6 The Role of Learning and Development in Language Evolution: A Connectionist Perspective

Morten H. Christiansen and Rick Dale

Introduction

Much ink has been spilled arguing over the idea that ontogeny recapitulates phylogeny. The discussions typically center on whether developmental stages reflect different points in the evolution of some specific trait, mechanism, or morphological structure. For example, the development trend from crawling to walking in human infants can be seen as recapitulating the evolutionary change from quadrupedalism to bipedalism in the hominid lineage. Closer to the area of the evolution of communication, casts have been taken to indicate that the vocal tract of newborn human infants more closely resembles those of australopithecines and extant primates than the adult human vocal tract—with the vocal tract of Neanderthals falling in between, roughly corresponding to that of a two-year-old human child (Lieberman, 1998). These data could suggest that the development of the vocal tract in human ontogeny is recapitulating the evolution of the vocal tract in hominid phylogeny. However, other researchers have strongly opposed such a perspective, arguing that evolution and development work along entirely different lines when it comes to language (Pinker and Bloom, 1990). In this chapter, we provide a different perspective on this discussion within the domain of linguistic communication, arguing that language evolution to a large extent has been shaped by language learning.

A rapidly growing body of work on the evolution of language is focusing on the role of learning—often in the guise of "cultural transmission"—in the evolution of linguistic communication (e.g., Batali, 1998; Christiansen, 1994; Deacon, 1997; Kirby and Hurford, 2002). Instead of concentrating on biological changes to accommodate language, this approach stresses the adaptation of linguistic structures to the biological substrate of the human brain. Languages are viewed as dynamical systems of communication that are subject to selection pressures arising from limitations on human learning and processing. From this perspective, language evolution can be construed as being shaped by language development, rather than vice versa.[1]

Computational simulations have proved to be a useful tool for investigating the impact of learning on the evolution of language. Connectionist models (also sometimes referred to as "artificial neural networks" or "parallel distributed processing models") provide a natural framework for exploring a learning-based perspective on language evolution because they have been applied extensively to model the development of language (see, e.g., Bates and Elman, 1993; MacWhinney, 2003; Plunkett, 1995; Seidenberg and MacDonald, 2001, for reviews). In this chapter, we show how language evolution may have been shaped by developmental constraints on language acquisition. First, we discuss

Oller & Griebel (eds.), Evolution of Communication Systems. MIT Press.

connectionist models in which the explanations of particular aspects of language evolution and linguistic change depend crucially on the learning properties of specific networks—properties that have also been pressed into service to explain similar aspects of language acquisition. We then present two simulations that directly demonstrate how network learning biases over generations can shape the language being learned. Finally, we conclude the chapter with a brief discussion of the possible theoretical advantages of approaching language evolution from a learning-based perspective.

Evolution Through Learning

Connectionist models can be thought of as a kind of "sloppy" statistical function approximator learning from examples to map a set of input patterns onto a set of associated output patterns. The two most important constraints on network learning (at least for the purpose of this chapter) derive from the architecture of the network itself and the statistical makeup of the input–output examples. Differences in network configuration (such as learning algorithms, connectivity, number of unit layers, etc.—see Bishop, 1995; Smolensky et al., 1996) provide important constraints on what can be learned. For example, temporal processing of words in sentences is better captured by recurrent networks in which previous states can affect current states, rather than in simple feed-forward networks in which current states are unaffected by previous states.

These architectural constraints interact with constraints inherent in the input–output examples from which the networks have to learn. In general, frequent patterns are more easily learned than infrequent patterns because repeated presentations of a given input–output pattern will strengthen the weights involved. For example, for a network learning the English past tense, frequently occurring mappings, such as *go* → *went*, are learned more easily than less frequent mappings, such as *lie* → *lay*.

However, low-frequency patterns may be more easily learned if they overlap in part with other patterns. This is because the weights involved in the overlapping features of such patterns will be strengthened by all the patterns that share those features, making it easier for the network to acquire the remaining unshared pattern features. In terms of the English past tense, this means that the partial overlap in the mappings from stem to past tense in *sleep* → *slept*, *weep* → *wept*, *keep* → *kept* (i.e., *-eep* → *-ept*) will make network learning of the these mappings relatively easy even though none of the words have a particularly high frequency of occurrence. Importantly, these two factors—the frequency and regularity (i.e., degree of partial overlap) of patterns—interact with each other. Thus, high-frequency patterns are easily learned independently of whether they are regular or not, whereas the learning of low-frequency patterns suffers if they are not regular (i.e., if they do not have partial overlap with other patterns).

This characteristic of learning in neural networks makes them suitable for capturing human language processing as many aspects of language acquisition and processing involve such frequency by regularity interactions (e.g., auditory word recognition, Lively et al., 1994; visual word recognition, Seidenberg, 1985; English past tense acquisition, Hare and Elman, 1995).

The frequency by regularity interaction also comes into play when processing sequences of words. In English, for example, embedded subject relative clauses such as *that attacked the reporter* in the sentence *The senator that attacked the reporter admitted the error* have a regular ordering of the verb (*attacked*) and the object (*the reporter*)—it is similar to the ordering in simple transitive sentences (e.g., *The senator attacked the reporter*). Embedded object relative clauses, on the other hand, such as *that the reporter attacked* in the sentence *The senator that the reporter attacked admitted the error* have an irregular verb–object ordering with the object (*the senator*) occurring before the verb (*attacked*). The regular nature of subject relative clauses—their patterning with simple transitive sentences—makes them easy to learn and process relative to the irregular object relative clauses; this is reflected in the similar way in which both humans and networks deal with the two kinds of constructions (MacDonald and Christiansen, 2002).

As we shall see next, the frequency by regularity interaction is also important for the connectionist learning-based approach to language evolution. From this perspective, structures that are either frequent or regular are more likely to be transferred from generation to generation of learners than structures that are irregular and have a low frequency of occurrence.

Learning-based Morphological Change

Although the first example comes from the area of morphological change, we suggest that the same principles are likely to have played a role in the evolution of morphological systems as well. Connectionist networks have been applied widely to model the acquisition of past tense and other aspects of morphology (for an overview, see Christiansen and Chater, 2001). The networks' sensitivity to the frequency by regularity interaction has proven crucial to this work. Simulations by Hare and Elman (1995) have demonstrated that these constraints on network learning can also help explain observed patterns of dramatic change in the English system of verb inflection over the past 1,100 years.

The morphological system of Old English (ca. 870) was quite complex, involving at least ten different classes of verb inflection (with a minimum of six of these being "strong"). The simulations involved several "generations" of neural networks, each of which received as input the output generated by a trained network from the previous generation. The first network was trained on data representative of the verb classes from Old English. However, training was stopped before learning could reach optimal performance.

The imperfect output of the first network was used as input for a second-generation net. This reflected the causal role of imperfect transmission from learner to learner in language change. Training for the second-generation network also was halted before learning reached asymptote. Output from the second network was then given as input to a third network, and so on, until seven generations were trained.

This training regime led to a gradual change in the morphological system. These changes can be explained by verb frequency in the training corpus and phonological regularity (i.e., phonological overlap between mappings, as in the *-eep* → *-ept* example above). As expected, given the frequency by regularity interaction, the results revealed that membership in small classes, irregular phonological characteristics, and low frequency all contributed to rapid morphological change. High frequency and phonologically regular patterns were much less likely to change. As the morphological system changed through generations, the pattern of simulation results closely resembled the historical change in English verb inflection from a complex past tense system to a dominant "regular" class and small classes of "irregular" verbs.

These simulations demonstrate how constraints on network learning can result in morphological change over time. Although these models cannot address such powerful influences as borrowing from foreign languages or other kinds of social change, we suggest that these learning-based pressures may have been an important force in shaping the evolution of morphological systems more generally. Next, we shall see how similar considerations may help explain the existence of word-order universals.

Learning-based Constraints on Word Order

Despite the considerable diversity that can be observed across the languages of the world, it is also clear that languages share a number of relatively invariant features in the way words are put together to form sentences. We propose that many of these invariant features—or linguistic universals—may derive from learning-based constraints, such as the frequency by regularity interaction. As an example consider the head of a phrase, the particular word in a phrase 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—with respect to the remaining clause material. English is considered to be a head-first language, meaning that the head is most frequently placed first in a phrase, as when the verb is placed before the object noun phrase in a transitive verb phrase such as *eat curry*. In contrast, speakers of Hindi would say the equivalent of *curry eat*, because Hindi is a head-last language.

Christiansen and Devlin (1997) trained simple recurrent networks (SRN; Elman, 1990) on corpora generated by thirty-two different grammars that differed in the regularity of their head ordering (i.e., irregular grammars would have a highly inconsistent mix of head-first and head-final phrases). The networks were trained to predict the next lexical category in a sentence. Importantly, these networks did not have built-in linguistic biases; rather, they were biased toward the learning of complex sequential structure. Nevertheless, the SRNs 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. The more irregular a grammar was, the more erroneous the network performance it elicited. The sequential biases of the networks made the corpora generated by regular grammars considerably easier to acquire than the corpora generated from irregular grammars.

Christiansen and Devlin further used frequency data on the world's natural languages (gleaned from the FANAL database; Dryer, 1992) concerning the specific syntactic constructions used in the simulations. They found that languages incorporating fragments the networks found hard to learn tended to be less frequent than languages the network learned more easily. This suggests that constraints on basic word order may derive from nonlinguistic 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.

In a similar vein, Van Everbroeck (1999) presented network simulations in support of an explanation for language type frequencies based on learning constraints. He 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) 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 (51% subject-object-verb and 23% subject-verb-object, respectively), 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 (object-verb-subject, 0.75%; object-subject-verb, 0.25%). Van Everbroeck argued that these results were a predictable product of network learning and processing constraints.

However, not all of Van Everbroeck's results were directly proportional to actual language type frequencies. For example, verb-subject-object languages account for only 10 percent of the world's language types, but the model's performance on it exceeded performance on the more frequent subject-first languages. In more recent simulations, Lupyan and Christiansen (2002) were able to fit language type frequencies appropriately once they took case markings into account. More important, from the viewpoint of this chapter, they were able to observe a frequency by regularity interaction when modeling the acquisition of English, Italian, Turkish, and Serbo-Croatian.

English relies strongly on word order to signal who does what to whom, and thus has a very regular mapping from words to grammatical roles (e.g., the subject noun always comes before the verb in declarative sentences). Italian has a slightly less regular pattern of word order, but both English and Italian make little use of case. Turkish, although it has a flexible (or irregular) word order, nonetheless has a very regular use of case markings to signal grammatical roles. Serbo-Croatian, on the other hand, has both an irregular word order and a somewhat irregular use of case.

Similar to children (Slobin and Bever, 1982), the networks initially showed the best performance on reversible transitive sentences in Turkish, with English and Italian quickly catching up, and with Serbo-Croatian lagging behind. 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. Of course, with repeated exposure the networks (and the children) learning Serbo-Croatian eventually caught up, as predicted by the frequency by regularity interaction.

Together, the simulations by Christiansen and Devlin, Van Everbroeck, and Lupyan and Christiansen provide support for a connection between learnability and frequency in the world's languages based on the learning and processing properties of connectionist networks. Languages that are more easily learned tend to proliferate, and we propose that such learning-based constraints are crucial to our understanding of how language may have evolved into its current form. However, one limitation regarding the three word-order models is that there is no actual transmission between generations of learners (as was the case in Hare and Elman, 1995). Next, we present a series of simulations in which we show how, through processes of linguistic adaptation, learning-based constraints on language acquisition can shape the language being learned.

The Evolutionary Emergence of Multiple-Cue Integration

An outstanding problem in developmental psycholinguistics is how children overcome initial hurdles in learning language. Upon first glance, these hurdles seem insurmountable: Children must disentangle a continuous stream of speech without any obvious information about syntactic structure. They have to learn to what grammatical categories words belong in their native language, and how to put those words together. However, grammatical categories and syntactic structure are not logically independent. A language's syntax assumes grammatical categories, and grammatical categories assume a particular syntactic distribution. The task of acquiring language therefore presents a "bootstrapping" problem.

A possible solution to this problem has been proposed (Gleitman and Wanner, 1982; Morgan and Demuth, 1996; Christiansen and Dale, 2001), and argues that multiple probabilistic cues in speech provide the child's entering wedge into syntax. Prosodic and phonological sensitivity emerges rapidly in children (for reviews, see Jusczyk, 1997; Kuhl, 1999), and this attunement offers opportunities for languages to contain prosodic and phonological information about linguistic structure. Christiansen and Dale (2001) offered computational support for the hypothesis that integrating multiple probabilistic cues (phonological, prosodic, and distributional) by perceptually attuned general-purpose learning mechanisms may hold the key to how children solve the bootstrapping problem. Multiple cues can provide reliable evidence about linguistic structure that is unavailable from any single source of information.

Much evidence suggests that such cues are present cross-linguistically (see Kelly, 1992, for a review) and are manifested in different combinations or "cue constellations." Our hypothesis is that in order for languages to increase their linguistic complexity without compromising learnability, they have evolved cue constellations that reflect their structure and cater to cognitive constraints imposed by the child's learning mechanisms. Here, we consider the evolution of such cues from a computational perspective. After reviewing the cues available for syntax acquisition, we present two language evolution simulations in which we explore how and why cues may have arisen. In the first, we demonstrate the ways in which cues could have emerged, given a language that is growing in vocabulary size. In the second, we offer an illustration of how growing grammatical complexity can strengthen the importance of cues for language acquisition.

Cues Available for Syntax Acquisition

Three sources of information may guide syntax acquisition: innate knowledge in the form of linguistic universals; language-external information that supplies the relationship between language and world; and language-internal information, such as aspects of phonological, prosodic, and distributional patterns within a language. Although some kind of innate knowledge may play a role in language acquisition, it cannot solve the bootstrapping problem. Even with built-in abstract knowledge about grammatical categories and syntactic rules (e.g., Pinker, 1984), the bootstrapping problem remains formidable: Children must map the right sound strings onto the right grammatical categories while determining the specific syntactic relations between these categories in their native language. Moreover, there now exists strong experimental evidence that children do not initially use abstract linguistic categories, but instead employ novel words as concrete items, thereby challenging the usefulness of hypothesized innate grammatical categories (Tomasello, 2000).

Language-external information may contribute substantially to language acquisition. Correlations between environmental observations relating prior semantic categories (e.g., objects and actions) and grammatical categories (e.g., nouns and verbs) may furnish a "semantic bootstrapping" solution (Pinker, 1984). However, given that children acquire linguistic distinctions with no semantic basis (e.g., gender in French; Karmiloff-Smith, 1979), semantics cannot be the only source of information involved in solving the bootstrapping problem. Another extralinguistic factor is cultural learning, where children may imitate the pairing of linguistic forms and their conventional communicative functions (Tomasello, 2000). Nonetheless, to break down the linguistic forms into relevant units, it appears that cultural learning must be coupled with language-internal learning. Moreover, because the nature of both language-external and innate knowledge is difficult to assess, it is unclear how this knowledge could be quantified: There are no computational models of how such knowledge might be applied to learning basic grammatical structure.

Though it is perhaps not the only source of information involved in bootstrapping the child into language, the potential contribution of language-internal information is more readily quantified. Phonological information—including stress, vowel quality, and duration—may help distinguish grammatical function words (e.g., determiners, prepositions, and conjunctions) from content words (nouns, verbs, adjectives, and adverbs) in English (e.g., Cutler, 1993). Phonological information may help distinguish between nouns and verbs. For example, nouns tend to be longer than verbs in English—a difference that even three-year-olds are sensitive to (Cassidy and Kelly, 1991). These and other phonological cues, such as differences in stress placement in multisyllabic words, have been found to exist cross-linguistically (see Kelly, 1992, for a review).

Prosodic information provides cues for word and phrasal/clausal segmentation, and may help uncover syntactic structure (e.g., Morgan, 1996). Acoustic analyses suggest that differences in pause length, vowel duration, and pitch indicate phrase boundaries in both English and Japanese child-directed speech (Fisher and Tokura, 1996). Infants seem highly sensitive to such language-specific prosodic patterns (for reviews, see e.g., Jusczyk, 1997; Morgan, 1996)—a sensitivity that may start in utero (Mehler et al., 1988). Prosodic information also improves sentence comprehension in two-year-olds (Shady and Gerken, 1999). Results from an artificial language learning experiment with adults show that prosodic marking of syntactic phrase boundaries facilitates learning (Morgan et al., 1987). Unfortunately, prosody is also partly affected by a number of nonsyntactic factors, such as breathing patterns, resulting in an imperfect mapping between prosody and syntax (Fernald and McRoberts, 1996). Nonetheless, infants' sensitivity to prosody provides a rich potential source of syntactic information (Morgan, 1996).

None of these cues in isolation suffice to solve the bootstrapping problem; rather, they must be integrated to overcome the partial reliability of individual cues. Previous connectionist simulations by Christiansen, Allen, and Seidenberg (1998) have pointed to efficient and robust learning methods for multiple-cue integration in speech segmentation.

Integration of phonological (lexical stress), prosodic (utterance boundary), and distributional (phonetic segment sequences) information resulted in reliable segmentation, outperforming the use of individual cues. The efficacy of multiple-cue integration has also been confirmed in artificial language learning experiments (e.g., McDonald and Plauche, 1995).

By age one, children's perceptual attunement is likely to allow them to utilize language-internal probabilistic cues (for reviews see, e.g., Jusczyk, 1997; Kuhl, 1999).[2] For example, infants appear to be sensitive to the acoustic differences between function and content words (Shi et al., 1999) and to the relationship between function words and prosody in speech (Shafer et al., 1998). Young infants can detect differences in number of syllables among isolated words (Bijeljac et al., 1993)—a possible cue to noun/verb differences. Moreover, infants are accomplished distributional learners (e.g., Saffran et al., 1996) and, importantly, they are capable of multiple-cue integration (Mattys et al., 1999). When solving the bootstrapping problem, children are also likely to benefit from specific properties of child-directed speech, such as the predominance of short sentences (Newport et al., 1977) and the cross-linguistically more robust prosody (Kuhl et al., 1997).

This review has indicated the range of language-internal cues available for language acquisition; that these cues affect learning and processing; and that mechanisms exist for multiple-cue integration. In an earlier paper (Christiansen and Dale, 2001), we reported on a series of simulations revealing the computational feasibility of the multiple-cue approach to syntax acquisition. SRNs that faced the task of learning grammatical structure *and* predicting cues actually benefited from the additional burden. Despite previous theoretical reservations about the value of multiple-cue integration (Fernald and McRoberts, 1996), the analysis of network performance revealed that learning under multiple cues results in faster, better, and more uniform learning. In another simulation, SRNs were able to distinguish between relevant cues and distracting cues, and performance did not differ from networks that received just reliable cues. Overall, these simulations offer support for the multiple-cue integration hypothesis in language acquisition. They demonstrate that learners can benefit from multiple cues, and are not distracted by irrelevant language-internal information.

Though Christiansen and Dale (2001) offered computational support for the benefit of multiple cues, they did not investigate how these cues may have emerged in language. The following two simulations address this question and illustrate how learning-based constraints can impinge on the evolution of languages.

Simulation 1: Growing Vocabulary

The following simulation implements a system of language adaptation: Grammars mutate, and are selected on the basis of their learnability. This approach echoes observations by

Christiansen (1994) and Deacon (1997) that language changes much more rapidly than its neurobiological substrate, and the child's brain serves as a kind of habitat in which natural selection applies to individual languages. Languages that were difficult to learn were selected against, and languages that were more easily learned, survived and propagated throughout a population of speakers. This method of simulating language change allows investigation into how cues evolved to contribute to this selection and benefit language learning. In what follows, we describe the networks and the language they learned, the conditions provided for transmitting language across generations, and the resulting patterns of cue constellations in the languages that evolved.

Networks and Grammar SRNs served as language learners in both simulations (see figure 6.1). This type of network has a context layer to which the activation of the hidden unit layer—the network's current internal state—is copied and fed back to the network at the next time step. This provides the network with the ability to learn and process the grammatical structure inherent in sequences of words.

Each SRN had initial weight randomization of [–.05, 0.05], with a learning rate of 0.1 and momentum of 0. Input to the networks consisted of individual words in the form of localist representations (one unit was activated for each word). When presented with a word, networks were required to predict the following word in a sentence, along with its corresponding cues. Networks consisted of 12 or 24 word units (depending on the

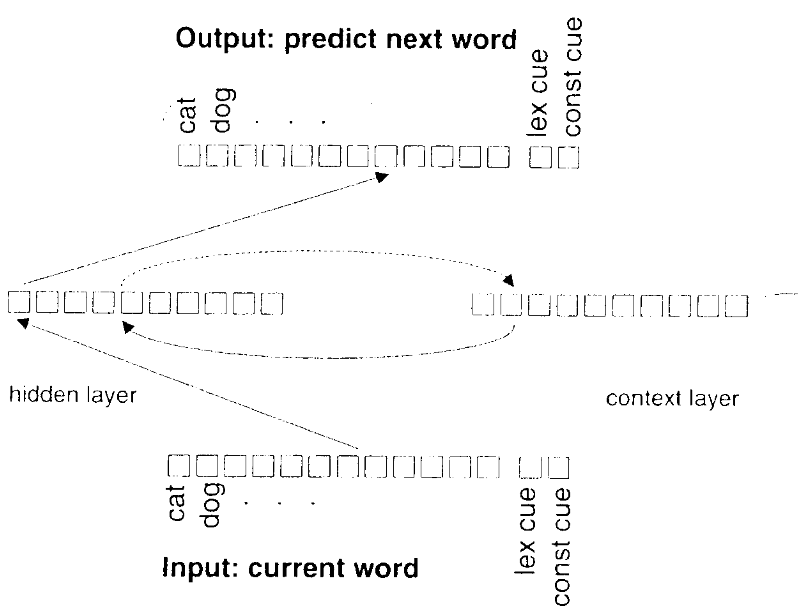

Figure 6.1
Diagram of one SRN agent. Solid lines indicate full connectivity between layers of nodes, and dashed lines indicate one-to-one copy-back connections (lex cue = lexical cue marker; const cue = constituent cue marker).

Table 6.1
The Phrase-Structure Grammar Used in Simulation 1

S → {NP VP}
NP → {PP N}
VP → {V NP}
PP → {P NP}

Note: Brackets indicate the order of these rules was permitted to change.

vocabulary size condition of the simulation) and two cue units, one representing a constituent cue (e.g., pauses) and another activated conjointly with words representing any lexical cue (e.g., primary stress). Each network had ten hidden units and ten context units.

Languages were defined by phrase-structure grammars, a system of rewrite rules defining how sentences are constructed. The phrase-structure grammar "template" used in this simulation is presented in table 6.1. Individual grammars had three changeable features allowing "mutation" with each generation. Head ordering was modified by shifting the constituent order of the four main rewrite rules: S(entence), N(oun)P(hrase), V(erb)P(hrase), and P(repositional)P(hrase). For example, a grammar with the rule PP → P NP, a head-first rule, could be made head-final by simply rewriting PP as NP P, with the head of the prepositional phrase in the final position.

The constituent cue was permitted to potentially mark the boundary of the four rewrite rules. This cue was modified by addition, deletion, and movement (from one rewrite rule to another). Finally, all words were permitted to be associated with the lexical cue. Cues could be added to words, deleted from them, or moved from one word to another. This process was applied across all words and was not specific to any particular grammatical category. The constituent cue was represented as a single unit activated separately after its corresponding phrase-structure rules. The lexical cue was a single unit coactivated with lexical items during training.

Two grammar templates were created for two separate sets of simulation runs. These grammars differed only in the size of their vocabulary, the first being half (12 words) of the second (24 words).

Procedure The grammar template was initially randomized to form five different languages, and each language was learned by five different networks (25 networks in total). Networks were trained on 3,000 randomly generated sentences of their respective grammar (approximately 15,000 word presentations). The performance of each language's five networks was averaged, and the language most easily learned produced linguistic "offspring" for the next generation of networks. Performance was based on a test corpus of 100 randomly generated sentences. The winning language, and four variations of it, served as the

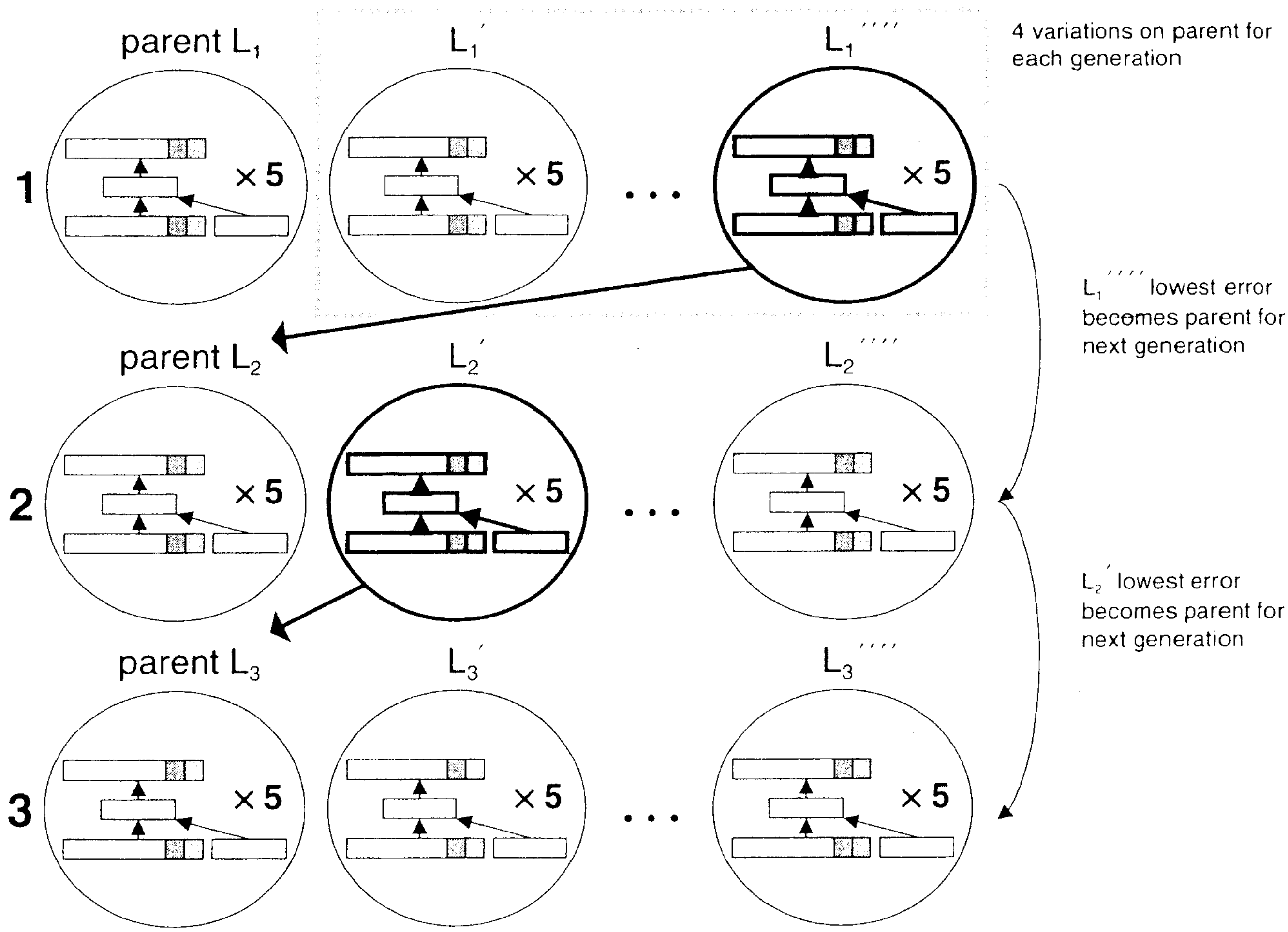

Figure 6.2
Evolutionary simulation schematic of the first three generations of 500.

five languages for the next generation. Variations were formed by randomly selecting two of the three features of the grammar to modify (as described above). The simulation was halted after 500 generations. Ten differently seeded simulation runs were performed. Figure 6.2 illustrates the procedure.

Analysis

Head Order Christiansen and Devlin (1997), as described previously, argued that head-order regularity is a consequence of learning constraints. SRNs in their simulation were better able to learn languages that had head-order regular rules. Similarly, in this simulation we observed head-order regularity of languages across generations. We associated with each winning grammar a score based on the proportion of rules consistently head-first or head-final.

Constituent Cue We observed the ways in which evolving languages incorporated the constituent cue, and its consonance with what is observed in child-directed speech.

Lexical Cue Length, stress, and other lexical cues in language benefit the child to the extent that they delimit grammatical classes. To measure this in the simulation, we performed a simple comparison of how the lexical cue associated with different classes. We used the magnitude of the maximum difference of association among grammatical categories. Formulaically, we measured cue relevance using

$$\max\left|\sum x_{i+}/\sum x_i - \sum x_{j+}/\sum x_j\right|$$

where x_i denotes a grammatical class and x_{i+} denotes words of that class that have an associated lexical cue. This approach to measuring cue relevance is beneficial for two reasons. First, if the cue is unimportant and does not become associated with any words, or becomes associated with all of them, the value of cue relevance will be 0 (0 percent relevant). If the cue separates any two-word classes completely, then cue relevance will be 1 (100 percent relevant). Second, this interval of [0, 1] allows us to graphically represent how the lexical cue becomes exploited across language generations.

Results Even though all simulation runs started with random grammars without consistent head ordering or use of the constituent and lexical cues, we expected that coherent cue constellations would emerge over generations.

Head Order Languages did not evolve head-ordering regularity in any runs of the simulation, in both vocabulary sizes.

Constituent Cue In all runs of the simulation, the constituent cue quickly delimited NP and VP rules, consistent with child-directed speech (Fisher and Tokura, 1996).

Lexical Cue Only in the simulation runs with a large vocabulary did languages exploit the lexical cue. As seen in figure 6.3, languages with a larger vocabulary remained highly consistent in use of the lexical cue, in which it clearly delimits two-word classes across generations. Using the area under these graphs as a measure of cue consistency across generations, it was found that larger vocabularies were more consistent in their use of the lexical cue than small ones ($p < .05$).

Simulation 2: Growing Grammatical Complexity

Simulation 1 suffers from a few limitations. First, the grammatical template used was very simple, and may not fully capture the importance of cues in emerging syntactic structure. Second, the simulations were unable to settle on a particular grammar, but would continuously change back and forth between several possible grammars. Finally, in contrast to

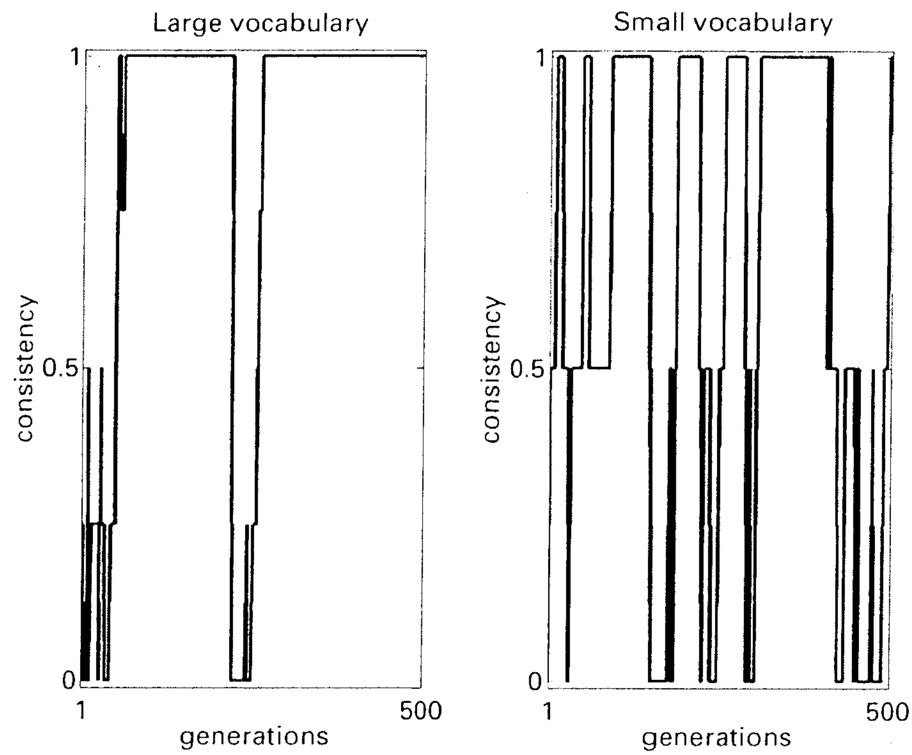

Figure 6.3
A small and large vocabulary run similarly seeded. Languages with larger vocabulary better exploit the lexical cue. The strength of a lexical cue varies from 0 to 1, and is computed by calculating the maximum extent to which a cue distinguishes any two lexical classes (0 = distinguishes no classes; 1 = signals with 100 percent probability two lexical classes in the vocabulary).

the simulation of Christiansen and Devlin (1997), we did not observe a strong effect of regular head ordering. Simulation 2 was intended to overcome these limitations.

Networks and Grammar The networks in this simulation were the same as those in simulation 1, and the same learning parameters were used. The selection process in this simulation, however, was based on a considerably more complex grammar template with two additional rules that encoded a recursive possessive phrase (PossP; see table 6.2). This grammar template is the same as the phrase-structure grammar used in Christiansen and Devlin.

Procedure The procedure mirrored that of simulation 1, with 3,000 sentences for training and 100 for testing. Mutation of the languages was accomplished in the same way, and winning grammars were again selected on the basis of their learnability. Runs of this simulation were halted after the winning language remained the same for 50 generations.

Analysis Head-order regularity, constituent cue use, and lexical cue consistency were measured as in simulation 1.

Table 6.2
The Phrase-Structure Grammar Used in Simulation 2

S → {NP VP}
NP → {PP N}
VP → {V NP}
PP → {P NP}
NP → {PossP N}
PossP → {Poss NP}

Note: Brackets indicate the order of these rules was permitted to change.

Results Nine of our ten simulation runs stabilized on one particular language variation. Of those nine, the following results were observed.

Head Order All languages had highly regular head ordering (i.e., at least five of the six rules were consistently head-first or head-final).

Constituent Cue As in simulation 1, the constituent cue consistently delimited plausible aspects of the grammar template. All runs of the simulation rapidly evolved languages delimiting NP boundaries, again consistent with child-directed speech.

Lexical Cue All stable languages had perfectly consistent lexical cues. Interestingly, six of the nine that stabilized, evolved lexical cues that separated function words from content words, much like English and other natural languages.

Summary of Simulations

These simulations explored two ways in which languages can evolve, and how these conditions influence the emergence of cues to service language acquisition. Simulation 1 revealed that constituent cues, such as pauses or pitch modulation, are highly important in initial syntactic structure and emerge quickly. A growing vocabulary, however, enabled languages to exploit subtler lexical cues, such as word length and lexical stress, to delimit grammatical classes. Simulation 2 revealed that growing grammatical complexity compels languages to incorporate both constituent and lexical cues for syntax acquisition. Together, these simulations illuminate how ontogenetic constraints can guide the evolution of languages. The learning-based constraints imposed by neural network learners shaped the form of the emerging languages across generations.

Conclusion

In this chapter, we have sought to turn the discussion of whether or not ontogeny recapitulates phylogeny on its head. At least when it comes to language, we have proposed that development to a large extent has shaped the evolution of our linguistic abilities, rather than vice versa. Consequently, we have emphasized the role of learning-based constraints in the evolution of linguistic structure, instead of biological changes to accommodate language. Connectionism provides a natural framework for studying a learning-based approach to language evolution, given its widespread application to the modeling of language development. Indeed, we have seen that the specific network properties which have proven crucial for modeling developmental patterns in language acquisition, such as the frequency by regularity interaction, also provide a basis for explaining language evolution.

We have presented two series of connectionist simulations in which learning biases over generations lead to the emergence of multiple-cue integration through linguistic adaptation. Importantly, the nature of the emergent cue systems was similar to the kind of cue systems that young infants have been shown to use in language acquisition. These cue systems appear to emerge to service growing linguistic structure. Fueled by constraints on learning, cue integration becomes a vehicle for facilitating the acquisition of complex linguistic structure. Languages employing cues become more likely to survive the processes of cultural transmission across generations, demonstrating how learning can shape evolution.

On a more theoretical level, our learning-based approach to language evolution may allow us to deal productively with Lewontin's (1998) scathing critique of evolutionary approaches to cognition, and to language evolution in particular: "Reconstructions of the evolutionary history and the causal mechanisms of the acquisition of linguistic competence . . . are nothing more than a mixture of pure speculation and inventive stories" (p. 111). He argues that we are unlikely to find solid evidence that there are heritable variations in linguistic abilities among individuals in the hominid lineage, and that these variations lead individuals with greater abilities to have more offspring. Lewontin's main concern is that we simply cannot test the hypotheses put forward to explain language evolution because of our limited knowledge about hominid evolution in general. However, if, as we have suggested here, language has evolved largely through cultural transmission constrained by limitations on human learning and processing, we can test these hypotheses through computational simulations and human experimentation (Christiansen and Ellefson, 2002; Christiansen et al., 2002).

Acknowledgment

The research reported in this chapter was supported in part by a grant from the Human Frontiers Science Program to Morten H. Christiansen.

Notes

1. For a more detailed description of this approach, see Kirby and Christiansen (2003). For a review placing the cultural transmission approach in the context of contemporary theories of language evolution, see Christiansen and Kirby (2003).

2. Acoustic and articulatory speech science has provided a strong historical basis for these psycholinguistic analyses, including some early clues that prosodic information may point toward syntactic structure. Oller (2000; chapter 4, this volume) presents a theory of communicative evolution that has grown partly out of this tradition.

References

Batali J (1998) Computational simulations of the emergence of grammar. In: Approaches to the Evolution of Language: Social and Cognitive Bases (Hurford JR, Studdert-Kennedy M, Knight C, eds.), 405–426. Cambridge: Cambridge University Press.

Bates E, Elman JL (1993) Connectionism and the study of change. In: Brain Development and Cognition: A Reader (Johnson MH, ed.), 623–642. Oxford: Blackwell.

Bijeljac R, Bertoncini J, Mehler J (1993) How do 4-day-old infants categorize multisyllabic utterances? Devel Psych 29: 711–721.

Bishop CM (1995) Neural Networks for Pattern Recognition. New York: Oxford University Press.

Cangelosi A (1999) Modeling the evolution of communication: From stimulus associations to grounded symbolic associations. In: Advances in Artificial Life (Proceedings ECAL99 European Conference on Artificial Life) (Floreano D, Nicoud J, Mondada F, eds.), 654–663. Berlin: Springer-Verlag.

Cassidy KW, Kelly MH (1991) Phonological information for grammatical category assignments. J Mem Lang 30: 348–369.

Christiansen MH (1994) Infinite languages and finite minds. Ph.D. thesis, University of Edinburgh.

Christiansen MH, Allen J, Seidenberg MS (1998) Learning to segment speech using multiple cues: A connectionist model. Lang Cog Proc 13: 221–268.

Christiansen MH, Chater N (2001) Connectionist psycholinguistics in perspective. In: Connectionist Psycholinguistics (Christiansen MH, Chater N, eds.), 19–75. Westport, Conn.: Ablex.

Christiansen MH, Dale R (2001) Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In: Proceedings of the 23rd Annual Conference of the Cognitive Science Society, 220–225. Mahwah, N.J.: Lawrence Erlbaum.

Christiansen MH, Dale R, Ellefson MR, Conway CM (2002) The role of sequential learning in language evolution: Computational and experimental studies. In: Simulating the Evolution of Language (Cangelosi A, Parisi D, eds.), 165–187. London: Springer-Verlag.

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

Christiansen MH, Ellefson MR (2002) Linguistic adaptation without linguistic constraints: The role of sequential learning in language evolution. In: Transitions to Language (Wray A, ed.), 335–358. Oxford: Oxford University Press.

Christiansen MH, Kirby S (2003) Language evolution: Consensus and controversies. Trends in Cognitive Sciences 7: 300–307.

Cutler A (1993) Phonological cues to open- and closed-class words in the processing of spoken sentences. J Psycholing Res 22: 109–131.

Deacon TW (1997) The Symbolic Species: The Co-evolution of Language and the Brain. London: Penguin Press.

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

Elman JL (1990) Finding structure in time. Cognit Sci 14: 179–211.

Fernald A, McRoberts G (1996) Prosodic bootstrapping: A critical analysis of the argument and the evidence. In: From Signal to Syntax (Morgan JL, Demuth K, eds.), 365–388. Mahwah, N.J.: Lawrence Erlbaum.

Fisher C, Tokura H (1996) Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. Child Devel 67: 3192–3218.

Gleitman L, Wanner E (1982) Language acquisition: The state of the state of the art. In: Language Acquisition: The State of the Art (Wanner E, Gleitman L, eds.), 3–48. Cambridge: Cambridge University Press.

Hare M, Elman JL (1995) Learning and morphological change. Cognition 56: 61–98.

Hebb DO (1949) Organization of Behavior: A Neuropsychological Theory. New York: John Wiley and Sons.

Jusczyk PW (1997) The Discovery of Spoken Language. Cambridge, Mass.: MIT Press.

Karmiloff-Smith A (1979) A Functional Approach to Child Language: A Study of Determiners and Reference. Cambridge: Cambridge University Press.

Kelly MH (1992) Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. Psych Rev 99: 349–364.

Kirby S, Christiansen MH (2003) From language learning to language evolution. In: Language Evolution (Christiansen MH, Kirby S, eds.), 272–294. New York: Oxford University Press.

Kirby S, Hurford J (2002) The emergence of linguistic structure: An overview of the iterated learning model. In: Simulating the Evolution of Language (Cangelosi A, Parisi D, eds.), 121–148. London: Springer-Verlag.

Kuhl PK (1999) Speech, language, and the brain: Innate preparation for learning. In: Neural Mechanisms of Communication (Konishi M, Hauser M, eds.), 419–450. Cambridge, Mass.: MIT Press.

Kuhl PK, Andruski JE, Chistovich IA, Chistovich LA, Kozhevnikova EV, Ryskina VL, Stolyarova EI, Sundberg U, Lacerda F (1997) Cross-language analysis of phonetic units in language addressed to infants. Science 277: 684–686.

Lewontin RC (1998) The evolution of cognition: Questions we will never answer. In: An Invitation to Cognitive Science, vol. 4: Methods, Models, and Conceptual Issues (Scarborough D, Sternberg S, eds.), 107–131. Cambridge, Mass.: MIT Press.

Lieberman DE (1998) Sphenoid shortening and the evolution of modern human cranial shape. Science 393: 158–162.

Lively SE, Pisoni DB, Goldinger SD (1994) Spoken word recognition. In: Handbook of Psycholinguistics (Gernsbacher MA, ed), 265–318. San Diego: Academic Press.

Livingstone D, Fyfe C (1999) Modelling the evolution of linguistic diversity. In: Advances in Artificial Life, Proceedings of ECAL99 European Conference on Artificial Life (Floreano D, Nicoud J, Mondada F, eds.), 704–708. Berlin: Springer-Verlag.

Lupyan G, Christiansen MH (2002) Case, word order, and language learnability: Insights from connectionist modeling. In: Proceedings of the 24th Annual Conference of the Cognitive Science Society, 596–601. Mahwah, N.J.: Lawrence Erlbaum.

MacDonald MC, Christiansen MH (2002) Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). Psych Rev 109: 35–54.

MacWhinney B (2003) Language acquisition. In: The Handbook of Brain Theory and Neural Networks, 2nd ed. (Arbib MA, ed.), 600–603. Cambridge, Mass.: MIT Press.

Mattys SL, Jusczyk PW, Luce P, Morgan JL (1999) Phonotactic and prosodic effects on word segmentation in infants. Cognit Psych 38: 465–494.

McDonald JL, Plauche M (1995) Single and correlated cues in an artificial language learning paradigm. Lang Speech 38: 223–236.

Mehler J, Jusczyk PW, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C (1988) A precursor of language acquisition in young infants. Cognition 29: 143–178.

Morgan JL (1996) Prosody and the roots of parsing. Lang Cog Proc 11: 69–106.

Morgan JL, Demuth K (1996) From Signal to Syntax. Mahwah, N.J.: Lawrence Erlbaum.

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

Newport EL, Gleitman H, Gleitman LR (1977) Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In: Talking to Children: Language Input and Acquisition (Snow CE, Ferguson CA, eds.), 109–149. Cambridge: Cambridge University Press.

Oliphant M (1999) The learning barrier: Moving from innate to learned systems of communication. Adapt Behav 7: 371–384.

Oller DK (2000) The Emergence of the Speech Capacity. Mahwah, N.J.: Lawrence Erlbaum.

Pinker S (1984) Language Learnability and Language Development. Cambridge, Mass.: Harvard University Press.

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

Plunkett K (1995) Connectionist approaches to language acquisition. In: Handbook of Child Language (Fletcher P, MacWhinney B, eds.), 36–72. Oxford: Blackwell.

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

Seidenberg MS (1985) The time course of phonological code activation in two writing systems. Cognition 19: 1–30.

Seidenberg MS, MacDonald MC (2001) Constraint satisfaction in language acquisition and processing. In: Connectionist Psycholinguistics (Christiansen MH, Chater N, eds.), 281–318. Westport, CT: Ablex.

Shady M, Gerken LA (1999) Grammatical and caregiver cues in early sentence comprehension. J Child Lang 26: 163–175.

Shafer VL, Shucard DW, Shucard JL, Gerken LA (1998) An electrophysiological study of infants' sensitivity to the sound patterns of English speech. J Speech, Lang Hearing Res 41: 874–886.

Shi R, Werker JF, Morgan JL (1999) Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. Cognition 72: B11–B21.

Slobin DI, Bever TG (1982) Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. Cognition 12: 229–265.

Smolensky P, Mozer MC, Rumelhart DE (eds.) (1996) Mathematical Perspectives on Neural Networks. Mahwah, N.J.: Lawrence Erlbaum.

Tomasello M (2000) The item-based nature of children's early syntactic development. Trends Cognit Sci 4: 156–163.

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