Language as Shaped by the Brain

Morten H. Christiansen
Department of Psychology
Cornell University
Ithaca, NY 14853
email: mhc27@cornell.edu

Nick Chater
Department of Psychology
University College London
London, WC1E 6BT
email: n.chater@ucl.ac.uk

and

Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501

Abstract

It is widely assumed that human learning and the structure of human languages are intimately related. This relationship is frequently suggested to be rooted in a language-specific biological endowment, which encodes universal, but arbitrary, principles of language structure (a universal grammar or UG). How might such a UG have evolved? We argue that UG could not have arisen either by biological adaptation or non-adaptationist genetic processes. The resulting puzzle concerning the origin of UG we call the *logical problem of language evolution*. Because the processes of language change are much more rapid than processes of genetic change, language constitutes a "moving target" both over time and across different human populations, and hence cannot provide a stable environment to which UG genes could have adapted. We conclude that a biologically determined UG is not evolutionarily viable. Instead, the original motivation for UG—the mesh between learners and languages—arises because language has been shaped to fit the human brain, rather than vice versa. Following Darwin, we view language itself as a complex and interdependent "organism," which evolves under selectional pressures from human learning and processing mechanisms. That is, languages are themselves undergoing severe selectional pressure from each generation of language users and learners. This suggests that apparently arbitrary aspects of linguistic structure may result from general learning and processing biases, independent of language. We illustrate how this framework can integrate evidence from different literatures and methodologies to explain core linguistic phenomena, including binding constraints, word order universals, and diachronic language change.

1. Introduction

Natural language constitutes one of the most complex aspects of human cognition, yet children routinely and rapidly learn to produce and understand it without any apparent effort. The rapidity of acquisition suggests that when the child makes a 'guess' about the structure of language on the basis of apparently limited evidence, she has an uncanny tendency to guess right. This strongly suggests that there must be a close relationship between the mechanisms by which the child acquires and processes the language, and the structure of language itself.

What is the origin of this presumed close relationship between the mechanisms children use in acquisition and the structure of language? One view is that specialized brain mechanisms adapted for language acquisition have been developed over long periods of natural selection (e.g., Pinker & Bloom, 1990). A second view rejects the idea that these specialized brain mechanisms have arisen through adaptation, and assumes that they have evolved through some non-adaptationist route, just as it has been argued that many biological properties are not the

product of adaptation (e.g., Bickerton, 1995; Gould, 1993; Jenkins, 2000). Both these viewpoints put the explanatory emphasis on putative specialist brain mechanisms—and ask how it is that these mechanisms are so well suited to learning natural language.

In this paper, we develop and argue for a third view, which takes the opposite starting point. It asks not, Why is the brain so well suited to learning language?, but instead, Why is language so well suited to being learned by the brain? We propose that *language* has adapted through a process of gradual change to be easy to learn to produce and understand. Thus the structure of human language must inevitably be shaped around human learning and processing biases. Language is easy for us to learn to produce and understand, not because our brains embody knowledge of language (e.g., in the form of an innate universal grammar), but because language has adapted to us. We argue that it is useful to view languages as 'organisms'; i.e., highly complex systems of interconnected constraints, that have evolved in a symbiotic relationship with humans. According to this view, whatever learning and processing biases people happen to have will tend to become embedded in the structure of language—because it will be easier to learn to understand and produce languages, or specific linguistic forms, that fit these biases.

The two main sections of this paper deal in turn with each of the issues outlined above. *The logical problem of language evolution* outlines and evaluates the adaptationist, non-adaptationist and finally the language-as-an-organism perspective on the evolution of language. The *Language as shaped by the brain: Case Studies* illustrates the "pay-off" of this viewpoint for understanding core linguistic phenomena, from binding constraints, to word order universals and subjacency, to the relationship between language change and language evolution. These cases are not meant to provide exhaustive explanations of these phenomena but, rather, are meant as preliminary demonstrations of how evidence from different, and currently largely separate, literatures on typology, grammaticalization, connectionist modeling, cognitive psychology, and construction-based grammar can fruitfully be integrated within our framework. Finally, we briefly conclude, noting the wider implications of this viewpoint, especially in radically recasting the problem of language acquisition.

2. The Logical Problem of Language Evolution

For a period spanning three decades, Chomsky (1965, 1972, 1980, 1986, 1988, 1993) has argued that a substantial innate endowment of language specific knowledge is necessary in order to provide sufficient constraints on language acquisition. These constraints form a *'Universal Grammar'* (UG); that is, an innate database consisting of a collection of universal grammatical principles that hold across all human languages. In this framework, language 'learning' involves setting of a number of parameters in UG according to the specifics of the particular language being learned. That is, a child's language ability gradually unfolds according to a genetic blueprint in much the same way as a chicken grows a wing. The staunchest proponents of this view even go as far as to claim that "doubting that there are language-specific, innate computational capacities today is a bit like being still dubious about the very existence of molecules, in spite of the awesome progress of molecular biology" (Piattelli-Palmarini, 1994: p. 335).

A central assumption of the theory of UG is that the universal principles governing language are arbitrary—i.e., not determined by functional considerations. That is, these principles cannot be explained in terms of learning, cognitive constraints, or communicative effectiveness. For example, consider the principles of binding, which have come to play a key role in generative linguistics (Chomsky, 1981). The principles of binding capture patterns of use of, among other

things, reflexive pronouns (e.g., *himself, themselves*) and accusative pronouns (*him*, *them*, etc.), which appear, at first sight at least, to defy functional explanation. For instance, consider examples (1)-(4), where the subscripts indicate co-reference, and asterisks indicate ungrammaticality.

(1) John$_i$ sees himself$_i$
(2) *John$_i$ sees him$_i$
(3) John$_i$ said he$_{i/j}$ won
(4) *He$_i$ said John$_i$ won

In (1), the pronoun *himself* must refer to John; in (2) it cannot. In (3), the pronoun *he* may refer to John or to another person; in (4), it cannot refer to John. These and many other cases, which we discuss further below, indicate that an extremely rich set of patterns govern the behavior of pronouns; and these patterns appear arbitrary—it appears that any number of alternative patterns would, from a functional standpoint, serve equally well. These patterns are captured, in modern generative linguistics, by the apparently arbitrary principles of binding theory (Chomsky, 1981); and this theory is often viewed as a core part of the innately specified UG. Later, we shall argue for a radically different viewpoint: that binding constraints have been shaped by the brain—and specifically by particular cognitive and pragmatic constraints underpinning language acquisition and processing. For now, though, we suppose that binding constraints, while apparently universal across natural languages, are entirely arbitrary—and hence may, according to Chomsky's viewpoint, be presumed to be part of the genetically coded UG.

If we suppose that the principles of UG are genetically specified, then this raises the question of the evolutionary origin of this genetic endowment. Two views have been proposed.

Adaptationists emphasize a gradual evolution of the human language faculty through *natural selection* (e.g., Briscoe, 2003; Corballis, 1992, 2003; Dunbar, 2003; Greenfield, 1991; Hurford, 1991; Jackendoff, 2002; Nowak, Komarova & Niyogi, 2001; Pinker, 1994, 2003; Pinker & Bloom, 1990). Linguistic ability confers added reproductive fitness, leading to a selective pressure for language genes; and richer language genes encode increasingly elaborate grammars.

Non-adaptationists (e.g., Bickerton, 1995—but see Bickerton, 2003; Chomsky, 1988; Jenkins, 2000; Lanyon, 2006; Piattelli-Palmarini, 1989) suggest that natural selection only played a minor role in the emergence of language in humans, focusing instead on a variety of alternative possible mechanisms.

We argue that both of these views, as currently formulated, face profound theoretical difficulties resulting in a 'logical problem of language evolution'. We argue that this is because, on analysis, it is mysterious how a highly elaborate fixed biological structure could become linked to what was, at least initially, a cultural product, proto-language, that must have been highly variable both over time and geographical locations.

In this section, we outline the first half of the argument of this paper: that adaptationist and non-adaptationist viewpoints face severe conceptual difficulties, and hence there is no currently viable account of how the brain came to be shaped for arbitrary language universals. We argue that the brain has *not* been so shaped—and hence neither adaptationist nor non-adaptationist solutions are required. Instead, language has been shaped around the brain; language reflects pre-existing, and hence non-language-specific, human learning and processing mechanisms. This

section provides the starting point for the next, in which we illustrate this viewpoint with case studies concerning core linguistic phenomena.

2.1. Linguistically-driven biological adaptation as the origin of innate universal grammar

The adaptationist position is probably the most widely held view of the origin of UG. We first describe adaptationism in biology and its proposed application to UG before outlining three conceptual difficulties for adaptationist explanations of language evolution.

2.1.1 Adaptation: The very idea
Adaptation is a candidate explanation for the origin of any innate biological structure. In general, the idea is that natural selection has favored genes that code for biological structures that increase 'fitness' (in terms of expected numbers of viable offspring).[1] Typically, a biological structure contributes to fitness by fulfilling some purpose—the heart is assumed to pump blood, the legs to provide locomotion, or UG to support language acquisition. If so, natural selection will generally favor biological structures that fulfill their purpose well, so that, over the generations, hearts will become well-adapted to pumping blood; legs well-adapted to locomotion; and any presumed biological endowment for language acquisition will become well-adapted to acquiring language.

Perhaps the most influential statement of the adaptationist viewpoint is by Pinker and Bloom (1990). They argue that "*natural selection is the only scientific explanation of adaptive complexity*. 'Adaptive complexity' describes any system composed of many interacting parts where the details of the parts' structure and arrangement suggest design to fulfill some function" (p. 709; their emphasis). As another example of adaptive complexity, they refer to the exquisite optical and computational sophistication of the vertebrate visual system. Pinker and Bloom note that such a complex and intricate mechanism has an extremely low probability of occurring by chance. Whatever the influence of non-adaptational factors (see below), they argue that there must additionally have been substantial adaptation to fine-tune a system as complex as the visual system. Given that language appears to be of comparable complexity equal to the visual system, Pinker and Bloom conclude that it is also highly improbable that language is the entirely the product of some nonadaptationist process (see also Pinker, 2003).

The scope and validity of the adaptationist viewpoint in biology is controversial (e.g., Dawkins, 1986; Gould, 2002; Gould & Lewontin, 1979; Hecht Orzak & Sober, 2001); and such controversy may be used to question adaptationist views of the origin of UG (e.g., Bickerton, 1995; Lewontin, 1998). Here, we take a different tack. We argue that, whatever the merits of adaptationist explanation in general, and as applied to vision in particular, the adaptationist account cannot extend to a putative UG.

2.1.2. Why universal grammar could not be an adaptation to language
Let us suppose that a genetic encoding of arbitrary universal properties of language did, as the adaptationist view holds, arise as an adaptation to the environment. Specifically, this adaptation is, of course, to be to the *linguistic* environment. But how could arbitrary properties of language be adaptive? Pinker and Bloom (1990) suggest that the constraints imposed by UG, such as the

---

[1] Strictly, the appropriate measure is the more subtle 'inclusive' fitness, which takes into account the reproductive potential not just of an organism, but also a weighted sum of the reproductive potentials of its kin, where the weighting is determined by the closeness of kinship (Hamilton, 1964). Moreover, mere reproduction is only of value to the degree that one's offspring have a propensity to reproduce; and so down the generations.

binding constraints mentioned above, can be construed as communicative protocols. As such, the specific nature of these standards does not matter as long as everyone (within a given speech community) adopts the same set of standards. For example, when using a modem it is important to use the right protocol, such as odd parity, handshake on, 7 bit, etc. There are many other combinations of settings that would be just as effective. What is important is that the computers that are to communicate with each other adopt the *same* protocol—otherwise communication will not be possible at all. Thus, when it comes to the specifics of UG, Pinker and Bloom suggest that "in the evolution of the language faculty, many 'arbitrary' constraints may have been selected simply because they defined parts of a standardized communicative code in the brains of some critical mass of speakers" (1990: p. 718)[2].

We will argue that this viewpoint faces three fundamental difficulties, concerning the dispersion of hominid populations; language change; and the question of *what* is genetically encoded. We briefly consider these in turn.

Problem 1: The dispersion of human populations
Pinker and Bloom's (1990) analogy with communications protocols, while apt, is, however, something of a double-edged sword. Communications protocols and other technical standards typically diverge rapidly unless there is concerted oversight and enforcement to maintain common standards. Maintaining and developing common standards is a substantial part of software and hardware development. In the absence of such pressures for standardization, standards would rapidly diverge. Given that language presumably evolved without top-down pressures for standardization, divergence between languages seem inevitable. To assume that 'universal' arbitrary features of language would emerge from adaptation by separate groups of language users, would be analogous to assuming that arbitrary features of communications protocols between computers might emerge from separate teams of scientists, working in separate laboratories (e.g., that different modem designers independently alight on odd parity, handshake on, 7 bit error correction, and so on). Note that this point would apply equally well, even if the teams of scientists emerged from a single group. Once cut off from each other, groups would develop in independent ways. Indeed, in biological adaptation, genes appear to rapidly evolve to deal with a specific local environment. Thus, Darwin observed rich patterns of variations in fauna (e.g., finches) across the Galapagos Islands, and interpreted these variations as adaptation to local island conditions. Hence, if language genes have adapted to local linguistic environments, we should expect a range of different biological encoded UGs, each specifically adapted to its local linguistic context. Indeed, one might expect, if anything, that language-genes would diverge especially rapidly—because the linguistic environment in each population is assumed to be itself shaped by the different language-genes in each subpopulation, thus amplifying the differences in linguistic environment. If this were the case, then people should have, at minimum, some specific predisposition to learn and process languages associated with their genetic lineage. This does not appear to be the case—and it is a key assumption of the generative linguistics perspective that the human language endowment does not vary in this way, and is universal across the species (Chomsky, 1980; Pinker, 1994).

The problem of divergent populations of language users arises across a range of different scenarios concerning the relationship between the evolution of language and the dispersion of

---

[2] In addition, Pinker and Bloom (1990) point out that it is often the case that natural selection has several (equally adaptive) alternatives to choose from to carry out a given function (e.g., both the invertebrate and the vertebrate eye support vision despite having significant architectural differences).

human populations. One scenario is that language evolution is recent, and occurred during the dispersion of modern humans (*homo sapiens sapiens*). In this case, whether language was discovered once, and then spread through human populations, or was discovered in various locations independently, there remains the problem that adaptations to language would not be coordinated across geographically dispersed groups. (It is tempting to suggest that all of these sublanguages will, nonetheless, obey universal grammatical principles, thus providing some constancy in the linguistic environment—but this appeal would, of course, be circular, as we are attempting to explain the *origin* of such principles.)

An alternative scenario is that language evolution pre-dates the dispersion of modern humans. If so, then it is conceivable that prior dispersions of hominid populations, perhaps within Africa, did lead to the emergence of diverse languages and diverse "universal grammars," adapted to learning and processing such languages; and then it is possible that a local population proved to be adaptively most successful, and came to displace other hominid populations. Thus, our current UG might conceivably be the only survivor of a larger family of such UGs—its universality would arise, then, because it was genetically encoded in the sub-population from which modern human descended[3]. This viewpoint is not without difficulties. In particular, the 100,000 years or so since the geographical dispersion of human populations would still seem to give time for substantial linguistic divergence, and, presuming that processes of genetic adaptation to language continued to operate[4], consequently result in different genetic bases for language. That is, the evolution of UG by adaptation would appear to require a process of adaptation for language prior to the dispersion of human populations; and an abrupt cessation of such adaptation after such dispersion.

Problem 2: Language change
Whatever the facts of the timing of the origin of language and hominid dispersion, the thesis that a genetically encoded UG arose through adaptation faces a second problem: that, even within a single population, linguistic conventions change rapidly. Hence the linguistic environment over which selectional pressures operate presents a "moving target" for natural selection. If linguistic conventions change more rapidly than genes change via natural selection, then genes that encode biases for particular conventions will be eliminated—because, as the language changes, the biases will be incorrect, and hence decrease fitness. More generally, in a fast changing environment, phenotypic flexibility to deal with various environments will typically be favored over genes that bias the phenotype narrowly toward a particular environment. Again, there is a tempting counter-argument—that the linguistic principles of UG will *not* change; and hence these aspects of language will provide a stable linguistic environment over which adaptation can operate. But, of course, this argument is circular—because the genetic endowment of UG is proposed to *explain* language universals; so it cannot be assumed that the language universals pre-date the emergence of the genetic basis for UG.

---

[3] One prominent view is that language emerged within the last 100,000 to 200,000 years (e.g., Corballis, 2003). Hominid populations over this period, and before, appear to have undergone waves of spread; "… modern languages derive mostly or completely from a single language spoken in East Africa around 100 kya … it was the only language then existing that survived and evolved with rapid differentiation and transformation." (Cavalli-Sforza & Feldman, 2003: p. 273)

[4] Human genome-wide scans have revealed evidence of recent positive selection for more than 250 genes (Voight, Kudaravalli, Wen & Pritchard, 2006), making it very likely that genetic adaptations for language would have continued in this scenario.
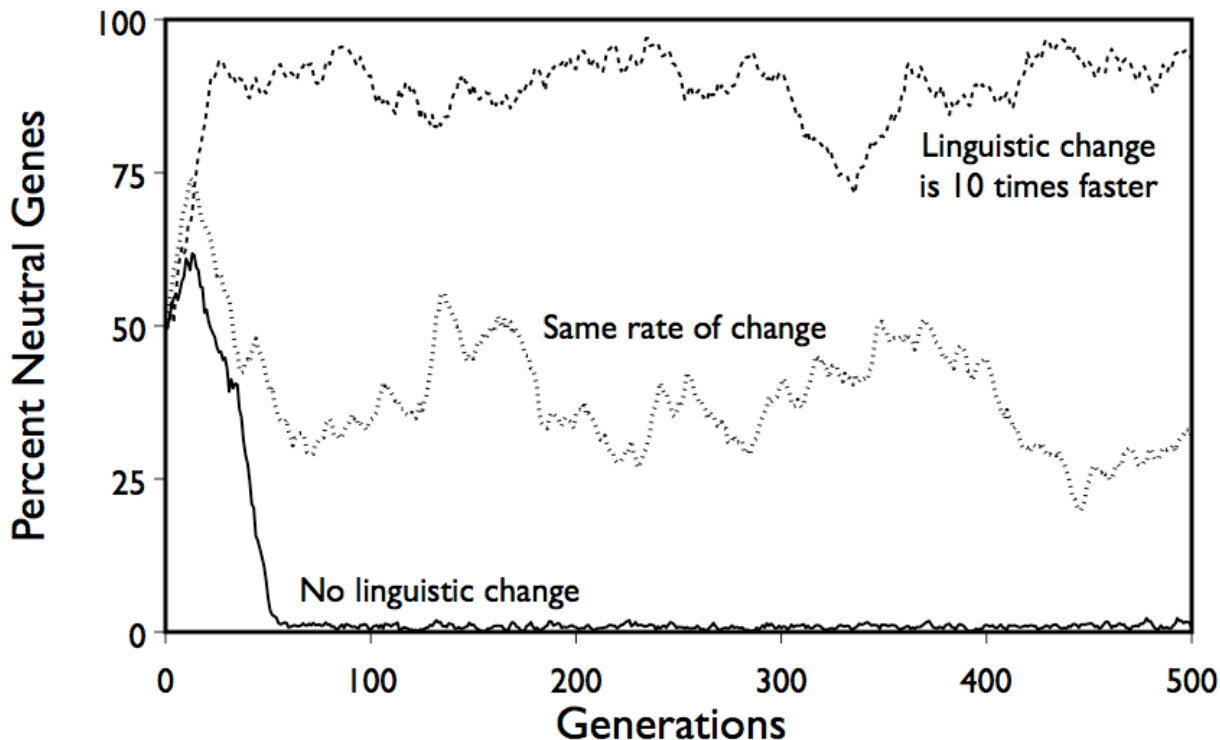
Figure 1. The effect of linguistic change on the genetic encoding of arbitrary linguistic principles. Results are shown from a simulation with a population size of 100 agents, a genome size of 20, survival of the top 50% of the population, and starting with 50% neutral genes. When there is no linguistic change, language genes emerge quickly but when language is allowed to change neutral genes become more advantageous. Similar results were obtained across a variety of different simulation parameters (Adapted from Christiansen, Reali & Chater, 2006).

Christiansen, Reali and Chater (2006) illustrate the problems raised by language change in a series of computer simulations. They assume the simplest possible set-up: that (binary) linguistic principles and language "genes" stand in one-to-one correspondence. Each gene has three alleles—one biased in favor of each version of the corresponding principle, and one neutral allele[5]. Agents learn the language by trial-and-error, where their guesses are biased according to which alleles they have. The fittest agents are allowed to reproduce, and a new generation of agents is produced by sexual recombination and mutation. When the language is fixed, there is a selection pressure in favor of the "correctly" biased genes, and these rapidly come to dominate the population, as illustrated by Figure 1. When, however, language is allowed to change, the effect reverses—biased genes are severely selected against when they are inconsistent with the linguistic environment, and neutral genes come to dominate the population. The selection in favor of neutral genes occurs even for low levels of language change (i.e., the effect occurs, to some degree, even if language change equals the rate of genetic mutation). But, of course, linguistic change (prior to any genetic encoding) is likely to have been much faster than this. After all, in the modern era, language change has been astonishingly rapid, leading, for example, to the wide phonological and syntactic diversity of the Indo-European language group, from a common ancestor perhaps 10,000 years ago (Gray & Atkinson, 2003). Language in hunter-

---

[5] This set-up closely resembles the one used by Hinton and Nolan (1987) in their simulations of the Baldwin effect, and to which Pinker and Bloom (1990) refer in support of their adaptationist account of language evolution.

gatherer societies changes at least as rapidly. Papua New Guinea, settled within the last 50,000 years has an estimated one quarter of the world's languages, and are enormously linguistically diverse, most of which originate in hunter-gatherer communities (Diamond, 1992)[6]. Thus, from the point of view of natural selection, it appears that language, like other cultural adaptations, changes far too rapidly to provide a stable target over which natural selection can operate.

Problem 3: What *is genetically encoded?*
Even if the first two difficulties for an adaptationist theory of the origin of a genetically specified UG could be solved, the view still faces a further puzzle: why is it that genetic adaptation occurred only to very abstract properties of language, rather than also occurring to its superficial properties? Given the spectacular variety of surface forms of the world's languages, in both syntax (including every combination of basic orderings of subject, verb and object, and a wide variety of less constrained word orders) and phonology (including tone and click languages, for example), why did not language genes adapt to these surface features? Why should genes become adapted to capture the extremely rich and abstract set of possibilities countenanced by the principles of UG, rather than merely encoding the actual linguistic possibilities in the specific language that was being spoken (i.e., the phonological inventory and particular morphosyntactic regularities of the early click-language from which the Khoisan family originated and which might be the first human language; e.g., Pennisi, 2004)? The unrelenting abstractness of the universal principles makes them difficult to reconcile with an adaptationist account.

One of the general features of biological adaptation is that it is driven by the constraints of the immediate environment—it can have no regard to distant or future environments that might one day be encountered. Hence, for example, the visual system is highly adapted to the laws of optics as they hold in normal environments. Thus, the human visual system mis-estimates the length of a stick in water, because it does not correct for the refraction of light through water (this being not commonly encountered in the human visual world). By contrast, the visual system of the archer fish, which must strike air-born flies with a water jet from below the water surface, does make this correction **(**Rossel, Corlija & Schuster, 2002). Biological adaptation produce systems designed to fit the environment in which adaptation occurs; there is, of course, no selectional pressure to fit environments that have not occurred, or might do so at some point in the future. Hence, if a UG did adapt to a past linguistic environment, it would seem inevitable that it would adapt to that language environment *as a whole*: thus adapting to its *specific* word order, phonotactic rules, inventory of phonemic distinctions, and so on. In particular, it seems very implausible that an emerging UG would be selected primarily for extremely abstract features, which apply equally to all possible human languages (not just the language evident in the linguistic environment in which selection operates). This would be analogous to an animal living in a desert environment somehow developing adaptations that are not specific to desert conditions, but that are equally adaptive in all terrestrial environments.

It is tempting to counter this argument by arguing that the principles of UG are just those that are invariant across languages, where contingent aspects of word order or phonology will vary across languages. Thus, we might suggest that only the highly abstract, language-universal, principles of UG will provide a stable basis upon which natural selection can operate. But this

---

[6] Some recent theorists have proposed that a further pressure for language divergence between groups is the sociolinguistic tendency for groups to 'badge' their in-group by difficult to fake linguistic idiosyncrasies (Baker, 2003; Nettle & Dunbar, 1997). Such pressures would increase the pace of language divergence, and thus exacerbate the problem of divergence for adaptationist theories of language evolution.

argument is again, of course, circular. We are trying to explain how a putative UG might become genetically fixed; and hence we cannot assume UG is already in place. Hence this counterargument is blocked.

In summary, Pinker and Bloom (1990), as we have seen, draw a parallel between the adaptationist account of the development of the visual system, and an adaptationist account of a putative language faculty. But the above argument indicates that the two cases are profoundly different. The principles of optics, and the structure of the visual world, have many invariant features across environments (e.g., Simoncelli & Olshausen, 2001); but the linguistic environment is vastly different from one population to another. Moreover, the linguistic environment, unlike the visual environment, will itself be altered in line with any genetic changes in the propensity to learn and use languages, thus amplifying differences between linguistic environments further. We conclude, then, that linguistically-driven biological adaptation cannot underlie the evolution of language.

It remains possible, though, that the development of language did have a substantial impact on biological evolution. The arguments we have given merely preclude the possibility that linguistic conventions that would originally *differ* across different linguistic environments could somehow become universal across all linguistic communities, by virtue of adaptation to the linguistic environment. This is because, in the relevant respects, the linguistic environment for the different populations is different, and hence any biological adaptations could only serve to entrench such differences further. But there might be features that are universal across linguistic environments (such as the means of producing speech; Lieberman, 1984; or the need for enhanced memory capacity, or complex pragmatic inferences; Givón & Malle, 2002) that might lead to biological adaptation. However, these language features are likely to be functional, to facilitate language *use*—and thus would typically not be considered part of UG.

It is consistent with our arguments that the emergence of language influenced biological evolution in a more indirect way. The possession of language might have fundamentally changed the patterns of collective problem solving and other social behavior of early humans, with a consequent shift in the selectional pressures on humans engaged in these new patterns of behavior. But what cannot emerge from biological adaptation to a varied pattern of linguistic environments are universal, arbitrary constraints on the structure of language. Thus, the adaptationist account of the biological origins of UG cannot succeed.

2.2 Non-adaptationist views of the origin of universal grammar

Some theorists advocating a strong language-specific genetic basis for language would concur with the conclusion of our arguments against adaptationist accounts of the evolution of language. For instance, Chomsky (1972, 1988, 1993) has for more than two decades expressed strong doubts about neo-Darwinian explanations of language evolution, and arguments for a radically non-adaptationist perspective have been advanced by Jenkins (2000) and Piattelli-Palmarini (1989, 1994).

Non-adaptationist positions rely on the complexity and intricacy of the putative UG as the premise for their arguments against adaptationist explanations of language evolution. UG appears to be so unique in terms of structure and properties, that it may seem unlikely to be a product of a process of natural selection amongst random mutations. We argue that, on the contrary, non-adaptationist attempts to explain a putative language-specific genetic endowment also fail. The implication is, of course, that there *is* little or no language-specific genetic endowment, as we discuss below—language has adapted to the brain, rather than the reverse.

Before turning to non-adaptationist explanation in language evolution, it is worth surveying proposed non-adaptationist mechanisms in biology. For example, *genetic drift* (Suzuki, Griffiths, Miller & Lewontin, 1989) refers to random fluctuations in gene frequencies in a population; *genetic hitch-hiking* (Maynard-Smith, 1978), that is, a mechanism by which non-selected genes might "catch a ride" with another gene that *was* selected for, if they are in close proximity to the selected gene along a chromosome; *spandrels* (Gould & Lewontin, 1979) are by-products with no previous function, but which come to serve some novel function; and *exaptation* (Gould & Vrba, 1982), that is, when something that was originally adapted to serve a particular function is put to use to serve a novel function.

To what extent can any non-adaptationist mechanism account for the development of a genetically encoded UG, as traditionally conceived? In particular, can such mechanisms account for the appearance of genetically specified principles that are presumed to be (a) idiosyncratic to language; and (b) of substantial complexity? We now argue that the probability that non-adaptationist factors played a substantial role in the evolution of UG is vanishingly small.

The argument involves a straightforward application of information theory. Suppose that constraints embodied in UG are indeed language-specific, and hence do not emerge as side-effects of existing processing mechanisms. This means that UG is generated *at random* by non-adaptationist processes.

Suppose that the information required to specify a language acquisition device, so that language can be acquired and produced, over and above the pre-linguistic biological endowment can be represented as a binary string of N bits (this particular coding assumption is purely for convenience). Then the probability of generating this sequence of N bits by chance is 2^{-N}. If the language-specific information could be specified using a binary string that would fit on one page of normal text (which would presumably be a considerable underestimate, from the perspective of most linguistic theory), then N would be over 2500. Hence the probability of generating the grammar by a random process would be less than 2^{-2500}. So to generate this machinery by chance (i.e., without the influence of the forces of adaptation) would be expected to require of the order of 2^{2500} individuals. But the total population of humans over the last two million or so years, including the present, is measured in billions, and is much smaller than 2^{20}. Hence, the probability of non-adaptationist mechanisms 'chancing' upon a specification of a language organ or language instinct through purely non-adaptationist means is astronomically unlikely[7].

The implication of this argument is that it is extremely unlikely that substantial quantities of linguistically idiosyncratic information have been specified by non-adaptationist means. Indeed, the point applies more generally to the generation of any complex, functional biological structures. Thus, it is not clear how any non-adaptationist account can explain the emergence of the level of complexity of UG.

Many authors who express skepticism concerning the role of adaptation implicitly recognize this kind of difficulty with the non-adaptationist viewpoint. That is, they recognize that it is unlikely that substantial amount of knowledge specific to language processing could have arisen through non-adaptationist processes (e.g., Chomsky, 1993). Instead, many apparently complex

---

[7] We have presented the argument in informal terms. A more rigorous argument is as follows. We can measure the amount of information embodied in universal grammar, U, over and above the information in pre-existing cognitive processes, C, by the length of the shortest code that will generate U from C. This is the conditional Kolmogorov complexity $K(U|C)$ (Li & Vitányi, 1997). By the coding theorem of Kolmogorov complexity theory (Li & Vitányi, 1997), the probability of randomly generating U from C is approximately $2^{-K(U|C)}$. Thus, if universal grammar has any substantial complexity, then it has a vanishingly small probability of being encountered by a random process, such as a non-adaptational mechanism.

and arbitrary aspects of cognition and language may have emerged out of the constraints on building any complex information processing system, given various physical and biological constraints. According to this picture, then, processes of language acquisition and processing would not emerge through non-adaptationist mechanisms of evolution; instead they would emerge spontaneously from physical and biological constraints (see Kaufmann, 1995, for a related viewpoint on evolutionary processes). Yet a different view is proposed by Gould (1993), who views language as a spandrel—i.e., as emerging as a byproduct of other cognitive processes.

These viewpoints, then, downplay the importance of non-adaptational mechanisms in specifying domain-specific information about language in the genome. And our argument above shows that only a very small amount of such information could plausibly have arisen through non-adaptationist mechanisms.

We may seem to be faced with a paradox. It seems clear that the mechanisms involved in acquiring and processing language are enormously intricate and moreover intimately connected to the structure of natural languages. The complexity of these mechanisms rules out, as we have seen in this section, a non-adaptationist account of their origin. But if these mechanisms arose through adaptation, this adaptation cannot, as we argued above, have been adaptation *to language*. But if the mechanisms that currently underpin language acquisition and processing were originally adapted to carry out other functions, then how is their apparently intimate relationship with the structure of natural language to be explained? How, for example, are we to explain that the language acquisition mechanisms seem particularly well-adapted to learning natural languages, but not to any of a vast range of conceivable non-natural languages (e.g. Chomsky, 1980)? As we now argue, the paradox can be resolved if we assume that the 'fit' between the mechanisms of language acquisition and processing and natural language has arisen because natural languages themselves have 'evolved' to be as easy to learn and process as possible.

2.3 The Adaptation of Linguistic Structure

We propose, then, to invert the perspective on language evolution, shifting the focus from the evolution of *language users* to the evolution of *languages* (see Figure 2). From this perspective, we argue that the mystery of the fit between human language acquisition and processing mechanisms and natural language may be unraveled; and we might, furthermore, understand how language attained its apparently 'idiosyncratic' structure. Instead of puzzling that humans can only learn a small subset of a huge set of possible languages, we take a different starting point: the observation that natural languages exist only because humans can produce, learn and process them. In order for languages to be passed on from generation to generation, they must adapt to the properties of the human learning and processing mechanisms; and the structures in each language form a highly interdependent *system*, rather than a collection of independent traits. We therefore propose to construe language as an *organism*, adapted through natural selection to fit a particular ecological niche: the human brain. Thus, the key to understanding the fit between language and the brain is understanding how language has been shaped by the brain, not the reverse.
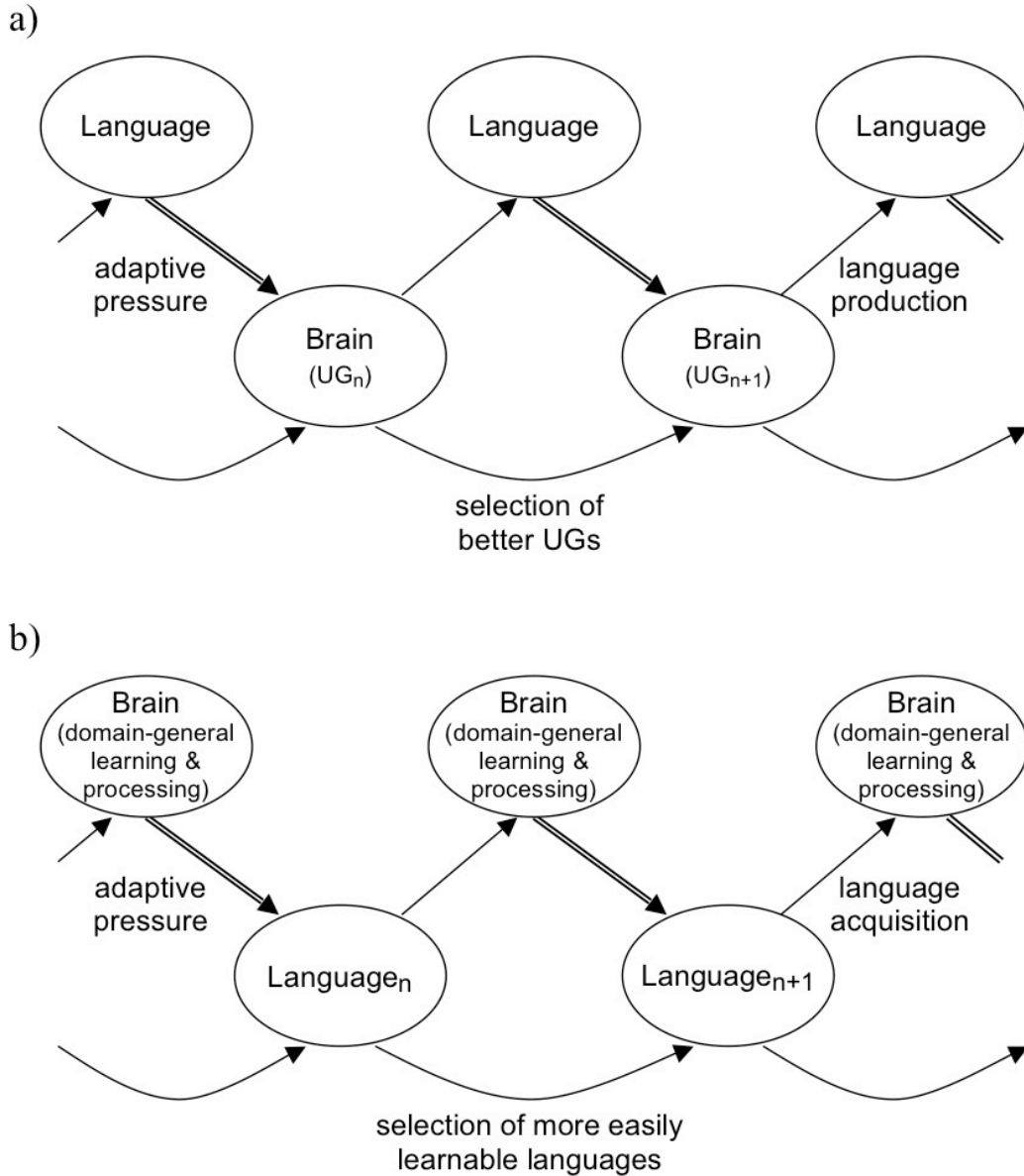
a)

b)

Figure 2. Illustration of two different views on the direction of causation in language evolution: a) biological adaptations of the brain to language, resulting in gradually more intricate UGs to provide the basis for increasingly more complex language production and comprehension; b) cultural adaptation of language to the brain, resulting in languages that are increasingly well suited to being acquired by domain-general learning and processing mechanisms.

This viewpoint does not imply that language does not provide selective pressure relevant to biological evolution. Good language skills may indeed enhance reproductive success. But the pressures working on language to adapt to 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 been shaped to fit

the human brain than the other way round. Languages that are hard for humans to learn cannot come into existence at all.

The biological perspective on language as an adaptive system has a prominent historical pedigree. One of the earliest proponents of the idea that languages evolve diachronically was the eighteenth-century language scholar, Sir William Jones, the first Western scholar to study Sanskrit and note its affinity with Greek and Latin (Cannon, 1991). Later, nineteenth-century linguistics was dominated by an organistic view of language (McMahon, 1994). Franz Bopp, one of the founders of comparative linguistics, regarded language as an organism that could be dissected and classified (Davies, 1987). Wilhelm von Humboldt—the father of generative grammar (Chomsky, 1965; Pinker, 1994)—argued that "… language, in direct conjunction with mental power, is a fully-fashioned *organism…"* (von Humboldt, 1836/1999, p. 90; original emphasis). 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; quoted 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 towards attainment of perfection, followed by decay.

Darwin (1900) also recognized the similarities between language and organisms[8]:

> The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel … We find in distinct languages striking homologies due to community of descent, and analogies due to a similar process of formation. The manner in which certain letters or sounds change when others change is very like correlated growth … Languages, like organic beings, can be classed in groups under groups; and they can be classed either naturally, according to descent, or artificially by other characters. Dominant languages and dialects spread widely, and lead to the gradual extinction of other tongues. A language, like a species, when once extinct, never … reappears … A struggle for life is constantly going on among the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand … The survival and preservation of certain favored words in the struggle for existence is natural selection. (p. 106)

In this sense, natural language is akin to an organism whose evolution has been constrained by the properties of human learning and processing mechanisms. A similar perspective on language evolution was revived, within a modern evolutionary framework, by Stevick (1963), and later by Nerlich (1989). Christiansen (1994) proposed that language be viewed as a kind of beneficial parasite that confers some selective advantage onto its human hosts without whom it cannot survive. Building on this work, Deacon (1997) further developed the analogy by construing language as a virus. More recently, Pennock (1999) has used the notion of language

---

[8] Darwin may have had several reasons for pointing to these similarities. Given that comparative linguistics at the time was considered to be a model science of the past on par with geology and comparative anatomy, he may have used comparisons between linguistic change—which was thought to be well understood at that time—and species change to corroborate his theory of evolution (Alter, 1998; Beer, 1996). Darwin may also have used these language-species comparisons to support the notion that less "civilized" human societies spoke less "civilized" languages, because he believed that this was predicted by his theory of human evolution (Raddick, 2000, 2002).

as an organism to argue the case of evolution by natural selection and against creationism, and van Driem (2005) proposed to view language in terms of meaning-based meme organisms.

But in what sense should language be viewed as an *organism*, rather than as a collection of independent traits, evolving relatively independently? The reason is that language is highly *systematic*—so much so, indeed, that much of linguistic theory is concerned with tracking the systematic relationships between different aspects of linguistic structure (see, for example, the discussion of word order and subjacency, below). Nonetheless, although language is an integrated system, it can, nonetheless, be viewed as comprising a complex set of 'features' or 'traits' which may or may not be passed on from one generation to the next (concerning lexical items, idioms, aspects of phonology, syntax and so on). Traits that are easy for learners to acquire and use will tend to become more prevalent; traits that are more difficult to acquire and use will tend to disappear. Thus, selectional pressure from language learners and users will shape the way in which the language evolves. Crucially, the systematic character of linguistic traits means that, to some degree at least, the fates of different traits in a language are intertwined. That is, the degree to which any particular trait is easy to learn or process will, to some degree, depend on the other features of the language—because language users will tend to learn and process each aspect of the language in the light of their experience with the rest. This picture is familiar in biology—the selectional impact of any gene depends crucially on the rest of the genome; and the selectional forces on each gene, for good or ill, are tied to the development and functioning of the entire organism.

To spell out the parallel, the idiolect of an individual speaker is analogous to an individual organism; a language (e.g., Mandarin, French) is akin to a species. A linguistic 'genotype' corresponds to the neural representation of an idiolect, instantiated by a collection of mental 'constructions' (in the sense of construction grammar; e.g., Goldberg, 2006) analogous to genes, and gives rise to linguistic behavior—the language 'phenotype'—characterized by a collection of utterances and interpretations. Just as the fitness of an individual gene depends on its interaction with other genes, so is the fitness landscape of an individual construction intertwined with those of other constructions; i.e., constructions are part of a (linguistic) system. A species in biology can interbreed; a 'language species' is mutually intelligible. To a first approximation, interbreeding and mutually intelligible linguistic interactions are the processes by which genetic material and constructions can propagate (see Sereno, 1991).

Pushing the analogy a little further, we follow Darwin in proposing language as a kind of beneficial parasite—technically, a *nonobligate symbiant*—that confers some advantages (including, presumably, selective advantages) to its human hosts, without whom it cannot survive. Symbiotic parasites and their hosts tend to become increasingly co-adapted (e.g., Dawkins, 1976). But note that the co-adaptation will be very lopsided, because the rate of linguistic change is far greater than the rate of biological change. Whereas Danish and Hindi needed less than 5,000 years to evolve from a common hypothesized proto-Indo-European ancestor into very different languages (McMahon, 1994), it took our remote ancestors approximately 100,000–200,000 years to evolve from the archaic form of *Homo sapiens* into the anatomically modern form, sometimes termed *Homo sapiens sapiens*. Indeed, as we argued above, the rapidity of language change, and the geographical dispersal of humanity suggests that the biological adaptation to language is negligible. This suggestion is further corroborated by work in evolutionary game theory, showing that when two species with markedly different rates of adaptation enter into a symbiotic relationship, the rapidly evolving species becomes highly cooperative and ends up being enslaved by the slowly evolving one (Frean & Abraham, 2004).

In the first part of our argument, we have argued that language is too variable, both in time and space, to provide a selectional pressure that might shape the gradual adaptation of an innate UG encoding arbitrary, but universal linguistic principles; and that such a putative innate UG would be too complex to credibly have arisen through non-adaptive mechanisms. Instead, we have proposed that the fit between language and the brain arises because language has evolved to be readily learned and processed by the brain. We now consider a series of case studies to illustrate how viewing language as shaped by the brain, rather than vice versa, provides an insightful perspective on human language.

3. Language as Shaped by the Brain: Case Studies

If language has adapted to the cognitive constraints of language learners and languages users, then this raises the question of the extent to which those constraints can be identified. That is, to what extent can linguistic structure previously ascribed to an innate UG be identified as having a cognitive basis? Clearly, establishing an answer to this question would require a vast program of research. In this section, we present three case studies to demonstrate that there are already indications that this framework for understanding language may be fruitful—and that some apparently arbitrary aspects of linguistic structure may be non-arbitrary with respect to language learning and processing. We focus on (i) how the apparently arbitrary constraints on binding might be explained in terms of processing limitations, (ii) how word order universals, and subjacency, may arise from cognitive constraints on sequential learning, and (iii) how processes of historical language change, including grammaticalization, may serve as a microcosm for language evolution.

3.1. Binding constraints as processing limitations

We noted above that binding, especially between reflexive and non-reflexive pronouns and noun phrases has been a theoretically central topic in generative linguistics (Chomsky, 1981); and the principles of binding appear both complex and arbitrary. Binding theory is thus a paradigm case of the type of information that has been proposed to be part of an innate UG. Binding therefore provides a challenge for theorists who do not assume UG. As we illustrate, however, there is a range of alternative approaches that provide a promising starting point for understanding binding as arising from processing and pragmatic factors. Let us begin with the following contrast (Black, 1999):

(5) That Sally$_i$ enjoyed herself$_i$ /*her$_i$ surprised her$_i$/*herself$_i$.

Why must the first pronoun be reflexive, but the second cannot be? According to generative grammar, the key concept here is *binding*. Roughly, a noun phrase *binds* a pronoun if it c-commands that pronoun, and they are co-referring. In an analogy between linguistic and family trees, an element c-commands its siblings and all their descendents. A noun phrase, NP, *A-binds* a pronoun if it binds it; and, roughly, if the NP is in either subject or object position. Now we can state simplified versions of Chomsky's three binding principles:

Principle A. Reflexives must be A-bound by an NP
Principle B. Pronouns must not be A-bound by an NP
Principle C. Full NPs must not be A-bound

Principle A says, roughly, that a reflexive pronoun (e.g., *herself*) must be used, if co-referring to a "structurally nearby" item (defined by c-command), in subject or object position. Principle B says, roughly, that a non-reflexive pronoun (e.g., *her*) must be used otherwise. These principles explain the pattern in (5). Principle C rules out co-reference such as (4). *John* cannot be bound to *he*. For the same reason, *John likes John*, or *the man likes John* do not allow co-reference between subject and object.

Need the apparently complex and arbitrary principles of binding theory be part of the child's innate UG? Or can these constraints be explained as a product of more basic processing or communicative constraints? One suggestion, due to O'Grady (2005), is that the language processing system seeks to resolve linguistic dependencies (e.g., between verbs and their arguments) at the first opportunity—a tendency that might not be specific to syntax, but might be an instance a general cognitive tendency to attempt to resolve ambiguities rapidly in linguistic (Clark, 1975) and perceptual input (Pomerantz & Kubovy, 1986). The use of a reflexive is assumed to signal that the pronoun co-refers with an available NP, given this dependency structure.

Thus, in parsing (5), the processor reaches *That Sally enjoyed herself*...and makes the first available dependency relationship, between the enjoyed, Sally, and herself. The use of the reflexive, *herself*, signals that co-reference with the available NP, *Sally*, is intended (cf. Principle A). The dependencies now resolved, the internal structure of the resulting clause is "closed off" and the parser moves on: *[That [Sally enjoyed herself]] surprised her/*herself*. The latter *herself* is not possible: because there is no appropriate noun phrase available to connect with (the only noun phrase is [*that Sally enjoyed herself*] which is used as argument to *surprised*, but which clearly cannot co-refer with the *herself*. But in *Sally enjoyed herself*, *Sally* is available as a noun-phrase when *herself* is encountered.

By contrast, plain pronouns can be viewed as *not* requiring that the immediate availability of the relevant constituent—and hence are used in complementary distribution to reflexive pronouns (cf. Principle B). It has been argued that this complementarity arises pragmatically (Levinson, 1987a; Reinhart, 1983); i.e., given that the use of reflexives is highly restrictive, they are, where appropriate more informative; hence, by not using them, the speaker signals that the co-reference is not appropriate (it is also possible, of course, that as with pragmatic patterns in general, this pattern may become increasingly conventionalized through use—a typical pattern in grammaticalization).

Finally, simple cases of Principle C can be explained by similar pragmatic arguments. Using *John sees John*, where the object can, in principle, refer to any individual named John, would be pragmatically infelicitous if co-reference were intended—because the speaker would have chosen a less informative option than using *himself* in object position. O'Grady (2005) and Reinhart (1983) consider more complex cases related to Principle C, in terms of a processing bias towards so-called 'upward feature-passing,' though we do not consider this here.

The linguistic phenomena involved in binding are extremely complex, and not fully captured by any theoretical account. We do not aim here to argue for any specific account of binding phenomena; but rather to indicate that many aspects of binding may arise from general processing or pragmatic constraints. To the extent that this is right, children may be able to acquire binding not on account of binding principles represented as part of an innate UG; but rather as results of processing and pragmatic constraints.

More generally, a growing bulk of work in computational and comparative linguistics has begun to substantiate the viewpoint that language structure emerges from processing constraints.

For example, Hawkins (1994, 2005) and Culicover (1999) propose specific measures of processing complexity (roughly, the number of linguistic constituents required to specify the correspondence between the syntactic structure of a sentence and a putative 'conceptual' structure), which they assume underpin judgments concerning linguistic acceptability. These proposals are particularly interesting in the present context, because they indicate an increasing emphasis on performance constraints, within the mainstream tradition in linguistics. Moreover, this work dovetails with recent work on how linguistic structure may arise from cognitive constraints on learning to which we turn next.

3.2 Word order universals as reflections of sequential learning constraints

Language processing involves extracting regularities from highly complex sequential input, pointing to an obvious connection between sequential learning and language: both involve the extraction and further processing of discrete elements occurring in complex temporal sequences. It is therefore not surprising that sequential learning tasks have become an important experimental paradigm for studying language acquisition and processing (sometimes under the guise of 'artificial grammar/language learning', Gómez & Gerken, 2000, or 'statistical learning', Saffran, 2003). Sequential learning have thus been demonstrated across a variety of language learning situations, including speech segmentation (Curtin, Mintz & Christiansen, 2005; Saffran Aslin & Newport, 1996; Saffran, Newport & Aslin, 1996), discovering complex word-internal structure between nonadjacent elements (Newport & Aslin, 2004; Onnis, Monaghan, Chater & Richmond, 2005; Peña, Bonnatti, Nespor & Mehler, 2002), acquiring gender-like morphological systems (Brooks, Braine, Catalano, Brody & Sudhalter, 1993; Frigo & McDonald, 1998), locating syntactic phrase boundaries (Saffran, 2001, 2002), using function words to delineate phrases (Green, 1979), integrating prosodic and morphological cues in the learning of phrase structure (Morgan, Meier & Newport, 1987), integrating phonological and distributional cues (Monaghan, Chater & Christiansen, 2005), and detecting long-distances relationships between words (Gómez, 2002; Onnis, Christiansen, Chater & Gómez, 2003).

If, as we have suggested, language has evolved to fit human sequential learning mechanisms, we might expect that constraints on the learning and processing of sequential structure may be reflected in the universal properties of human language. Importantly, many of the cognitive constraints that have shaped the evolution of language would still be at play in our current language ability. Thus, the study of how artificial sequential material is learned may reveal selectional pressures operating on the evolution of natural languages (e.g., Reber, 1992). After first discussing evidence pointing to common cognitive and neural mechanisms for sequential learning and language, we report on a series of modeling and experimental results that indicate how constraints on sequential learning may have given rise to certain word order universals relating to head-ordering, interactions between case and word order flexibility, and subjacency.

3.2.1 Shared mechanisms for sequential learning and language
If language acquisition and processing shares mechanisms with sequential learning in other domains, this suggests that breakdown of language should be associated with impaired sequential learning. This prediction is particularly interesting, because breakdown in sequential learning does not co-occur with many cognitive impairments but is generally considered "to remain robust in the face of time, lack of attentional resources and psychological disorder" (Cleeremans

et al., 1998, p. 407; Reber, 1992).[9] We tested this prediction using an artificial grammar learning task involving agrammatic aphasic patients who typically have damage in or around Broca's area and have severe problems with the hierarchical structure of sentences (Christiansen, Kelly, Shillcock & Greenfield, 2006). Although both aphasic patients and normal controls, matched for age, socio-economic status, and abstract reasoning abilities, were able to successfully complete a training task in which they had to say whether two successively presented symbol strings were the same or different, only the control group was able to correctly determine which of a set of novel strings were generated by the same rules as the training strings and which were not. These results are consistent with other studies showing that language disability (Plante, Gómez & Gerken, 2002) and specific language impairment (Hsu, Christiansen, Tomblin, Zhang & Gómez, 2006) are associated with impaired sequential learning. A more direct causal link between language and sequential learning was established by Hoen et al. (2003), who provided agrammatic aphasics with training on a visual sequence-learning task and found that increased performance on this task resulted in improvements in the patients' abilities to understand certain complex linguistic constructions.

More generally, the close relationship between sequential learning and grammatical ability has been further corroborated by recent neuroimaging studies with normal populations, showing that subjects trained on an artificial language have the same event-related potential (ERP) brainwave patterns to ungrammatical artificial-language sentences as to ungrammatical natural-language sentences (Friederici et al, 2002). Moreover, novel incongruent musical sequences elicit ERP patterns that are statistically indistinguishable from syntactic incongruities in language (Patel, Gibson, Ratner, Besson & Holcomb, 1998). Results from a magnetoencephalography (MEG) experiment further suggest that Broca's area plays a crucial role in the processing of music sequences (Maess, Koelsch, Gunter & Friederici, 2001). Finally, event-related functional magnetic resonance imaging (fMRI) results have shown that the same brain area—Broca's area—is involved in an artificial grammar learning task and in normal natural language processing (Petersson, Forkstam & Ingvar, 2004). Together, these studies indicate that there is considerable overlap in the neural mechanisms involved in language and sequential learning (see also Wilkins & Wakefield, 1995; Ullman, 2004, for similar perspectives).

3.2.2 Word-order universals

If sequential learning and language share common mechanisms, we might expect that basic word order universals might arise from constraints on sequential learning. To consider this question, let us begin with the *heads* of phrases: the word that determines the properties and meaning of the phrase as a whole (such as the noun *boy* in the noun phrase *'the boy with the bicycle'*). Across the world's languages, there is a statistical tendency toward a basic format in which the head of a phrase consistently is placed in the same position — either first or last — across different types of phrase. 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 in a transitive verb-phrase such as *'eat curry'*. A head-last language, such as Hindi, typically uses the opposite order, and hence the equivalent of *'curry eat'*. Likewise, head-first languages tend to have prepositions before the noun-phrase in prepositional phrases (such as

---

[9] Indeed, sequential learning is preserved across vastly different impaired populations, including those with Alzheimer's disease (Reber, Martinez & Weintraub, 2003), amnesia (Knowlton & Squire, 1996), closed-head injury (McDowall & Martin, 1996), Parkinson's disease (Meulemans, Peigneux & Van der Linden, 1998), and psychiatric problems such as schizophrenia, depression or chronic alcoholism (Abrams & Reber, 1988).

'with a fork'), whereas head-last languages tend to have postpositions following the noun-phrase in postpositional phrases (such as 'a fork with'). In the UG framework, head-ordering consistency has been explained by an innate module embodying the principles of X-bar theory, specifying constraints on the phrase structure of languages (e.g., Chomsky, 1986), and which has been suggested to be a product of natural selection (Pinker 1994).
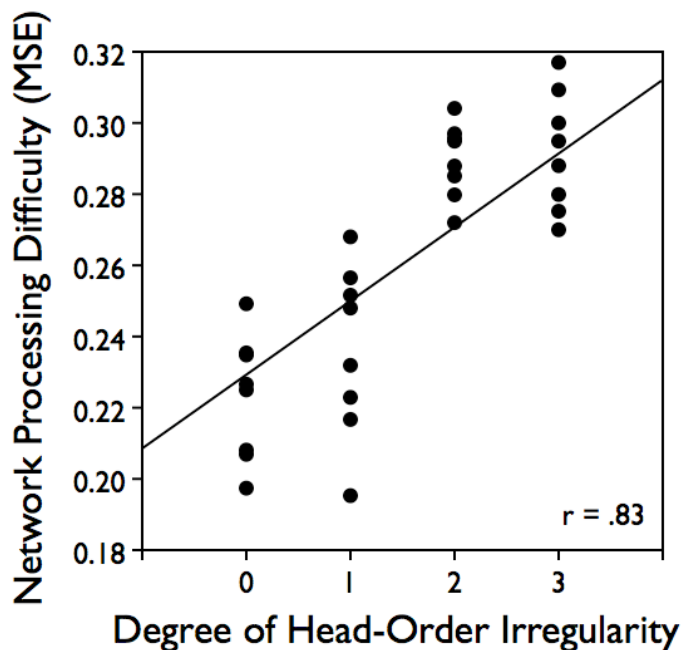

Figure 3. Using the degree of head-order irregularity to predict network-processing difficulty: higher degrees of head-order irregularity result in increased processing difficulty. (Adapted from Christiansen & Devlin, 1997).

A very different picture emerges if we consider that word order has evolved to fit human sequential learning mechanisms. To explore this viewpoint, we first consider computational simulations of sequential learning and how these illustrate how learning mechanisms can bias word order. We then consider experimental data on biases in human sequential learning, aimed at directly testing how far these conform to word-order universals. Christiansen and Devlin (1997) trained simple recurrent networks (Elman, 1990; SRN) on corpora generated by 32 different grammars that differed in head-order regularity (i.e., irregular grammars would mix head-first and head-last phrases). The networks were trained to predict the next lexical category in a sentence. Although these networks had no built-in linguistic biases, their predictions were sensitive to the amount of head-order regularity found in the grammars, such that there was a strong correlation between the degree of head-order regularity of a given grammar and the degree to which the network had learned to master the language. As shown in Figure 3, the more irregular the grammar, the more errorful the network's performance. Thus, the networks' sequential biases made the corpora generated by regular grammars considerably easier to acquire than the corpora generated from irregular grammars. Christiansen and Devlin further analyzed frequency data on the world's natural languages (from the FANAL database, Dryer, 1992) concerning the specific syntactic constructions used in the simulations. They found that languages incorporating fragments that the networks found hard to learn tended to be less

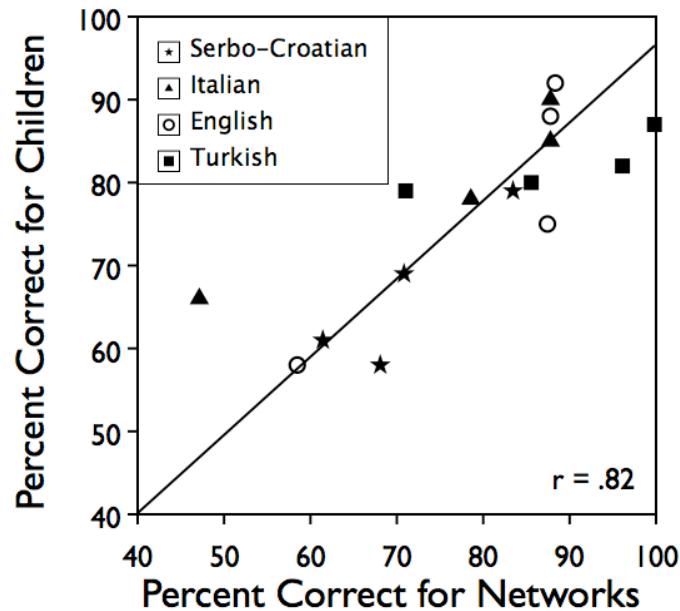frequent than languages the networks learned more easily.

Figure 4. Using network performance as a function of training to predict the improvements in children's performance with increasing age in Turkish, English, Italian and Serbo-Croatian. (Network results from Lupyan & Christiansen, 2002, and child data from Slobin & Bever, 1982).

In a similar vein, Van Everbroeck (1999) trained recurrent networks (a variation on the SRN) to produce the correct grammatical role assignments (i.e., who does what to whom) for noun-verb-noun sentences, presented one word at a time. Forty-two different language types were used to represent cross-linguistic variation in word order (e.g., subject-verb-object or SVO), and noun/verb inflection. Results of the simulations coincided with many observed trends in the distribution of the world's languages. Subject-first languages, which make up the majority of language types (SOV: 51% and SVO: 23%), were easily learned by the networks. Object-first languages, on the other hand, were not well learned, and have very low frequency in the world's languages (OVS: 0.75% and OSV: 0.25%). Van Everbroeck argued that these results were a product of network learning and processing constraints. Using rule-based language induction, Kirby (1999) arrived at a similar account of typological universals.

While highly encouraging, not all of Van Everbroeck's results were directly proportional to actual language-type frequencies. For example, VSO languages only account for 10% of the world's language types, but the model's performance on this word order exceeded performance on the more frequent subject-first languages. Using SRNs, Lupyan and Christiansen (2002) were able to fit language-type frequencies appropriately once they took case-markings into account. They were also able to model data from a study by Slobin and Bever (1982) showing differences in performance across English, Italian, Turkish, and Serbo-Croatian when children were asked to act out reversible transitive sentences, such as *'the horse kicked the cow'*, using familiar toy animals. Similar to the children, the networks initially showed the best performance in Turkish, with English and Italian quickly catching up, and with Serbo-Croatian lagging behind. The close

match between network performance across training and that of children across age is illustrated by Figure 4. Because of their regular use of case and word order, respectively, Turkish and English were more easily learned than Italian and, in particular, the highly irregular Serbo-Croatian language. With repeated exposure, the networks learning Serbo-Croatian eventually caught up, as do the children learning these languages.

To determine whether these sequential learning biases would result in the emergence of consistent head ordering across successive generations of learners, Reali and Christiansen (2006) trained SRNs to map words onto grammatical roles. Prior to the introduction of language, the SRNs were first evolved "biologically" to improve their ability to perform a sequential learning task. Specifically, the initial weights from the best learner at each generation were chosen as the basis for the next, with copies of the parent's weights mutated slightly. After 500 generations, the SRNs had evolved a considerably better ability to deal with sequential structure, following Conway & Christiansen's (2001) suggestion that improved sequential learning abilities may be an important preadaptation for language[10]. A language with no word order constraints was selected from a total of 729 languages with differing degrees of word order flexibility. Crucially, both language and networks were allowed to change while the networks at the same time also had to maintain the same level of performance on the sequential learning task as obtained after initial evolution of sequential learning biases (on the assumption that this skill would still have been crucial for hominid survival after the emergence of language). Over generations, a consistent head-ordering emerged due to linguistic adaptation rather than biological adaptations (of initial weights). Indeed, the pressure toward maintaining a high level of sequential learning performance prevented the SRNs from adapting biologically to language.

If sequential learning is a fundamental human skill (Lashley, 1951), as assumed in the above simulations, it should be possible to uncover the source of some of the universal constraints on language by studying human performance on sequential learning tasks. In a series of sequential learning experiments (Christiansen, 2000; Christiansen & Reeder, 2006), participants were trained on sequences generated by either a regular or irregular grammar from Christiansen and Devlin (1997). When tested on novel sequences, the participants trained on the regular grammar were significantly better at distinguishing grammatical from ungrammatical items compared to participants trained on the irregular grammar. Identical results were obtained with sequences of nonwords, letters, and visual nonsense shapes, indicating that the results were due to non-linguistic sequential learning constraints rather than the influence of a putative UG. Together, the simulations and experimental results suggest that sequential learning constraints may provide an alternative explanation of head-order consistency to X-bar theory. Specifically, constraints on basic word order may derive from non-linguistic constraints on the learning and processing of complex sequential structure. Grammatical constructions with highly irregular head-ordering may simply be too hard to learn and would therefore tend to disappear.

3.2.3 Subjacency

Given the assumption that what is acquired during language acquisition is the ability to process language, our approach converges with work in functional linguistics on how processing

---

[10] Conway and Christiansen (2001) reviewed evidence on sequential learning abilities in non-human primates and concluded that although the performance of non-human primates on learning fixed sequences and certain types of statistical structure is similar to that of humans, the former has problems dealing with the kind of hierarchical sequential structure characteristic of human languages (see also Hauser, Chomsky & Fitch, 2002, for a similar suggestion). This difference may also help explain why only humans have complex linguistic abilities.

constraints may shape word-order universals (e.g., Hawkins, 1994, 2005), which Kirby (1999) explored in an evolutionary context. Another linguistic universal that has similarly been suggested to derive from processing constraints is the subjacency principle (e.g., Berwick & Weinberg, 1984; O'Grady, 2005). According to Pinker and Bloom (1990), subjacency is a classic example of an arbitrary linguistic constraint that makes sense only from a linguistic perspective and as such is considered to have evolved as part of UG. Informally, subjacency provides constraints on long-distance dependencies in complex question formation within generative grammar. Consider the two questions in (7) and (8), which can be derived from sentence (6). The question in (7) is ungrammatical because subjacency blocks *what* from referring to the thing that everyone likes (i.e., *dogs*), whereas (8) is grammatical because *who* refers to the person(s) Betty asked *'why everyone likes dogs'* and which does not violate subjacency.

(6) Betty asked why everyone likes dogs.
(7) *What did Betty ask why everyone likes?
(8) Who did Betty ask why everyone likes dogs?

Ungrammaticality due to subjacency violations also shows up, on this account, for questions such as (10) derived from (9). Here the subjacency principle suggests that the noun phrase '*the news*' blocks *what* from referring to dogs, as indicated by the grammatical question in (11).

(9) Betty heard the news that everybody likes dogs.
(10) *What did Betty hear the news that everybody likes?
(11) What did Betty hear that everybody likes?

Ellefson and Christiansen (2000) explored an alternative explanation which suggests 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. 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 (similar to examples 8 and 11), whereas the two complex questions in the 'unnatural language' violated the subjacency constraints (similar to examples 7 and 10).

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 trained in the natural condition were marginally better than the subjects trained in the unnatural condition at classifying the grammaticality of 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.

To further support the suggestion that the difference in learnability between the two languages is brought about by sequential learning constraints and not pre-existing language abilities, Ellefson and Christiansen conducted a set of SRN simulations. 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 behavioral results, corroborating the suggestion that constraints on the learning and processing of sequential structure can 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. When language itself is viewed as a "organic" system sensitive to adaptive pressures, natural selection will favor combinations of linguistic constructions that can be acquired relatively easily given existing learning and processing mechanisms. Consequently, difficult to learn language fragments such as the unnatural language will tend to disappear. In conclusion, rather than having an innate UG principle to rule out subjacency violations, we suggest that they may have been eliminated altogether through an evolutionary process of linguistic adaptation constrained by prior cognitive limitations on sequential learning and processing.

Of course, the situation surrounding subjacency is more complex than captured by Ellefson and Christiansen's results. However, this approach dovetails nicely with work in construction-based grammars, suggesting that subjacency violations tend to be much more graded than generally assumed by generative linguistics and are subject to a number of referential, pragmatic, and processing influences (e.g., Goldberg, 2006; O'Grady, 2005). We see this kind of converging perspective as providing a possible route to a more complete picture of language universals, such as subjacency and basic word order, while, as we see next, also dovetailing with ongoing work in grammaticalization seeking to explain diachronic change.

3.3. Language change and evolution through grammaticalization

According to the view that language evolution is determined by the development of a genetic basis for language, there is a sharp divide between questions of language evolution (how the genetic endowment is shaped by natural selection), and historical language change (which is viewed as variation within the genetically determined limits of possible human languages). By contrast, if language has evolved to fit prior cognitive and communicative constraints, then it may be that historical processes of language change provide a model of language evolution; indeed, historical language change may be language evolution in microcosm.

Recent theory in diachronic linguistics has focused on grammaticalization (e.g., Bybee, Perkins & Pagliuca, 1994; Heine, 1991; Hopper & Traugott, 1993): the process by which functional items, including closed class words and morphology, develop from what are initially open-class items. This transitional process involves a 'bleaching' of meaning; phonological reduction; and increasingly rigid dependencies with other items. Thus, the English number *one* is likely to be the root to *a(n)*. The Latin *cantare habeo* ('I have (something) to sing') mutated into *chanterais*, *cantaré, cantarò* ('I will sing' in French, Spanish, Italian). The suffix corresponds phonologically to 'I have' in each language (respectively, *ai*, *he*, *ho*—the 'have' element has collapsed into inflectional morphology, Fleischman, 1982). The same processes of grammaticalization can also cause certain content words over time to get bleached of their meaning and become grammatical particles. For example, the use of *go* and *have* as auxiliary

verbs (as in *I am going to sing* or *I have forgotten my hat*) have been bleached of their original meanings concerning physical movement and possession. The processes of grammaticalization appear gradual, and to follow historical patterns which suggest that there are systematic selectional pressures in operation in language change. More generally, these processes provide a possible origin of grammatical structure from a proto-language initially involving perhaps unordered and uninflected strings of content words.

This perspective contrasts sharply with the view of language typical in the generative grammar framework. From a historical perspective, for example, it is natural to view many aspects of syntax as emerging from processing or pragmatic factors. Thus, revisiting our discussion of binding constraints (Section 3.1), we might view complementary distributions of pronouns as initially arising from pragmatic factors; and the resulting pattern may be acquired and modified by future generations of learners, to some degree independently of those initial factors (e.g., Levinson, 1987b; Givón, 1979). Thus, binding constraints might be a complex product of many forces, including pragmatic factors and learning and processing biases—and hence the subtlety of those constraints should not be entirely surprising. But from the present perspective, the fact that such a complex system of constraints is readily learnable, is neither puzzling, nor indicative of an innately specified genetic endowment; rather the constraints are learnable because they have been shaped by the very pragmatic, processing and learning constraints with which the learner is endowed.

Understanding the cognitive and communicative basis for the direction of grammaticalization and related processes is an important challenge. But equally, the suggestion that this type of observable historical change may be continuous with language evolution opens up the possibility that research on the origin of language may not be a theoretically isolated island of speculation, but may connect directly with some of the most central topics in linguistics: the nature of language change (e.g., Zeevat, 2006). Indeed, grammaticalization has become the center of many recent perspectives on the evolution of language as mediated by cultural transmission across hundreds (perhaps thousands) of generations of learners (e.g., Bybee et al., 1994; Givón, 1998; Heine & Kuteva, 2002; Schoeneman, 1999; Tomasello, 2003). Although our approach also emphasizes the importance of grammaticalization in the evolution of complex syntax, it differs from other approaches in that we see this diachronic process as being constrained by limitations on learning and processing. Indeed, there have even been intriguing attempts to explain some aspects of language change with reference to the learning properties of connectionist networks, of the kind discussed above. Most notably, Hare & Elman (1995) demonstrated how cross-generational learning in neural networks can model the gradual historical change in English verb inflection from a complex past tense system in Old English to the dominant "regular" class and small classes of "irregular" verbs of modern English.

More generally, we are skeptical that language evolution can be explained entirely as a function of cultural transmission without the need for innate constraints on learning (as e.g., proposed by Arbib, 2005; Bybee, 2002; Donald, 1998). Without constraints on cultural transmission we would expect to find few commonalities among languages. Yet, the languages of the world—despite their many differences—also share many systematic similarities in their structure and usage. Although the space of logically possible ways in which languages could be structured and used is vast, the world's languages only occupy a small fraction of this space. For example, of the world's languages more than 50% have an SOV word order whereas less than 0.25% have an OSV word order (Dryer, 1992). If the processes of language emergence are focused within the cultural domain then linguistic universals should be unlikely because it is

possible to imagine a multitude of culturally useful, and equally adaptive, constraints on linguistic form. That is, cultural transmission on its own cannot explain the existence of universal linguistic patterns; rather, innate constraints are needed to explain why language is structured the way it is, and why language is so readily learnt. However, as we have noted, these constraints do not necessarily have to be linguistic in nature. Rather, innate cognitive and pragmatic constraints on learning and processing, existing prior to the emergence of language, provided a niche within which cultural transmission could take place. These constraints subsequently became "fossilized" in the structure of language because linguistic forms that fit these constraints were more readily learned, and hence propagated more effectively from speaker to speaker.

Human language has thus been shaped by selectional pressure from thousands of generations of language users. Linguistic variants which children find easier to learn to understand and produce; variants which are more economical, expressive and generally effective in communication, persuasion, and perhaps signally of status and social group, will be favored. Just as with the multiple selectional pressures operative in biological evolution, the matrix of factors at work in driving the evolution of language is complex. Nonetheless, as we have seen, candidate pressures can be proposed (e.g., the pressure for incrementality, minimizing memory load, regularity, brevity, and so on); and regular patterns of language change that may be responses to those pressures can be identified (e.g., the processes of successive entrenchment, generalization and erosion of structure evident in grammaticalization). We have argued that these processes have shaped language to fit human cognition. Thus, the logical problem of language evolution that appears to confront attempts to explain how a genetically specified linguistic endowment could become encoded, does not arise; it is not cognition that has somehow evolved to language, but the reverse.

4. Concluding Remarks

In this paper, we have presented a theory of language evolution in which language is construed as an organism, which has adapted to be symbiotic with human cognitive processes. From this perspective, the close 'fit' between language learners and the structure of natural language that motivates many theorists to posit a language-specific biological endowment may instead arise from processes of adaptation operating on language itself. Moreover, we have argued that there are fundamental difficulties with postulating a language-specific biological endowment. It is implausible that such an endowment could evolve through adaptation (because the prior linguistic environments would be too diverse to give rise to universal principles). It is also unlikely that a language-specific endowment arose through non-adaptational genetic mechanisms, because the probability of a functional language system arising essentially by chance is vanishingly small. Instead, we have suggested that some apparently arbitrary aspects of language structure may arise from the properties of human sequential learning mechanisms, and have illustrated the point by showing how analogs of important, and apparently arbitrary, linguistic regularities (e.g., binding, head order, and subjacency) can emerge from learning and processing constraints.

4.1 The logical problem of language evolution meets the logical problem of language acquisition

The present viewpoint has interesting theoretical implications concerning language acquisition. Children acquire the full complexity of natural language over a relatively short amount of time,

from exposure to noisy and partial samples of language. The ability to develop complex linguistic abilities from what appears to be such poor input data has led many to speak of the "logical" problem of language acquisition (e.g., Baker & McCarthy, 1981; Hornstein & Lightfoot, 1981). One solution to the problem is to assume that learners have some sort of biological 'head-start' in language acquisition—that their learning apparatus is precisely meshed with the structure of natural language. This viewpoint is, of course, consistent with theories according to which there is a genetically specified language organ, module or instinct (e.g., Chomsky, 1986, 1993; Crain, 1991; Piattelli-Palmarini, 1989, 1994; Pinker, 1994; Pinker & Bloom, 1990). But it is also consistent with the present view that languages have evolved to be learnable. According to this view, the mesh between language learning and language structure has occurred not because specialized biological machinery embodies the principles that govern natural languages (UG), but rather that the structure of language has evolved to fit with pre-linguistic learning and processing principles.

If language has evolved to be learnable, then the problem of language acquisition may have been mis-analyzed. Language acquisition is frequently viewed as a standard problem of induction (e.g., Gold, 1967; Osherson, Stob & Weinstein, 1986; Jain, Osherson, Royer & Sharma, 1999; Pinker, 1984, 1989), where there is a vast space of possible grammars that are consistent with the linguistic data to which the child is exposed. Accordingly, it is often readily concluded that the child must have innate knowledge of language structure to constrain the space of possible grammars to a manageable size. But, if language is viewed as an organism adapted to the learner, then language learning is by no means a standard problem of induction. To give an analogy, according to the standard view of induction, the problem of language acquisition is like being in an unreasonable quiz show, where you have inadequate information, but must somehow guess the "correct" answer. But according to the present view, by contrast, there is no externally given "correct" answer; instead, the task is simply to give the same answer as everybody else— because the structure of language will have adapted to conform to this most 'popular' guess. This is a much easier problem—whatever learning biases people have, so long as these biases are *shared* across individuals, learning should proceed successfully. Moreover, the viewpoint that children learn language using general-purpose cognitive mechanisms, rather than language-specific mechanisms, has also been advocated on independent grounds (e.g., Arbib, 2005; Davidson, 2003; Deacon, 1997; Donald, 1998; Elman et al., 1996; MacWhinney, 1999; Ragir, 2002; Schoenemann, 1999; Seidenberg, 1997; Seidenberg & MacDonald, 2001; Tomasello, 2000a, 2000b, 2000c; 2003).

From this perspective, the problem of language acquisition is very different from learning, say, some aspect of the physical world. In learning naïve physics, the constraints to be learned (e.g., how rigid bodies move, how fluids flow, and so on) are defined by processes outside the cognitive system. External processes define the 'right' answers, to which the learner must converge. But in language acquisition, the structure of the language to be learned is itself determined by the learning of generations of previous learners. Because learners have similar learning biases, this means that the first wild guesses that the learner makes about how some linguistic structure works are likely to be the right guesses. More generally, in language acquisition, whatever biases the learner has, so long as they are shared by other learners, are likely to be the helpful biases in acquiring the language—because the language has been shaped by processes of selection to conform with those biases to the greatest extent possible. This also means that the problem of the poverty of the stimulus (e.g., Chomsky, 1980; Crain, 1991; Crain & Pietroski, 2001) does not arise, because language has been shaped to be learnable from the

kind of noisy and partial input available to young children. Thus, language acquisition is constrained by substantial biological constraints—but these constraints emerge from cognitive machinery that is not language-specific.

This alternative characterization of language acquisition additionally offers a different perspective on linguistic phenomena that has typically been seen as requiring a UG account for their explanation, such as specific language impairment (SLI) and creolization. These phenomena are beyond the scope of this paper, so we can only sketch how they may be approached. For example, the acquisition problems in SLI may, on our account, be largely due to deficits in underlying sequential learning mechanisms that support language (see Ullman & Pierpont, 2005, for a similar perspective), rather than impaired language-specific modules (e.g., Gopnik & Crago, 1991; Pinker, 1994; Van der Lely & Battell, 2003). Consistent with this perspective, recent studies have shown that children and adults with SLI have impaired sequential learning abilities (e.g., Hsu et al, 2006; Evans & Saffran, 2005). Although processes of creolization, in which children acquire consistent linguistic structure from noisy and inconsistent input, have been seen as evidence of UG (e.g., Bickerton, 1984), we suggest that creolization may be better construed as arising from cognitive constraints on learning and processing. The rapid emergence of a consistent SOV word order in the Al-Sayyid Bedouin Sign Language (Sandler, Meir, Padden & Aronoff, 2004) is consistent with this suggestion and the earlier mentioned word order simulation results (Section 3.2.2). Additional research is required to flesh these accounts out in detail as well as to address the full force of the poverty of stimulus problem, but a growing bulk of work indicates that such accounts are indeed possible (e.g., Chater & Vitányi, in press; Goldberg, 2006; Hudson Kam & Newport, 2005; O'Grady, 2005; Reali & Christiansen, 2005; Tomasello, 2003).

4.2 The scope of the argument

We have argued that the standard conception of the problem of language evolution—as explaining the origin of language-specific genes, coding for arbitrary, though universal, properties of natural language—is based on a false premise. Language should be viewed as shaped by the brain; the brain has not adapted to encode arbitrary universal principles of language. The pre-linguistic learning and processing biases, however capricious, will be reflected in language—because language has evolved to be learned and processed by people whose brains embody these constraints. But, intriguingly, it seems that many apparently arbitrary aspects of language can be explained by relatively natural cognitive constraints—and hence that language may be rather less arbitrary than at first supposed.

How broad is our argument? It is important to emphasize what the arguments are not intended to show. In particular, we are not suggesting that biological adaptation is not relevant for language. Indeed, it seems likely that a number of preadaptations for language might have occurred (see Hurford, 2003, for a review), such as the ability to represent discrete symbols (Deacon, 1997; Tomasello, 2003), to reason about other minds (Malle, 2002), to understand and share intentions (Tomasello, 2003; Tomasello, Carpenter, Call, Behre & Moll, 2005), to perform pragmatic reasoning (Levinson, 2000), and the emergence of an exceptionally prolonged childhood (Locke & Bogin, 2006). Similarly, biological adaptations might have led to improvements to the cognitive systems that support language, including increased working memory capacity (Gruber, 2002), domain-general capacities for word learning (Bloom, 2001), and complex hierarchical sequential learning abilities (Calvin, 1994; Conway & Christiansen, 2001; Greenfield, 1991; Hauser et al., 2002), though these adaptations are likely to have been for

improved cognitive skills rather than for language. Nonetheless, some language-specific adaptations may have occurred as well, but given our arguments above these would only be for functional features of language, and not the arbitrary features of UG. For example, changes to the human vocal tract may have resulted in more intelligible speech (Lieberman, 1984, 1991, 2003—though see also Hauser & Fitch, 2003).

Our argument may, though, have applications beyond language. The arguments we have outlined against adaptationism in language evolution appear to apply equally to rule out putative co-evolution of the brain with any rapidly changing and highly varied aspect of human culture—from marriage practices and food sharing practices, to music and art, to folk theories of religion, science or mathematics. We speculate that, in each case, the apparent fit between culture and the brain arises primarily because culture has been shaped to fit with our prior cognitive biases. Thus, by analogy with language, we suggest that nativist arguments across these domains might usefully be re-evaluated, from the perspective that culture may have adapted to cognition much more substantially than cognition has adapted to culture.

In summary, we have argued that the notion of UG is subject to a logical problem of language evolution, whether it is suggested to be the result of gradual biological adaptation or other nonadaptationist factors. Instead, we have proposed to explain the close fit between language and learners in terms of linguistic adaptation of language to learners, which, in turn, offers a reappraisal of the logical problem of language acquisition.

Acknowledgments

References

Abrams, M. & Reber, A.S. (1988). Implicit learning: Robustness in the face of psychiatric disorders. *Journal of Psycholinguistic Research*, *17*, 425-439.

Alter, S. (1998). *Darwinism and the linguistic image: Language, race, and natural theology in the nineteenth century*. Baltimore, MD: Johns Hopkins University Press.

Arbib, M.A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral & Brain Sciences, 28*, 105-124.

Baker, C. L. & McCarthy, J. J. (Eds.) (1981). *The logical problem of language acquisition*. Cambridge, MA: MIT Press.

Baker, M.C. (2003). Language differences and language design. *Trends in Cognitive Sciences, 7*, 349-353.

Beer, G. (1996). Darwin and the growth of language theory. In *Open fields: Science in cultural encounter* (pp. 95-114). Oxford: Oxford University Press.

Berwick, R.C & Weinberg, A.S (1984). *The grammatical basis of linguistic performance: language use and acquisition*. Cambridge, MA: MIT Press.

Bickerton, D. (1984). The language bio-program hypothesis. *Behavioral and Brain Sciences*, *7*, 173-212.

Bickerton, D. (1995). *Language and human behavior*. Seattle, WA: University of Washington Press.

Bickerton, D. (2003) Symbol and structure: a comprehensive framework for language evolution. In M.H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 77-93). Oxford: Oxford University Press.

Black, Cheryl A. (1999). *A step-by-step introduction to the Government and Binding theory of syntax*. http://www.sil.org/americas/mexico/ling/E002-IntroGB.pdf, February 1999.

Bloom, P. (2001). Précis of *How children learn the meanings of words. Behavioral and Brain Sciences*, *24*,1095-1103.

Briscoe, E.J. (2003) Grammatical assimilation. In M.H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 295-316). Oxford: Oxford University Press.

Brooks, P.J., Braine, M.D.S., Catalano, L., Brody, R.E., & Sudhalter, V. (1993). Acquisition of gender-like noun subclasses in an artificial language: The contribution of phonological markers to learning. *Journal of memory and language*, *32*, 76-95.

Bybee, J.L. (2002). Sequentiality as the basis of constituent structure. In T. Givón, & B. Malle (Eds.), The evolution of language out of pre-language (pp. 107-132). Philadelphia, PA: John Benjamins.

Bybee, J.L., Perkins, R.D. & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect and modality in the languages of the world.* Chicago: University of Chicago Press.

Calvin, W.H. (1994). The emergence of intelligence. *Scientific American*, *271*, 100-107.

Cannon, G. (1991). Jones's "Spring from some common source": 1786–1986. In S. M. Lamb and E. D. Mitchell (eds.), *Sprung from some common source: Investigations into the pre-history of languages*. Stanford, CA: Stanford University Press.

Cavalli-Sforza, L.L. & Feldman, M.W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nature Genetics, 33,* 266-275.

Chater, N. & Vitányi, P. (in press). 'Ideal learning' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology*.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass: MIT Press.

Chomsky, N. (1972). *Language and mind*. Harcourt, Brace and World (extended edition).

Chomsky, N. (1980). *Rules and representations* New York: Columbia University Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Forris Publications.

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

Chomsky, N. (1988). *Language and the problems of knowledge. The Managua Lectures*. Cambridge, Mass: MIT Press.

Chomsky, N. (1993). *Language and thought*. Wakefield, RI: Moyer Bell.

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

Christiansen, M.H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word universals. J. L. Dessalles & L. Ghadakpour (Eds.), *The Evolution of Language: 3rd International Conference* (pp. 45-48). Paris, France: Ecole Nationale Supérieure des Télécommunications.

Christiansen, M.H. & Devlin, J.T. (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* (pp. 113-118). Mahwah, NJ: Lawrence Erlbaum.

Christiansen, M.H., Kelly, L., Shillcock, R. & Greenfield, K. (2006). *Impaired artificial grammar learning in agrammatism*. Submitted manuscript.

Christiansen, M.H., Reali, F. & Chater, N. (2006). The Baldwin effect works for functional, but not arbitrary, features of language. In A. Cangelosi, A. Smith & K. Smith (Eds.), *Proceedings of the Sixth International Conference on the Evolution of Language* (pp. 27-34). London: World Scientific Publishing.

Christiansen, M.H. & Reeder, P.A. (2006). *Cognitive constraints on word order universals: Evidence from connectionist modeling and artificial grammar learning.* Manuscript in preparation.

Clark, H.H. (1975). Bridging. In R. C. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing*. New York: Association for Computing Machinery.

Cleeremans, A., Destrebecqz, A., & Boyer, A. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences, 2*, 406-416.

Conway, C.M., & Christiansen, M.H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, 5, 539-546.

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

Corballis, M.C. (2003). From hand to mouth: The gestural origins of language. In M.H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 201-218). Oxford: Oxford University Press.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences, 14*, 597-650.

Crain, S., & Pietroski, P. (2001). Nature, nurture and universal grammar. *Linguistics and Philosophy, 24*, 139–186.

Culicover, P.W. (1999). *Syntactic nuts*. Oxford: Oxford University Press.

Curtin, S., Mintz, T.H. & Christiansen, M.H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition, 96,* 233-262.

Darwin, C. (1900). *The descent of man, and selection in relation to sex* (2nd Edition). New York: P.F. Collier and Son.

Davidson, I. (2003). The archaeological evidence of language origins: States of art. In M.H. Christiansen, & S. Kirby (Eds.), *Language evolution* (pp. 140-157). New York: Oxford University Press.

Davies, A. M. (1987). 'Organic" and "Organism" in Franz Bopp. In H. M. Hoenigswald and L. F. Wiener (Eds.), *Biological metaphor and cladistic classification* (pp. 81–107). Philadelphia, PA: University of Pennsylvania Press).

Dawkins, R. (1976). *The selfish gene*. New York: Oxford University Press.

Dawkins, R. (1986). *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. Harmondsworth, UK: Penguin.

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

Diamond, J. (1992). *The third chimpanzee: The evolution and future of the human animal*. New York: Harper Collins.

Donald, M. (1998). Mimesis and the executive suite: Missing links in language evolution. In J.R. Hurford, M. Studdert-Kennedy and C. Knight (Eds.), *Approaches to the evolution of language* (pp. 44-67). Cambridge, U.K.: Cambridge University Press.

Dryer, M. S. (1992). The Greenbergian word order correlations, *Language, 68*, 81–138.

Dunbar, R.I.M. (2003). The origin and subsequent evolution of language. In M.H. Christiansen, & S. Kirby (Eds.), *Language evolution* (pp. 219-234). New York: Oxford University Press.

Ellefson, M.R. & 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* (pp. 645-650). Mahwah, NJ: Lawrence Erlbaum.

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

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Evans, J., & Saffran, J. R. (2005). *Statistical learning in children with Specific Language Impairment*. Paper presented at the Boston University Conference on Language Development, Boston, MA, November, 2005.

Fleischman, S. (1982). *The future in thought and language: Diachronic evidence from Romance*. Cambridge, U.K.: Cambridge University Press.

Frean, M.R. & Abraham, E.R. (2004). Adaptation and enslavement in endosymbiont-host associations. *Physical Review E, 69,* 051913.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 529-534.

Frigo, L. & McDonald, J.L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39,* 218-245.

Givón, T. (1979) *On understanding grammar*. New York: Academic Press.

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

Givón, T. & Malle, B. F. (Eds.) (2002). *The evolution of language out of pre-language.* Amsterdam: Benjamins.

Gold, E. (1967). Language identification in the limit. *Information and Control, 16*, 447-474.

Goldberg, A.E. (2006). *Constructions at work: The nature of generalization in language*. New York: Oxford University Press.

Gómez, R.L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431-436.

Gómez, R.L., & Gerken, L.A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*, 178-186.

Gopnik, M. & Crago, M.B. (1991). Familial aggregation of a developmental language disorder. *Cognition*, *39*, 1-50.

Gould, S.J. (1993). *Eight little piggies: Reflections in natural history*. New York: Norton.

Gould, S. J. (2002). *The structure of evolutionary theory*. Cambridge, MA: Harvard University Press.

Gould, S. J. & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London (Series B)*, *205*, 581-598.

Gould, S.J. & Vrba, E.S. (1982). Exaptation - a missing term in the science of form. *Paleobiology*, *8*, 4-15.

Gray, R. D. & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, *426*, 435-439.

Green, T. R. G. (1979). Necessity of syntax markers: 2 Experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior, 18,* 481-496.

Greenfield, P.M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, *14*, 531-595.

Gruber, O. (2002). The co-evolution of language and working memory capacity in the human brain. In M.I. Stamenov & V. Gallese (Eds.), *Mirror neurons and the evolution of brain and language* (pp. 77–86). Amsterdam: John Benjamins.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, 7, 1-52.

Hare, M. & Elman, J.L. (1995). Learning and morphological change. *Cognition, 56*, 61-98

Hauser, M.D., Chomsky, N. & Fitch, W.T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569-1579.

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. 158-181). Oxford: Oxford University Press.

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

Hawkins, J.A. (2005). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hecht Orzak, S. & Sober, E. (Eds.) (2001). *Adaptationism and optimality*. Cambridge: Cambridge University Press.

Heine, B. (1991). *Grammaticalization*. Chicago: University of Chicago Press.

Heine, B. & Kuteva, T. (2002). On the evolution of grammatical forms. In A. Wray (ed.), *Transitions to language* (pp. 376-397). Oxford, U.K.: Oxford University Press.

Hinton, G.E. & Nowlan, S.J. (1987). How learning can guide evolution. *Complex Systems*, *1*, 495-502.

Hoen, M., Golembiowski, M., Guyot, E., Deprez, V., Caplan, D. & Dominey, P.F. (2003). Training with cognitive sequences improves syntactic comprehension in agrammatic aphasics. *NeuroReport, 14,* 495-499.

Hopper, P. & Traugott, E. (1993). *Grammaticalization*. Cambridge, UK: Cambridge University Press.

Hornstein, N. & Lightfoot, D. (Eds.) (1981). *Explanations in linguistics: The logical problem of language acquisition*. London: Longman.

Hsu, H.-J., Christiansen, M.H., Tomblin, J.B., Zhang, X. & Gómez, R.L. (2006). *Statistical learning of nonadjacent dependencies in adolescents with and without language impairment*. Poster presented at the 2006 Symposium on Research in Child Language Disorders, Madison, WI.

Hudson Kam, C.L. & Newport, E.L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151-195.

Hurford, J.R. (1991). The evolution of the critical period for language learning. *Cognition*, *40*, 159-201.

Hurford, J.R. (2003). The language mosaic and its evolution. In M.H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 38-57). Oxford: Oxford University Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. New York: Oxford University Press.

Jain, S., Osherson, D., Royer, J., & Sharma, A. (1999). *Systems that learn* (2nd ed.). Cambridge, MA: M.I.T. Press.

Jenkins, L. (2000). *Biolinguistics: Exploring the biology of language.* Cambridge, Cambridge University Press.

Kauffman, S. A. (1995). *The origins of order: Self-organization and selection in evolution.* Oxford: Oxford University Press.

Kirby, S. (1999). *Function, selection and innateness: The emergence of language universals.* Oxford: Oxford University Press.

Knowlton, B.J. & Squire, L.R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 169-181.

Lanyon, S.J. (2006). A saltationist approach for the evolution of human cognition and language. In A. Cangelosi, A.D.M. Smith, & K. Smith (Eds.), *The Evolution of Language* (pp. 176-183). Singapore: World Scientific.

Lashley, K.S. (1951). The problem of serial order in behavior. In L.A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112-146). New York: Wiley.

Levinson, S.C (1987a). Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics*, *23*, 379-434.

Levinson S.C. (1987b). Minimization and conversational inference. In M. Papi and J. Verschueren (eds.), *The Pragmatic Perspective: Proceedings of the International Conference on Pragmatics at Viareggio* (pp. 61-129). Amsterdam: J. Benjamins.

Levinson, S.C (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Lewontin, R.C. (1998). The evolution of cognition: Questions we will never answer. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science, Volume 4: Methods, models, and conceptual issues.* Cambridge, MA: MIT Press.

Li, M. & Vitányi, P. (1997). *An introduction to Kolmogorov complexity theory and its applications* (2nd ed). Berlin: Springer.

Lieberman, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard University Press.

Lieberman, P. (1991). Speech and brain evolution. *Behavioral and Brain Science*, *14*, 566-568.

Lieberman, P. (2003). Motor control, speech, and the evolution of human language. In M.H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 255-271). New York: Oxford University Press.

Locke, J.L. & Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral & Brain Sciences, 29*, 259-280.

Lupyan, G. & Christiansen, M.H. (2002). Case, word order, and language learnability: Insights from connectionist modeling. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 596-601). Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. (Ed.) (1999). *The emergence of language*. Mahwah, NJ: Erlbaum.

Maess, B., Koelsch, S., Gunter, T.C. & Friederici A.D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, *4*, 540–545.

Malle, B.F. (2002). The relation between language and theory of mind in development and evolution. In T. Givón, & B. Malle (Eds.), *The evolution of language out of pre-language* (pp. 265-284). Philadelphia, PA: John Benjamins.

Maynard-Smith, J. (1978). Optimization theory in evolution. *Annual Review of Ecology and Systematics*, *9*, 31-56.

McDowall, J. & Martin, S. (1996). Implicit learning in closed head injured subjects: Evidence from an event sequence learning task. *New Zealand Journal of Psychology, 25,* 2-6.

McMahon A.M.S. (1994). *Understanding language change*. Cambridge: Cambridge University Press.

Meulemans, T., Peigneux, P. & Van der Linden, M. (1998). Preserved artificial grammar learning in Parkinson's disease. *Brain and Cognition, 37,* 109-112.

Monaghan, P., Chater, N. & Christiansen, M.H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition, 96,* 143-182.

Morgan, J.L., Meier, R.P., & Newport, E.L. (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.

Nettle, D. & Dunbar, R.I.M. (1997). Social markers and the evolution of reciprocal exchange. *Current Anthropology, 38,* 93-99.

Newport, E.L. & Aslin, R.N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127-162.

Nowak, M.A., Komarova, N.L. & Niyogi, P., 2001. Evolution of universal grammar. *Science, 291,* 114-118.

O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Mahwah, NJ: Erlbaum.

Onnis, L., Christiansen, M.H., Chater, N. & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial grammar learning. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 886-891). Mahwah, NJ: Lawrence Erlbaum.

Onnis, L., Monaghan, P., Chater, N. & Richmond, K. (2005). Phonology impacts segmentation in speech processing. *Journal of Memory and Language, 53,* 225-237.

Osherson, D., Stob, M. and Weinstein, S. (1986). *Systems that learn*. Cambridge, Mass: MIT Press.

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

Peña, M., Bonnatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298,* 604-607.

Pennisi, E. (2004). The first language? *Science, 303,* 1319-1320.

Pennock, R.T (1999). *Tower of Babel: The evidence against the new creationism*. Cambridge, Mass: MIT Press.

Percival, W.K. (1987). Biological analogy in the study of languages before the advent of comparative grammar. In H.M. Hoenigswald & L.F. Wiener (Eds.), *Biological metaphor and cladistic classification* (pp. 3–38). Philadelphia, PA: University of Pennsylvania Press.

Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science, 28,* 383-407.

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

Piattelli-Palmarini, M. (1994). Ever since language and learning: Afterthoughts on the Piaget-Chomsky debate. *Cognition, 50,* 315-346.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, Mass: Harvard University Press.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, Mass: MIT Press.

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

Pinker, S., (2003). Language as an adaptation to the cognitive niche. In M. H. Christiansen and S. Kirby (Eds.), *Language evolution* (pp. 16-37). Oxford: Oxford University Press.

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

Plante, E., Gómez, R.L., & Gerken, L.A. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders, 35*, 453-462.

Pomerantz, J. R. & Kubovy. M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.) *Handbook of Perception and Human Performance. Volume 2: Cognitive Processes and Performance*. (pp. 36-1-36-46) New York: Wiley.

Raddick, G. (2000). Review of S. Alter's *Darwinism and the Linguistic Image. British Journal for the History of Science, 33,* 122–124.

Raddick, G. (2002). Darwin on language and selection. *Selection, 3*, 7–16.

Ragir, S. (2002). Constraints on communities with indigenous sign languages: Clues to the dynamics of language origins. In A. Wray (Ed.), *Transitions to language* (pp. 272-294). Oxford: Oxford University Press.

Reali, F. & Christiansen, M.H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science, 29,* 1007-1028.

Reali, F. & Christiansen, M.H. (2006). *The role of biological and linguistic adaptation in language evolution.* Manuscript under revision.

Reber, A.S. (1992). An evolutionary context for the cognitive unconscious. *Philosophical Psychology, 5*, 33-51.

Reber, P.J., Martinez, L.A. & Weintraub, S. (2003). Artificial grammar learning in Alzheimer's disease. *Cognitive, Affective and Behavioral Neuroscience, 3*, 145-153.

Reinhart, T. (1983). *Anaphora and semantic interpretation*. Chicago: Chicago University Press.

Rossel, S., Corlija, J., & Schuster, S. (2002). Predicting three-dimensional target motion: How archer fish determine where to catch their dislodged prey. *Journal of Experimental Biology, 205*, 3321-3326.

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

Saffran J.R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172-196.

Saffran, J.R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12,* 110-114.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996a). Statistical learning by 8-month-old infants. *Science, 274*, 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606-621.

Sandler, W., Meir, I., Padden, C. & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences, 102*, 2661-2665.

Schoenemann, P.T. (1999). Syntax as an emergent characteristic of the evolution of semantic complexity. *Minds and Machines, 9*, 309-346.

Seidenberg, M.S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599-1604.

Seidenberg, M.S. & MacDonald, M. (2001). Constraint-satisfaction in language acquisition. In M.H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics* (pp. 281-318). Westport, CT: Ablex.

Sereno, M.I. (1991). Four analogies between biological and cultural/linguistic evolution. *Journal of Theoretical Biology, 151*, 467-507.

Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics as neural representation. *Annual Review of Neuroscience, 24*, 1193-1215.

Slobin, D.I., & Bever, T.G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition, 12*, 229-265.

Stevick R.D. (1963). The biological model and historical linguistics. *Language, 39*, 159–169.

Suzuki, D.T., Griffiths, A.J.F., Miller, J.H. & Lewontin, R.C. (1989). *An introduction to genetic analysis* (4th edition). New York, NY: W. H. Freeman.

Tomasello, M., (2000a). Do you children have adult syntactic competence? *Cognition, 74*, 209-253.

Tomasello, M., (2000b). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4*, 156-163.

Tomasello, M., (Ed). (2000c). *The new psychology of language: Cognitive and functional approaches*. Hillsdale, NJ: Erlbaum.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral & Brain Sciences, 28*, 675-691.

Ullman, M.T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231-270.

Ullman, M. T. & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex, 41,* 399-433.

Van der Lely, H. K. J. & Battell, J. (2003). WH-movement in children with grammatical SLI: A test of the RDDR hypothesis. *Language, 79,* 153-181.

van Driem, G. (2005). The language organism: The Leiden theory of language evolution. In J.W. Minett & W.S.-Y. Wang (Eds.), *Language acquisition, change and emergence: Essays in evolutionary linguistics* (pp. 331-340). Hong Kong: City University of Hong Kong Press.

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

Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *PloS Biology, 4*, e72.

von Humboldt, W. (1999). *On language: On the diversity of human language construction and its influence on the metal development of the human species.* Cambridge, U.K.: Cambridge University Press.

Wilkins, W.K. & Wakefield, J. (1995). Brain evolution and neurolinguistic preconditions. *Behavioral & Brain Sciences, 18*, 161-182.

Zeevat, H. (2006). Grammaticalisation and evolution. In A. Cangelosi, A. D. M. Smith, & K. Smith (Eds.) *The Evolution of Language* (pp. 372-378). Singapore: World Scientific.