrne & Russon, 1998). However, despite these impressive sequential learning abilities, non-human primates also display certain limitations in comparison to humans. In some tasks, non-humans need considerably longer training in order to adequately learn sequential information (cf., Lock & Colombo, 1996). More importantly, non-human subjects often display sequential learning and behavior that is less complex and less developed compared to human children and adults (e.g., Oshiba, 1997), especially with regards to the learning of hierarchical structure (e.g., Johnson-Pynn, Fragaszy, Hirsh, Brakke, & Greenfield, 1999; Spinozzi & Langer, 1999). We suggest that such limitations may help explain why non-human primates have not developed complex, human-like language.

The limitations of the non-human primates on sequential learning and processing are also likely to play a role in the explanation of the limited success of the numerous ape language learning experiments. Indeed, we see these experiments as complex versions of the ALL tasks used with humans[3]. Much like some human ALL experiments, the non-human primates must learn to associate arbitrary visual symbols (lexigrams), manual signs, or spoken words with objects, actions, and events. Some of these studies have shown that apes can acquire complex artificial languages with years of extensive training. Although some of the "stars" of these experiments—such as the female gorilla Koko (Patterson, 1978) and the male bonobo Kanzi (Savage-Rumbaugh, Shanker, & Taylor, 1998)—have demonstrated remarkable abilities for learning the artificial language they have been exposed to, they nevertheless also seem to experience problems with complex sequential structures. Non-human primates, in particular the apes, possess sequential learning abilities of a reasonable complexity and appear to be able to utilize these abilities in complex ALL tasks. Yet the language abilities of these apes remain relatively limited compared to those of young children. On our account, the better sequential learning and processing abilities observed in humans are likely to be the product of evolutionary changes occurring after the branching point between early hominids and the ancestors of extant apes. These evolutionary improvements in sequential learning have then subsequently provided the basis for the evolution of language.

Evidence from computational modeling

An important question for all evolutionary accounts of language pertains to the feasibility of the underlying assumptions. For example, our approach emphasizes

[3] Early ape language experiments attempted to teach non-human primates actual human language (e.g., Kellogg & Kellogg, 1933). The animals were spoken to and treated in a manner similar to human infants and young children. However, this approach was subsequently abandoned because of lack of success and replaced by the artificial language methodology used today.

the role of linguistic adaptation over biological adaptation in the evolution of language. As we have mentioned earlier, computational modeling provides a very fruitful means with which to test the assumptions of a given approach. As a final line of evidence in support of our perspective on language evolution we therefore review some recent modeling efforts that demonstrate its computational feasibility[4].

Several recent computational modeling studies have shown how the adaptation of linguistic structure can result in the emergence of complex languages with features very similar to what is observed in natural languages. Batali's (1998) "negotiation" model explored the appearance of systematic communication in a social group of agents in the form of simple recurrent networks (SRN; Elman, 1990). An SRN is essentially a standard feed-forward neural network equipped with an extra layer of so-called context units. At a particular time step *t*, an input pattern is propagated through the hidden unit layer to the output layer. At the next time step, *t+1*, the activation of the hidden unit layer at time *t* is copied back to the context layer and paired with the current input. This means that the current state of the hidden units can influence the processing of subsequent inputs, providing a limited ability to deal with sequentially presented input incorporating hierarchical structure. Although these network agents were not initially equipped with a system of communication, the generated sequences gradually exhibited systematicity. Batali also demonstrated that this communication system enabled the agents to convey novel meanings. Importantly, there was no "biological" adaptation (e.g., selection of better learners); instead, the communication system emerged from linguistic adaptation driven by the social interaction of agents. Kirby offered a similar account for the evolution of typological universals (Kirby, 1998), and systematic communication in agents without prior grammatical encoding (Kirby, 2000; 2001). Using abstract rule-based descriptions of individual language fragments, Kirby demonstrated that fairly complex properties of language could arise under an adaptive interpretation of linguistic selection.

Livingstone (2000) and Livingstone and Fyfe (1999) used a similar technique to show that linguistic diversity can arise from an imperfect cultural transmission of language among a spatially organized group of communicating agents. In their simulations, neural network agents, able only to communicate with others in close proximity, exhibited a dialect continuum: intelligibility was high in clusters of agents, but diminished significantly as the distance between two agents increased. In a similar simulation without such spatial distribution (where any agent is equally probable to communicate with all others), diversity rapidly converged onto a global language. This work demonstrates how linguistic diversity may arise through linguistic adaptation across a spatially distributed population of agents, perhaps giving rise to different languages over time. Some of these emergent languages are likely to be more easily accommodated by sequential learning and processing mechanisms than other languages. This sequential learnability difference is, *ceteris*

[4] To keep our discussion brief, we focus on the computational modeling of linguistic adaptation, side-stepping the issue of the origin of language. For simulations relevant to this perspective, see e.g., Arbib (this volume) and Parisi and Cangelosi (this volume).

paribus[5], likely to result in different frequency distributions across languages. Simulations by Van Everbroek (1999) substantiate this hypothesis. He used a variation of the SRN to investigate how sequential learning and processing limitations might be related to the distribution of the world's language types. He constructed example sentences from 42 artificial languages, varying in three dimensions: word order (e.g., subject-verb-object), nominal marking (accusative vs. ergative), and verbal marking. The networks easily processed language types with medium to high frequency, while low frequency language types resulted in poor performance. These simulations support a connection between the distribution of language types and constraints on sequential learning and processing, suggesting that frequent language types are those that have successfully adapted to these learning and processing limitations.

The computational modeling results lend support to the suggestion that the evolution of language may have been shaped by linguistic adaptation to pre-existing constraints on sequential learning and processing. When these results are viewed together with the evidence showing a breakdown of sequential learning in agrammatic aphasia, the ALL demonstrations of linguistic constraints as reflections of sequential learning limitations with similar neural substrates, and the existence of relatively complex sequential learning abilities in apes, they all appear to converge on the language evolution account we have put forward here. Next, we present two case studies that provide further evidence for the idea that constraints on sequential learning may underlie many universal linguistic constraints.

Explaining Basic Word Order Constraints

Across the languages of the world there are certain universal constraints on the way in which languages are structured and used. These so-called *linguistic universals* help explain why the known human languages only take up a small fraction of the vast space defined by the logically possible linguistic subpatterns. From the viewpoint of the UG approach to language, the universal constraints on the acquisition and processing of language are essentially arbitrary (e.g., Pinker & Bloom, 1990). That is, given the Chomskyan perspective on language, these constraints appear arbitrary because it is possible to imagine a multitude of alternative, and equally adaptive, constraints on linguistic form. For instance, Piattelli-Palmarini (1989) contends that there are no (linguistic) reasons not to form yes-no questions by reversing the word order of a sentence instead of the normal inversion of subject and auxiliary. On our account, however, these universal constraints are in most cases *not* arbitrary. Rather, they are determined predominately by the properties of the human learning and processing mechanisms that underlie our language capacity. This can explain why we do not reverse the

[5] Of course, other factors are likely to play a role in whether or not a given language may be learnable. For example, the presence of concord morphology may help overcome some sequential learning difficulties as demonstrated by an ALL experiment by Morgan et al. (1987). Nonetheless, sequential learning difficulties are hypothesized to be strong predictors of frequency in the absence of such ameliorating factors.

word order to form yes-no questions; it would put too heavy a load on memory to store a whole sentence in order to be able to reverse it.

Head-order consistency

There is a statistical tendency across human languages to conform to a form in which the head of a phrase consistently is placed in the same position—either first or last—with respect to the remaining clause material. English is considered to be a head-first language, meaning that the head is most frequently placed first in a phrase, as when the verb is placed before the object noun phrase (NP) in a transitive verb phrase (VP) such as *"eat curry"*. In contrast, speakers of Hindi would say the equivalent of *"curry eat"*, because Hindi is a head-last language. Likewise, head-first languages tend to have *pre*positions before the NP in prepositional phrases (PP) (such as *"with a fork"*), whereas head-last languages tend to have *post*positions following the NP in PPs (such *as "a fork with"*). Within the Chomskyan approach to language (e.g., Chomsky, 1986) such head direction consistency has been explained in terms of an innate module known as X-bar theory which specifies constraints on the phrase structure of languages. It has further been suggested that this module emerged as a product of natural selection (Pinker, 1994). As such, it comes as part of the UG with which every child is supposedly born. All that remains for a child to "learn" about this aspect of her native language is the direction (i.e., head-first or head-last) of the so-called head-parameter.

The evolutionary perspective that we have proposed above suggests an alternative explanation in which head-order consistency is a by-product of non-linguistic constraints on the learning of hierarchically organized temporal sequences. In particular, if recursively consistent combinations of grammatical regularities, such as those found in head-first and head-last languages, are easier to learn (and process) than recursively inconsistent combinations, then it seems plausible that recursively inconsistent languages would simply "die out" (or not come into existence), whereas the recursively consistent languages should proliferate. As a consequence languages incorporating a high degree of recursive inconsistency should be far less frequent among the languages of the world than their more consistent counterparts. In other words, languages may need to have a certain recursive consistency across their different grammatical regularities in order for the former to be learnable by learning devices with adapted sensitivity to sequential information. Languages that do not have this kind of consistency in their grammatical structure may not be learnable, and they will, furthermore, be difficult to process (cf. Hawkins, 1994).

From this perspective, Christiansen and Devlin (1997) provided an analysis of the interactions in a recursive rule set with consistent and inconsistent ordering of the heads[6]. A recursive rule set is a pair of rules for which the expansion of one rule involves the second rule, and vice versa; e.g.,

---

[6] The fact that we use rules and (later) syntactic trees to characterize the language to be acquired should not be taken as suggesting that we believe that the end-product of the

$$A \rightarrow a\ (B)$$
$$B \rightarrow b\ A$$

This analysis showed that head-order inconsistency in a recursive rule set, such as,

$$A \rightarrow a\ (B)$$
$$B \rightarrow A\ b$$

creates center-embedded constructions, whereas a consistent ordering of heads creates right-branching constructions for head-first orderings and left-branching constructions for head-last orderings. Center-embeddings are difficult to process because constituents cannot be completed immediately, forcing the language processor to keep lexical material in memory until it can be discharged. For the same reason, center-embedded structures are likely to be difficult to learn because of the distance between the material relevant for the discovery and/or reinforcement of a particular grammatical regularity. This means that recursively inconsistent rule sets are likely to be harder to learn than recursively consistent rule sets.

To explore the notion of recursive inconsistency further, Christiansen and Devlin created the grammar skeleton shown in Table 8.1. The curly brackets around the constituents on the right-hand side of rules 1-5 indicate that the order of these constituents can be either as is (i.e., head-first) or the reverse (i.e., head-last). From this grammar skeleton, it is therefore possible to produce (2^5=) 32 different grammars with varying degrees of head-order consistency. There are two possibilities for recursive inconsistency: a) the PP recursive rules set (rules 1 and 2), and b) the PossP (possessive phrase) recursive rule set (rules 4 and 5). Since a PP can occur inside both NPs and VPs, an inconsistency within this rule set was predicted to impair learning more than an inconsistency violation within the PossP recursive rule set. Grammars that involved inconsistent PP recursive rule sets were therefore assigned an inconsistency penalty of 2 and grammars with inconsistent PossP recursive rule sets a penalty of 1. The top panel of Figure 8.1 shows the predicted learning difficulty of each grammar, ranging between 0 to 3.

Table 8.1 The grammar skeleton used by Christiansen and Devlin (1997). Curly brackets indicate that the ordering of the constituents can be either as is (i.e., head-first) or in reverse (i.e., head-last), whereas parentheses indicate optional constituents.

S	→	NP VP	
NP	→	{N (PP)}	(1)
PP	→	{adp NP}	(2)
VP	→	{V (NP) (PP)}	(3)
NP	→	{N (PossP)}	(4)
PossP	→	{Poss NP}	(5)

acquisition process is a set of rules. We merely use rules and syntactic trees as convenient descriptive devices, approximating the particular grammatical regularities that we are considering.

Connectionist simulations

In order to test the hypothesis that non-linguistic constraints on sequential learning restrict the set of languages that are easily learnable, Christiansen and Devlin conducted a series of connectionist simulations in which SRNs were trained on sentences generated from each of the 32 grammars. The networks were trained to predict the next lexical category in a sentence, using sentences generated by the 32 grammars. Each unit in the input/output layers corresponded to one of seven lexical categories or an end of sentence marker: singular/plural noun (N), singular/plural verb (V), singular/plural possessive genitive affix (Poss), and adposition (adp). Although these input/output representations abstract away from many of the complexities facing language learners, they suffice to capture the fundamental aspects of grammar learning important to our hypothesis. Network performance was measured in terms of the networks' ability to predict the probability distribution of possible next items given prior sentential context. The bottom panel of Figure 8.1 shows SRN performance, averaged over 25 networks, for each of the 32 different grammars. A comparison between the top and bottom panels in Figure 8.1 reveals that the grammars that were predicted to be harder to learn because of high recursive inconsistency are the ones that the SRNs showed decreased performance on. A regression analysis confirmed this observation, showing a strong correlation between the degree of head-order consistency of a given grammar and the degree to which the network had learned to master the grammatical regularities underlying that grammar: The higher the inconsistency, the more erroneous the final network performance was. The sequential biases of the networks made the corpora generated by consistent grammars considerably easier to acquire than the corpora generated from inconsistent grammars.

This is an important result because it is not obvious that the SRNs should be sensitive to inconsistencies at the structural level. The SRN did not have any built-in linguistic biases; rather, it was designed for the learning of complex sequential structure (e.g., Cleeremans, 1993). Moreover, recall that the networks only were presented with lexical categories one at a time, and that structural information about grammatical regularities had to be induced from the way the lexical categories combine in the input. No explicit structural information was provided, yet the networks were sensitive to the recursive inconsistencies. In this connection, it is worth noting that Christiansen and Chater (1999) have shown that increasing the size of the hidden/context layers (beyond a certain minimum) does not affect SRN performance on center-embedded constructions (i.e., structures which are recursively inconsistent structures). This suggests that Christiansen and Devlin's results may not be dependent on the specific size of the SRNs they used, nor is it likely to depend on the size of the training corpus.

Typological analyses by Christiansen and Devlin using the FANAL database (Dryer, 1992) with typological information about some 625 languages further corroborated our account. Languages that incorporated fragments that the networks found hard to learn tended to be less frequent among the languages of the world compared to languages the networks learned more easily. This suggests that constraints on basic word order may derive from non-linguistic constraints on the learning and processing of complex sequential structure, perhaps obviating the

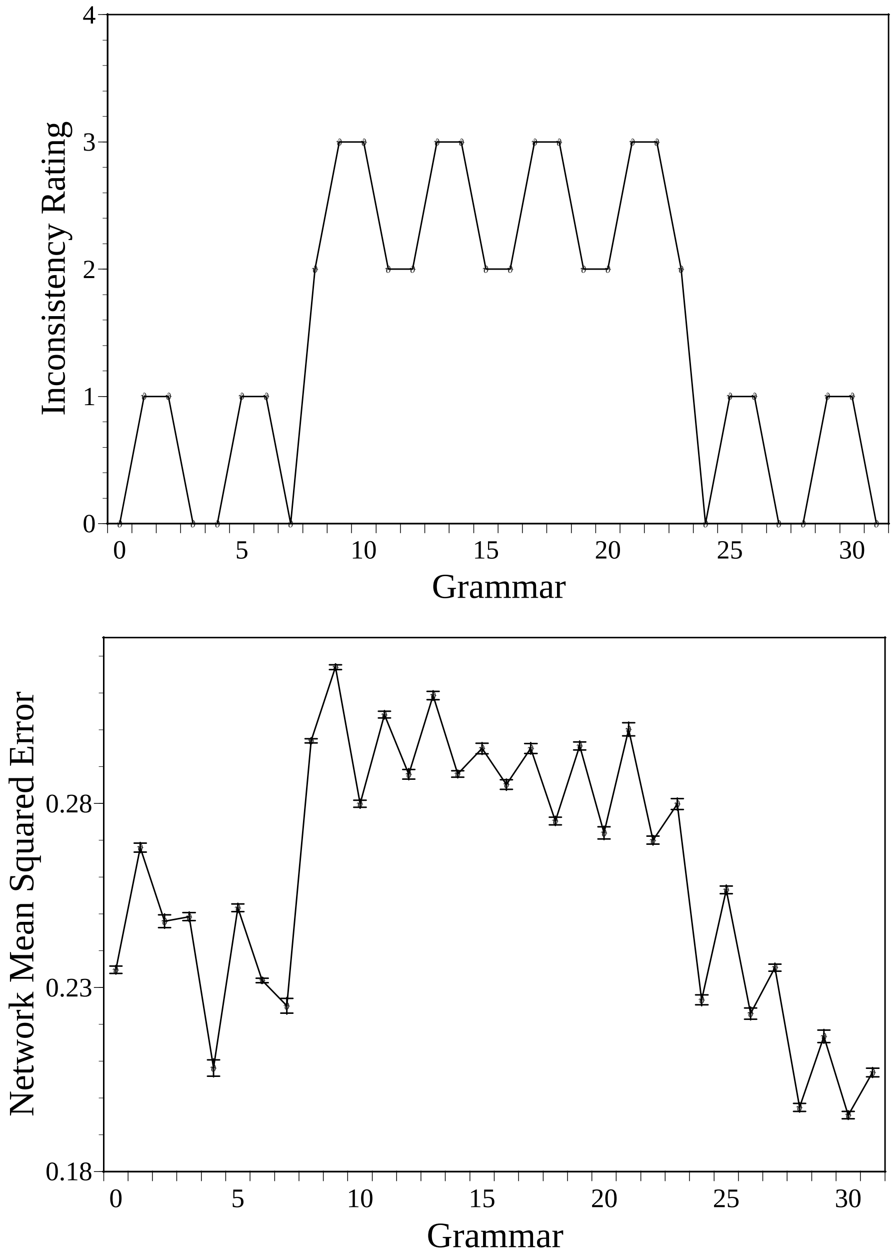

Figure 8.1 The predicted learning difficulty for the 32 grammars from Christiansen and Devlin (1997) (top panel) shown with the difficulty that the network experienced with each grammar (bottom panel). Error bars indicate standard error of the mean.

need for an innate X-bar module to explain such word order universals. Grammatical constructions incorporating a high degree of head-order inconsistency may be too hard to learn and would therefore tend to disappear.

Artificial language learning experiment

The final set of evidence supporting our explanation of basic word order universals comes from a recent ALL study by Christiansen (2000). In one experiment, Christiansen took two of the grammars that Christiansen and Devlin had used for their network simulation—a consistent and an inconsistent grammar (see Table 8.2)—and trained subjects on sentences (represented as consonant strings) derived from the two grammars. Training and test materials were controlled for length and differences in the distribution of bigram and trigram fragments. In the training phase of the experiment, subjects read and reproduced consonant strings on a computer. After training, subjects were informed that the strings were generated by a complex set of rules, and that they would be presented with additional strings; some of which were generated by the same rule set (i.e., grammatical), and some which were not (i.e., ungrammatical). They were then asked to decide which of the new strings were generated by the same rule set as before, and which were not. The results showed that the subjects trained on strings from the consistent grammar were significantly better at distinguishing grammatical from ungrammatical items than the subjects trained on the inconsistent grammar.

Together, Christiansen's ALL experiment and the three sets of evidence from Devlin and Christiansen converge in support of our claim that basic word order universals (head-ordering) can be explained in terms of non-linguistic constraints on sequential learning and processing. This research thus suggests that universal word order correlations may emerge from non-linguistic constraints on learning, rather than being a product of innate linguistic knowledge. In the next section, we show how constraints on complex question formation may be explained in a similar manner.

Table 8.2 The two grammars used for stimuli generation in Christiansen (2000). The vocabulary is: {X, Z, Q, V, S, M}

Consistent Grammar			Inconsistent Grammar		
S	→	NP VP	S	→	NP VP
NP	→	(PP) N	NP	→	(PP) N
PP	→	NP post	PP	→	pre NP
VP	→	(PP) (NP) V	VP	→	(PP) (NP) V
NP	→	(PossP) N	NP	→	(PossP) N
PossP	→	NP Poss	PossP	→	Poss NP

Subjacency through Linguistic Adaptation

According to Pinker and Bloom (1990), subjacency is one of the classic examples of an arbitrary linguistic universal that makes sense only from a linguistic perspective. Subjacency provides constraints on complex question formation. Informally, "Subjacency, in effect, keeps rules from relating elements that are 'too far apart from each other', where the distance apart is defined in term of the number of designated nodes that there are between them" (Newmeyer, 1991, p. 12). Consider the following sentences:

1. Sara heard (the) news that everybody likes cats.
   N V N comp N V N
2. What (did) Sara hear that everybody likes?
   Wh N V comp N V
3. *What (did) Sara hear (the) news that everybody likes?
   Wh N V N comp N V

According to the subjacency principle, sentence 3 is ungrammatical because too many boundary nodes are placed between the noun phrase complement (NP-Comp) and its respective "gaps".

The subjacency principle, in effect, places certain restrictions on the ordering of words in complex questions. The movement of Wh-items (*what* in Figure 8.2) is limited with respect to the number of so-called bounding nodes that it may cross during its upward movement. In English, the bounding nodes are S and NP (circled in Figure 8.2). Put informally, as a Wh-item moves up the tree it can use comps as temporary "landing sites" from which to launch the next move. The subjacency principle states that during any move only a single bounding node may be crossed. Sentence 2 is therefore grammatical because only one bounding node is crossed for each of the two moves to the top comp node (Figure 8.2, top panel). Sentence 3 is ungrammatical, however, because the Wh-item has to cross two bounding nodes—NP and S—between the temporary comp landing site and the topmost comp, as illustrated in bottom panel of Figure 8.2.

Not only do subjacency violations occur in NP-complements, but they can also occur in Wh-phrase complements (Wh-Comp). Consider the following examples:

4. Sara asked why everyone likes cats.
   N V Wh N V N
5. Who (did) Sara ask why everyone likes cats?
   Wh N V Wh N V N
6. *What (did) Sara ask why everyone likes?
   Wh N V Wh N V

According to the subjacency principle, sentence 6 is ungrammatical because the interrogative pronoun has moved across too many bounding nodes (as was the case in 3).

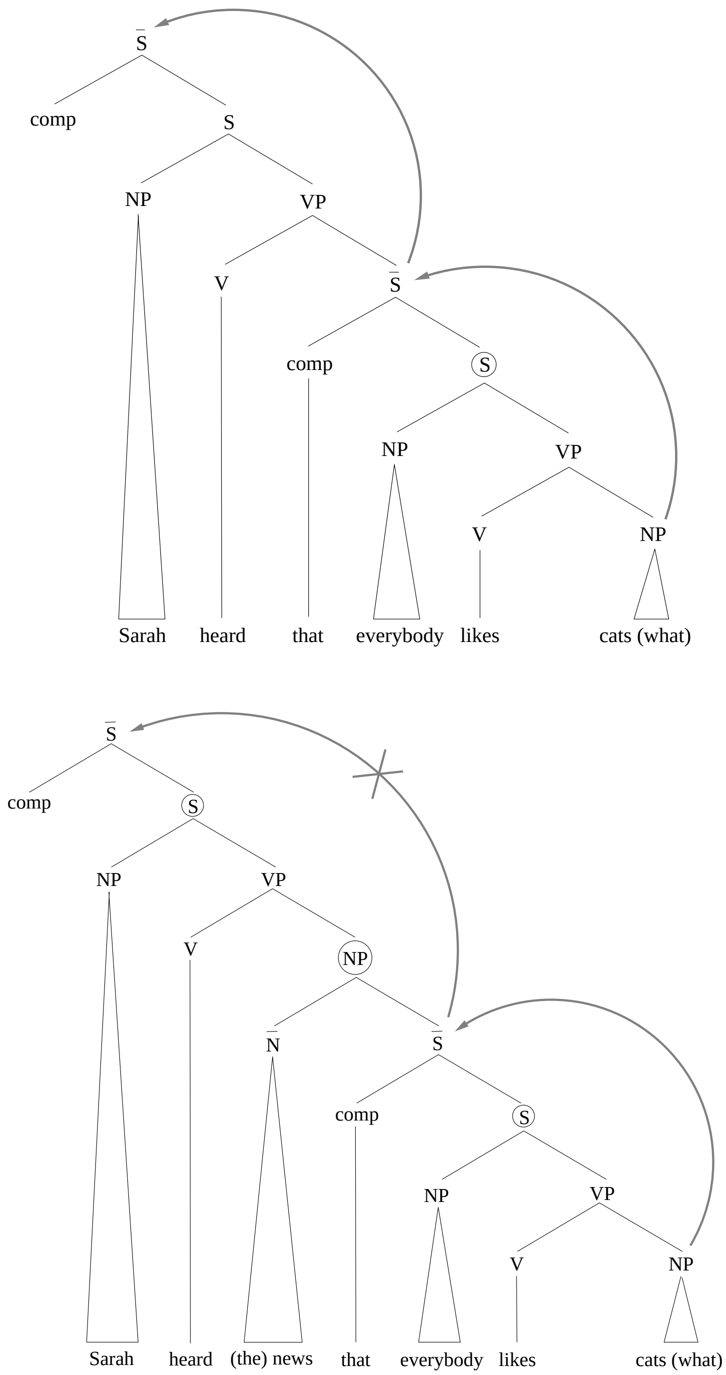

Figure 8.2 Syntactic trees showing grammatical (top panel) and ungrammatical (bottom panel) movement.

Table 8.3 The structure of the natural and unnatural languages in Ellefson and Christiansen (2000). The vocabulary is: {X, Z, Q, V, S, M}

Natural	Unnatural
N V N	N V N
Wh N V	Wh N V
N V N comp N V N	N V N comp N V N
N V Wh N V N	N V Wh N V N
Wh N V comp N V	*Wh N V N comp N V
Wh N V Wh N V N	*Wh N V Wh N V

Artificial language learning experiment

Ellefson and Christiansen (2000) explored an alternative explanation, suggesting that subjacency violations are avoided, not because of a biological adaptation incorporating the subjacency principle, but because language itself has undergone adaptations to root out such violations in response to non-linguistic constraints on sequential learning. They created two artificial languages to test this idea. As shown in Table 8.3, both languages consisted of six sentence types of which four were identical across the two languages. The two remaining sentence types involved complex question formation. In the natural language the two complex questions were formed in accordance with subjacency, whereas the two complex questions in the unnatural language violated the subjacency constraints. All training and test items were controlled for length and fragment information. As in the previous ALL experiment, subjects were not told about the linguistic nature of the stimuli until they received the instructions for the test phase.

The results showed that the subjects trained on the natural language had learned the language significantly better than the subjects trained on the unnatural language. Subjects in the natural condition performed marginally better than the subjects in the unnatural condition at classifying strings related to the two complex questions. Interestingly, the natural group was significantly better at classifying the remaining four sentence types in comparison with the unnatural group—despite the fact that both groups were trained on exactly the same items and saw exactly the same test items. The presence of the two unnatural question formation sentence types affected the learning of the other four test items. In other words, the presence of the subjacency violations in two of the sentence types in the unnatural language affected the learning of the language as a whole, not just the two complex question items. From the viewpoint of language evolution, languages such as this unnatural language would lose out in competition with other languages such as the natural language because the latter is easier to learn.

Connectionist simulations

In principle, one could object that the reason why Ellefson and Christiansen found differences between the natural and the unnatural groups is because the former in some way was able to tap into an innately specified subjacency principle when learning the language. Another possible objection is that the natural language follows the general pattern of English whereas the unnatural language does not, and that our human results could potentially reflect an "English effect". To counter these possible objections, and to support the suggestion that the difference in learnability between the two languages is brought about by constraints arising from sequential learning, Ellefson and Christiansen conducted a set of connectionist simulations of the human data using SRNs—a sequential learning device that clearly does not have subjacency constraints built-in. They used one network for each subject, and found that the networks were significantly better at learning the natural language in comparison with the unnatural language. Thus, the simulation results closely mimicked the ALL results, corroborating the suggestion that constraints on the learning and processing of sequential structure may explain why subjacency violations tend to be avoided: These violations were weeded out because they made the sequential structure of language too difficult to learn. Thus, rather than having an innate UG principle to rule out subjacency violations, we suggest that they may have been eliminated altogether through linguistic adaptation.

General Discussion

In this chapter we have argued for a view of language evolution, acquisition, and processing that places these phenomena within the more general domain of sequential learning. We hypothesize that constraints on sequential learning help define a cognitive niche within which languages have had to adapt. A considerable amount of evidence that supports this view has been discussed. ALL studies of normal human subjects illuminate the importance of complexity and consistency in learning artificial languages. From our perspective, the experiments suggest that languages have evolved these and other properties to facilitate learning. Over time this process of linguistic adaptation has resulted in the structural constraints on language use that we observe today. The association between sequential learning and natural language is further evidenced by ALL experiments that demonstrate the accompanying breakdown of sequential skills in agrammatics (Christiansen et al., in preparation). Related artificial language experiments have also demonstrated that non-human primates can achieve relative proficiency in complex sequential tasks. This is not surprising in our view, since the fundamental role of sequential learning implies a long phylogenetic history: Primate studies in natural contexts have readily provided evidence of these complex sequential abilities.

In a similar vein, computational models allow researchers to explore and test hypotheses about factors contributing to language evolution in maximally controlled circumstances. Many of the models discussed in this paper have incorporated sequential learning mechanisms that shape language evolution. Put simply, these models can be viewed as investigations into the constraints on language imposed by sequential learning in a social environment. The results of these computational efforts dovetail with our view. We discussed Christiansen and Devlin's (1997) simulation, which showed how constraints on sequential learning can explain basic word order constraints. Also, Ellefson and Christiansen (2000), provided an explanation, based on sequential learning constraints, for why subjacency violations tend to be avoided across the languages of the world. Together, the results from all these computational models suggest that constraints arising from general cognitive processes, such as sequential learning and processing, are likely to play a larger role in language evolution than has traditionally been assumed. What we observe today as linguistic universals may be stable states that have emerged through an extended process of linguistic evolution.

As is customary in many scientific endeavors, the sources of evidence reported here abstracts away from many potentially important details. First, ALL experiments provide a highly idealized window into real language acquisition and processing. The artificial languages are highly simplified and often lack the social context in which language is normally acquired. Nonetheless, ALL studies have yielded important insights into the acquisition of language (for a review, see Gomez & Gerken, 2000). The computational modeling of language evolution shares many of the same limitations as ALL. For example, most of the modeled communication systems are equally simplified and embedded within a social context that is often reduced to a collection of abstract semantic features. However, we see the limitations on current use of ALL and computational modeling as unavoidable growing pains associated with a field very much in its infancy. Both lines of research methodologies are essential sources of information for furthering our understanding of language evolution. Their limitations underscore the necessity of an interdisciplinary approach. Only by amassing converging evidence from multiple lines of investigation can evolutionary hypotheses be supported, or discarded. Computational modeling and ALL experiments, rendered more sophisticated and naturalistic, hold considerable promise as two essential sources of evidence for studying language evolution.

References

Aitchison J (1998) On discontinuing the continuity-discontinuity debate. In: Hurford J, Studdert-Kennedy M, Knight C (eds) (1998) *Approaches to the evolution of language.* Cambridge University Press, Cambridge UK, pp 17-29

Batali J (1998) Computational simulations of the emergence of grammar. In: Hurford J, Studdert-Kennedy M, Knight C (eds) (1998) *Approaches to the evolution of language.* Cambridge University Press, Cambridge UK, pp 405-426

Burnett J (1773) *The origin and process of language (vols I-IV).* A Kincaid, Edinburgh

Byrne RW, Russon AE (1998) Learning by imitation: A hierarchical approach. *Behavioral & Brain Sciences*, 21: 667-721

Chomsky N (1986) *Knowledge of language.* Praeger, New York

Christiansen MH (1994) *Infinite languages finite minds: Connectionism learning and linguistic structure.* Unpublished doctoral dissertation, Centre for Cognitive Science, University of Edinburgh UK

Christiansen MH (2000) Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In: Dessalles JL, Ghadakpour L (eds)

The Evolution of Language: 3rd International Conference. Ecole Nationale Superieure des Telecommunications, Paris, pp 45-48

Christiansen MH, Chater N (1999) Toward a connectionist model of recursion in human linguistic performance . *Cognitive Science*, 23: 157-205

Christiansen MH, Chater N (in preparation) *Language as an organism: Language evolution as the adaptation of linguistic structure*. Manuscript in preparation

Christiansen MH, Devlin JT (1997) Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations. In: *Proceedings of the 19th Annual Cognitive Science Society Conference*. Lawrence Erlbaum Associates, Mahwah NJ, pp 113-118

Christiansen MH, Kelly L, Shillcock RC, Greenfield K (in preparation) *Artificial grammar learning in agrammatism*. Manuscript in preparation

Cleeremans A (1993) *Mechanisms of implicit learning: Connectionist models of sequence processing*. MIT Press, Cambridge MA

Conway C, Christiansen MH (2001) *Sequential learning in non-human primates*. Manuscript submitted for publication

Corballis MC (1992) On the evolution of language and generativity. *Cognition*, 44: 197-226

Davies AM (1987) 'Organic' and 'organism' in Franz Bopp. In: Hoenigswald HM, Wiener LF (eds) Biological metaphor and cladistic classification. University of Pennsylvania Press, Philadelphia PA, pp 81-107

Deacon TW (1997) *The symbolic species: The co-evolution of language and the brain*. WW Norton, New York

Dryer MS (1992) The Greenbergian word order correlations. *Language*, 68: 81-138

Ellefson MR, Christiansen M H (2000) Subjacency constraints without universal grammar: Evidence from artificial language learning and connectionist modeling. In: *The proceedings of the 22nd annual conference of the cognitive science society*, Erlbaum, Mahwah NJ, pp 645-650

Elman JL (1990) Finding structure in time. *Cognitive Science*, 14: 179-211

Givón T (1998) On the co-evolution of language, mind and brain *Evolution of Communication*, 2: 45-116

Gomez R L & Gerken L (2000) Infant artificial language learning and language acquisition. *Trends in Cognitive Science*, 4(5): 178-186

Grossman M (1980) A central processor for hierarchically structured material: Evidence from Broca’s aphasia. *Neuropsychologia*, 18: 299-308

Hauser MD, Newport EL, Aslin RN (2001) Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, 78: 53-64

Hawkins JA (1994) *A performance theory of order and constituency*. Cambridge University Press, Cambridge UK

Johnson-Pynn J, Fragaszy DM, Hirsh EM, Brakke KE, Greenfield PM (1999) Strategies used to combine seriated cups by chimpanzees (Pan troglodytes), bonobos (Pan paniscus) and capuchins (Cebus apella). *Journal of Comparative Psychology*, 113: 137-148

Kellogg WN, Kellogg LA (1933) *The ape and the child: A study of environmental influence on early behavior*. Hafner, New York

Kimura D (1988) Review of what the hands reveal about the brain. *Language and Speech*, 31: 375-378

Kiparsky P (1976) Historical linguistics and the origin of language. *Annals of the New York Academy of Sciences*, 280: 97-103

Kirby S (1998) Fitness and the selective adaptation of language In: Hurford J, Studdert-Kennedy M, Knight C (eds) (1998) *Approaches to the evolution of language*. Cambridge University Press, Cambridge UK, pp 359-383

Kirby S (2000) Language evolution without natural selection: From vocabulary to syntax in a population of learners. In: Knight C, Hurford JR, Studdert-Kennedy M (eds) *The evolutionary emergence of language: Social function and the origins of linguistic form*. Cambridge University Press, Cambridge UK, pp 303-323

Kirby S (2001) Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation and Cognitive Science*, 5(2): 102-110

Langacker RW (1987) *Foundations of Cognitive Grammar: Theoretical perspectives Vol 1*. Stanford University Press, Stanford CA

Livingstone D (2000) A modified-neural theory for the evolution of linguistic diversity. Paper presented at *The Evolution of Language: 3rd International Conference*, Paris, France

Livingstone D, Fyfe C (1999) Modelling the evolution of linguistic diversity. In In: Floreano D, Nicoud JD, Mondada F (eds) *Proceedings of ECAL99 the Fifth European Conference on Artificial Life (Lecture Notes in Artificial Intelligence)* Springer-Verlag, Berlin, pp 704-708

Lock A, Colombo M (1996) Cognitive abilities in a comparative perspective. In: Lock A, Peters CR (eds) *Handbook of human symbolic evolution*. Clarendon Press, Oxford, pp 596-643

Maess B, Koelsch S, Gunter TC, Friederici AD (2001) Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, 4: 540-545

McMahon AMS (1994) *Understanding language change*. Cambridge University Press, Cambridge UK

Morgan JL, Demuth K (eds) (1996) *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Lawrence Erlbaum, Mahwah, NJ

Morgan JL, Meier RP, Newport EL (1987) Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, 19: 498-550

Nerlich B (1989) The evolution of the concept of 'linguistic evolution' in the 19th and 20th century. *Lingua*, 77: 101-112

Newmeyer F (1991) Functional explanation in linguistics and the origins of language. *Language and Communication*, 11: 3–28

Nowak MA, Komarova NL, Niyogi P (2001) Evolution of universal grammar. *Science*, 291: 114-118

Ohshiba N (1997) Memorization of serial items by Japanese monkeys a chimpanzee and humans. *Japanese Psychological Research*, 39: 236-252

Orlov T, Yakovlev V, Hochstein S, Zohary E (2000) Macaque monkeys categorize images by their ordinal number. *Nature*, 404: 77-80

Osterhout L, Holcomb PJ (1992) Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory & Language*, 31:785-806

Patel AD, Gibson E, Ratner J, Besson M, Holcomb PJ (1998) Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10: 717-733

Patterson FG (1978) The gestures of a gorilla: Language acquisition in another pongid. *Brain and Language*, 5: 72-97

Percival WK (1987) Biological analogy in the study of languages before the advent of comparative grammar. In Hoenigswald HM, Wiener LF (eds) Biological metaphor and cladistic classification. University of Pennsylvania Press, Philadelphia PA , pp 3-38

Piatelli-Palmarini M (1989) Evolution selection and cognition: From "learning" to parameter setting in biology and in the study of language. *Cognition*, 31: 1-44

Pinker S (1994) *The language instinct: How the mind creates language*. William Morrow, New York

Pinker S, Bloom P (1990) Natural language and natural selection. *Behavioral and Brain Sciences*, 13: 707-727

Reber AS (1993) *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford University Press, New York

Saffran JR (2000) Non-linguistic constraints on the acquisition of phrase structure. In: Gleitman LR, Joshi AK (eds) *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Erlbaum, Mahwah NJ, pp 417-422

Saffran JR (2001) The use of predictive dependencies in language learning. *Journal of Memory and Language*, 44: 493-515

Saffran JR, Aslin RN, Newport EL (1996) Statistical learning by 8-month-old infants. *Science*, 274: 1926-1928

Savage-Rumbaugh S, Shanker SG, Taylor TJ (1998) *Apes language and the human mind*. Oxford University Press, New York

Schleicher A (1863) *Die Darwinsche Theorie und die Sprachwissenschaft*. Böhlau, Weimer

Spinozzi G, Langer J (1999) Spontaneous classification in action by a human-enculturated and language-reared bonobo (Pan paniscus) and common chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, 113 : 286-296

Steinhauer K, Friederici AD, Pfeifer E (2001) ERP recordings while listening to syntax errors in an artificial language: Evidence from trained and untrained subjects. Poster presented at the *14th Annual CUNY Conference on Human Sentence Processing*, Philadelphia PA

Stevick RD (1963) The biological model and historical linguistics. *Language*, 39: 159-169

Swartz KB, Chen S, Terrace HS (2000) Serial learning by rhesus monkeys: II. Learning four-item lists by trial and error. *Journal of Experimental Psychology: Animal Behavior Processes*, 26: 274-285

Van Everbroeck E (1999) Language type frequency and learnability: A connectionist appraisal. In: *Proceedings of the 21st Annual Cognitive Science Society Conference*. Erlbaum, Mahwah NJ, pp 755-760

Webb J (1669) *An historical essay endeavouring the probability that the language of the empire of china is the primitive language*. London