Subjacency Constraints without Universal Grammar: Evidence from Artificial Language Learning and Connectionist Modeling

Michelle R. Ellefson (ellefson@siu.edu)

Morten H. Christiansen (morten@siu.edu)
Department of Psychology
Southern Illinois University - Carbondale
Carbondale, IL 62901-6502 USA

Abstract

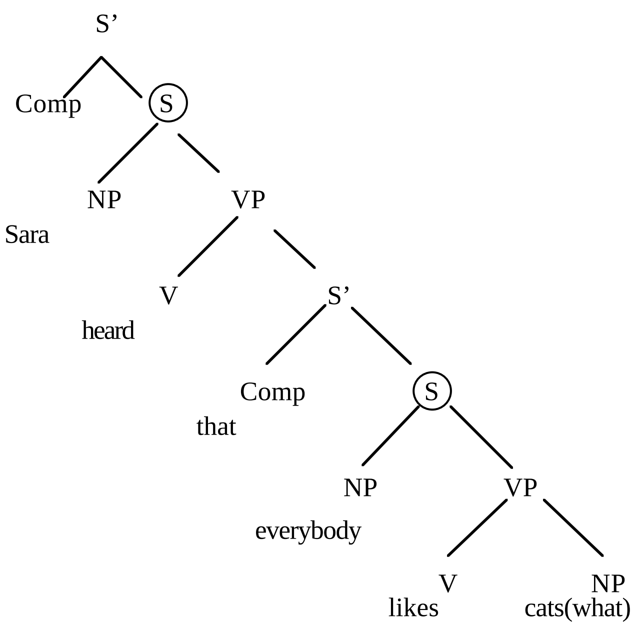
The acquisition and processing of language is governed by a number of universal constraints, many of which undoubtedly derive from innate properties of the human brain. However, language researchers disagree about whether these constraints are linguistic or cognitive in nature. In this paper, we suggest that the constraints on complex question formation, traditionally explained in terms of the linguistic principle of subjacency, may instead derive from limitations on sequential learning. We present results from an artificial language learning experiment in which subjects were trained either on a "natural" language involving no subjacency violations, or on an "unnatural" language that incorporated a limited number of subjacency violations. Although two-thirds of the sentence types were the same across both languages, the natural language was acquired significantly better than its unnatural counterpart. The presence of the unnatural subjacency items negatively affected the learning of the unnatural language as a whole. Connectionist simulations using simple recurrent networks, trained on the same stimuli, replicated these results. This suggests that sequential constraints on learning can explain why subjacency violations are avoided: they make language more difficult to learn. Thus, the constraints on complex question formation may be better explained in terms of innate *cognitive* constraints, rather than linguistic constraints deriving from an innate Universal Grammar.

Introduction

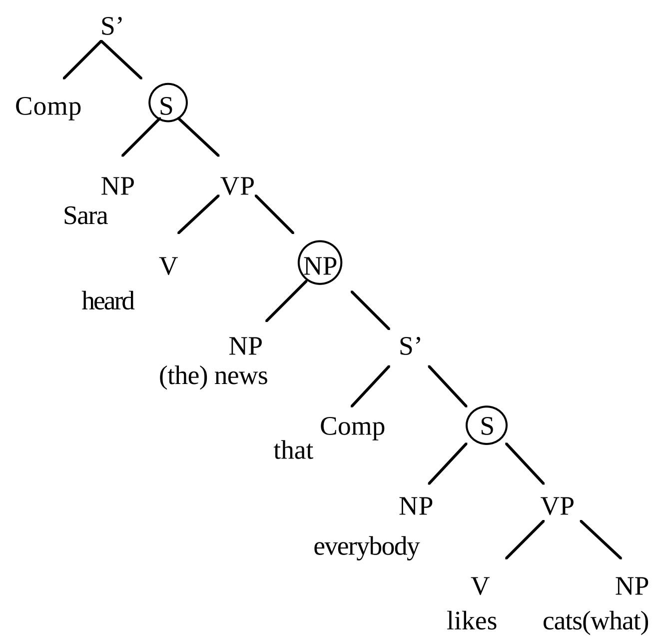
One aspect of language that any comprehensive theory of language must explain is the existence of linguistic universals. The notion of language universals refers to the observation that although the space of logically possible linguistic subpatterns is vast; the languages of the world only take up a small part of it. That is, there are certain universal tendencies in how languages are structured and used. Theories of language evolution seek to explain how these constraints may have evolved in the hominid lineage. Some theories suggest that the evolution of a Chomskyan Universal Grammar (UG) underlies these universal constraints (e.g., Pinker & Bloom, 1990). More recently, an alternative perspective is gaining ground. This approach advocates a refocus in evolutionary thinking; stressing the adaptation of linguistic structures to the human brain rather than vice versa (e.g., Christiansen, 1994; Kirby, 1998). Language has evolved to fit sequential learning and processing mechanisms existing prior to the appearance of language. These mechanisms presumably also underwent changes after the emergence of language, but the selective pressures are likely to have come not only from language but also from other kinds of complex hierarchical processing, such as the need for increasingly complex manual combination following tool sophistication. On this account, many language universals may reflect non-linguistic, cognitive constraints on learning and processing of sequential structure rather than innate UG.

This perspective on language evolution also has important implications for current theories of language acquisition and processing in that it suggests that many of the cognitive constraints that have shaped the evolution of language are still at play in our current language ability. If this is correct, it should be possible to uncover the source of some linguistic universal in human performance on sequential learning tasks. Christiansen (2000; Christiansen & Devlin, 1997) has previously explored this possibility in terms of a sequential learning explanation of basic word order universals. He presented converging evidence from theoretical considerations regarding rule interactions, connectionist simulations, typological language analyses, and artificial language learning in normal adults and aphasic patients, corroborating the idea of cognitive constraints on basic word order universals.

In this paper, we take a similar approach to one of the classic linguistic universals: subjacency. We first briefly discuss some of the linguistic data that have given rise to the subjacency principle. Next, we present an artificial language learning experiment that investigates our hypothesis that limitations on sequential learning rather than an innate subjacency principle provide the appropriate constraints on complex question formation. Finally, we report on a set of connectionist simulations in which networks are trained on the same material as the humans, and with very similar results. Taken together, the results from the artificial language learning experiment and the connectionist simulations support our idea that subjacency violations are avoided, not because of an innate subjacency principle, but because of cognitive constraints on sequential learning.



2a. Sara heard that everybody likes cats.
2. What (did) Sara hear that everybody likes?



1. Sara heard (the) news that everybody likes cats.
3. * What (did) Sara hear (the) news that everybody likes?

Figure 1. Syntactic trees showing grammatical (2) and ungrammatical (3) Wh-movement.

Why Subjacency?

According to Pinker and Bloom (1990), subjacency is one of the classic examples of an arbitrary linguistic constraint that makes sense only from a linguistic perspective. Informally, The subjacency principle involves the assumption of certain principles governing the grammaticality of sentences. "Subjacency, in effect, keeps rules from relating elements that are ‘too far apart from each other’, where the distance apart is defined in term of the number of designated nodes that there are between them" (Newmeyer, 1991, p. 12). Consider the following sentences:

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

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

The subjacency principle, in effect, places certain restrictions on the ordering of words in complex questions. The movement of wh-items (*what* in Figure 1) is limited as far as the number of so-called bounding nodes that it may cross during its upward movement. In Figure 1, these bounding nodes are the S and NP’s which are circled. Put informally, as a wh-item moves up the tree it can use comps as temporary “landing sites” from which to launch the next move. The subjacency principle states that during any move only a single bounding node may be crossed. Sentence 2 is therefore grammatical because only one bounding node is crossed for each of the two moves to the top comp node. Sentence 3 is ungrammatical, however, because the wh-item has to cross two bounding nodes—NP and S—between the temporary comp landing site and the topmost comp.

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

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

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

In the remainder of this paper, we explore an alternative explanation of the restrictions on complex question formation. This alternative explanation suggests that subjacency violations are avoided, not because of a biological adaptation

Table 1. The Structure of the Natural and Unnatural Languages (with Examples)

NAT		UNNAT	
Sentence	Letter String Example	Sentence	Letter String Example
1. N V N	Z V X	1. N V N	Z V X
2. Wh N V	Q Z M	2. Wh N V	Q Z M
3. N V N comp N V N	Q X M S X V	3. N V N comp N V N	Q X M S X V
4. N V Wh N V N	X M Q X M X	4. N V Wh N V N	X M Q X M X
5. Wh N V comp N V	Q X V S Z M	5*. Wh N V N comp N V	Q X V X S Z M
6. Wh N V Wh N V N	Q Z V Q Z V Z	6*. Wh N V Wh N V	Q Z V Q Z V

Note: Nouns (N) = {Z, X}; Verbs (V) = {V, M}; comp = S; Wh = Q.

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

Artificial Language Experiment

Artificial language learning has been shown to be an effective tool in the understanding of the acquisition of language (e.g., Gomez & Gerken, 1999; Saffran, Aslin, & Newport, 1996). More recently, artificial language learning has been used to explore how languages themselves may have evolved in the human species.

Subjects

Sixty undergraduates were recruited from an introductory psychology class at Southern Illinois University, and earned course credit for their participation.

Materials

We created two artificial languages, natural (NAT) and unnatural (UNNAT). Each artificial language consisted of a set of letter strings. The letters in the strings each represented a specific grammatical class (see Table 1). The letters Z and X represented nouns. V and M stood for verbs. The letter S designated a complementizer. Interrogative pronouns were denoted by the letter Q. These strings were constructed based on the sentence structure of the six examples discussed above. Unique letter strings were created for training and testing sessions.

Training Stimuli Twenty letter strings, 10 of each for NAT and UNNAT, were created to represent grammatical and ungrammatical complex question formation structures (SUB). The grammatical SUB items used for the NAT training, while the ungrammatical SUB items were used for UNNAT training. An example of SUB letter strings for both conditions can be seen in Table 1 as sentences 5 and 6.

An additional 20 general training items were constructed to represent grammatical structures (GEN). These items were the same for both groups. Examples of GEN letter strings for both conditions are sentences 1 through 4 in Table 1. In summary, 10 unique SUB and 20 GEN letters strings were created for the training session.

Test Stimuli An additional set of novel letter strings was created for the test session. For each group there were 30 grammatical items and 30 ungrammatical items. Twenty-eight novel SUBs were constructed. For these unique SUB letter strings there were 14 each, of grammatical and ungrammatical complement structures. For UNNAT the ungrammatical SUBs were scored as grammatical and the grammatical SUBs were scored as ungrammatical. In the NAT condition the grammatical SUBs were scored as grammatical and the ungrammatical SUBs were scored as ungrammatical. Testing in both groups also included 16 novel grammatical GEN items and 16 novel ungrammatical GEN items in which one of the letters, except those in the first and last position, were changed.

A test item can be divided into a number of two and three letter fragments. The relative frequency with which these fragments occur in the training set can affect how the test item will be classified by the human subjects. We therefore controlled our stimuli for five different kinds of fragment information to ensure that the structural differences between the two languages would be the only remaining explanation for the expected differential learning of them.

1) *Associative chunk strength* is measured as the sum of the frequency of occurrence in the training items of each of the fragments in a test item, weighted by the number of fragments in that item (Knowlton & Squire, 1994). E.g., the associative chunk strength of the item ZVX would be calculated as the sum of the frequencies of the fragments ZV, VX and ZVX divided by 3. Two-tailed t-tests indicated that there were no differences across the languages in associative chunk strength for the grammatical ($t<1$) and the ungrammatical ($t<1$) items.

2) *Anchor strength* is measured as the relative frequency of initial and final fragments in similar anchor positions in the training items (Knowlton & Squire, 1994). E.g., the anchor strength of the item QXMSXV is calculated as the sum of

the frequencies of the fragments QX and QXM in initial positions in the training items and of the fragments XV and SXV in final positions in the training items. Again, there were no differences across the two languages in the anchor strength of the grammatical ($t(58)=1.75$, $p>.085$) or the ungrammatical items ($t<1$).

3) *Novelty* is measured as the number of fragments that did not appear in any training item (Redington & Chater, 1996). E.g., if the fragments XVS and VS from the item QXVSZM never occurred in a training item, then the test item would receive a novelty score of 2. Here there is a significant difference between the novelty scores for the grammatical test items in the NAT language (.43) and the UNNAT language (0) ($t(58)=3.50$, $p<.001$). However, given that items with novel fragments will seem less familiar they are more likely to not to be accepted as grammatical, making it more difficult to correctly classify the test items from the NAT language. Thus this difference provides a bias against our hypothesis that the NAT language should be easier to learn. There were no differences between the ungrammatical items ($t<1$).

4) *Novel fragment position* is measured as the number of fragments that occur in novel absolute positions where they did not occur in any training item (Johnstone & Shanks, 1999). E.g., if the fragment VQZ from the item QZVQZV never occurred in this absolute position in any of the training items then this item would be assigned a novel fragment position score of 1. There were no differences between the novel fragment scores for the grammatical ($t(58)=1.54$, $p>.13$) or ungrammatical items ($t<1$) across the two languages.

5) *Global similarity* is measured as the number of letters that a test item is different from the nearest training item (Vokey & Brooks, 1992). E.g., if the test item QZM has QZV as its closest training item then it would be assigned a global similarity score of 1. There were no differences between the two languages for the grammatical ($t=0$) and ungrammatical ($t<1$) items.

Procedures

Subjects were randomly assigned to one of three conditions (NAT, UNNAT, and CONTROL). NAT and UNNAT were trained using the natural and unnatural languages, respectively. The CONTROL group completed only the test session. During training, individual letter strings were presented briefly on a computer. After each presentation, participants were prompted to enter the letter string using the keyboard. Training consisted of 2 blocks of the 30 items, presented randomly. During the test session, participants decided if the test items were created by the same (grammatical) or different (ungrammatical) rules as the training items. Testing consisted of 2 blocks of 60 items, again presented randomly.

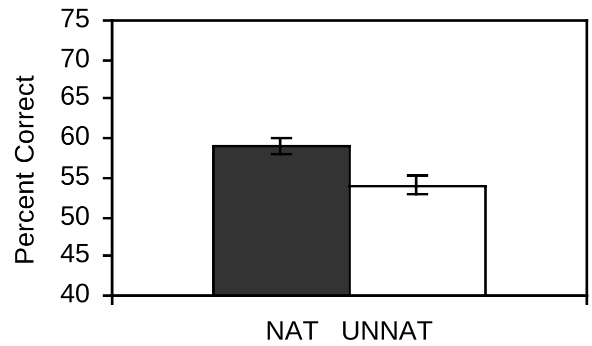


Figure 2. Overall correct classification for NAT and UNNAT languages.

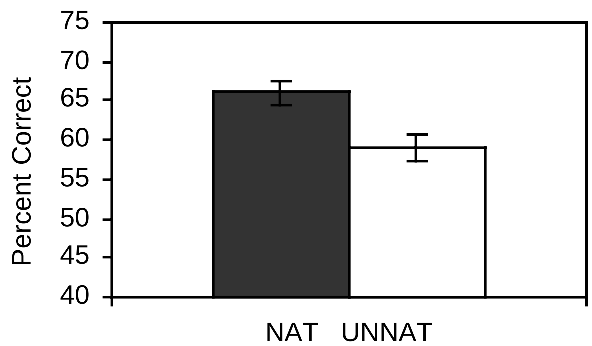


Figure 3. Correct classification of GEN items for NAT and UNNAT langauges.

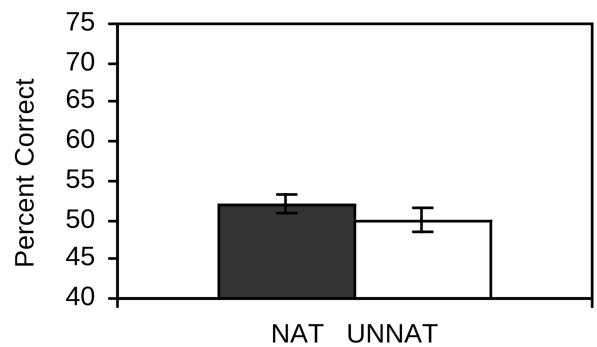


Figure 4. Correct classification of SUB items for NAT and UNNAT languages.

Results and Discussion

Control Group Since the test items were the same for all groups, but scored differently depending on training condition, the control data was scored from the viewpoint of both the natural and unnatural languages. Differences between correct and incorrect classification from both language perspectives were non-significant with all t-values <1 (range of correct classification: 59%–61%). Thus, there was no inherent bias in the test stimuli toward either language.

Experimental Group An overall t-test indicated that NAT (59%) learned the language significantly better than UNNAT (54%) (Figure 2; $t(38)=3.27$, $p<.01$). This result indicates that the UNNAT was more difficult to learn than the NAT. Both groups were able to differentiate the grammatical and ungrammatical items (NAT: $t(38)=4.67$, $p<.001$; UNNAT: $t(38)=2.07$, $p<.05$). NAT correctly classified 70% of the grammatical and 51% of the ungrammatical items. UNNAT correctly classified 61% of the grammatical and 47% of the ungrammatical items. NAT (66%) exceeded UNNAT (59%) at classifying the common GEN items (Figure 3; $t(38)=2.80$, $p<.01$). Although marginal, NAT (52%) was also better than UNNAT (50%) at classifying SUB items (Figure 4; $t(38)=1.86$, $p=.071$). Note that the presence of the SUB items affected the learning of the GEN items. Even though both groups were tested on exactly the same GEN items, the UNNAT performed significantly worse on these items. Thus, the presence of the subjacency violations in the UNNAT language affected the learning of the *language as a whole*, not just the SUB items. From the viewpoint of language evolution, languages such as UNNAT would loose out in competition with other languages such as NAT because the latter is easier to learn.

Connectionist Model

In principle, one could object that the reason why we found differences between the NAT and the UNNAT groups is because the NAT group is in some way tapping into an innately specified subjacency principle when learning the language. To counter this possible objection and to support our suggestion that the difference in learnability between the two languages is brought about by constraints arising from sequential learning, we present a set of connectionist simulations of our human data.

Networks

For the simulations, we used simple recurrent networks (SRNs; Elman, 1991) because they have been successfully applied in the modeling of both non-linguistic sequential learning (e.g., Christiansen & Devlin, 1997; Cleeremans, 1993) and language processing (e.g., Christiansen, 1994; Elman, 1991). SRNs are standard feed-forward neural networks equipped with an extra layer of so-called context units. The SRNs used in our simulations had 7 input/output units (corresponding to each of the 6 letters plus an end of sentence marker) as well as 8 hidden units and 8 context units. At a particular time step t, an input pattern is propagated through the hidden unit layer to the output layer. At the next time step, $t+1$, the activation of the hidden unit layer at time t is copied back to the context layer and paired

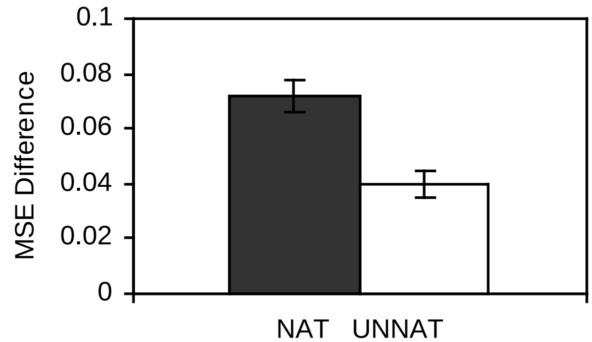


Figure 5. MSE differences for grammatical (low error) and ungrammatical (high error) items for NAT and UNNAT networks.

with the current input. This means that the current state of the hidden units can influence the processing of subsequent inputs, providing an ability to deal with integrated sequences of input presented successively.

Materials

For the simulations we used the same training and test items as in the artificial language learning experiment.

Procedures

Forty networks with different initial weight randomizations (within $\pm$ .5) were trained to predict the next consonant in a sequence. The networks were randomly assigned to the NAT and UNNAT training conditions, and given 20 pass through a random ordering of the 30 training items appropriate for a given condition. The learning rate was set to .1 and the momentum to .95. After training, the networks were tested separately on the 30 grammatical and 30 ungrammatical test items (again, according to their respective grammar).

Following successful training, an SRN will tend to output a probability distribution of possible next items given the previous sentential context. Performance was measured in terms of how well the networks were able to approximate the correct probability distribution given previous context. The results are reported in terms of the Mean Squared Error (MSE) between network predictions for a test set and the empirically derived, full conditional probabilities given the training set (Elman, 1991). This error measure provides an indication of how well the network has acquired the grammatical regularities underlying a particular language, and thus allows for a direct comparison with our human data.

Results and Discussion

The results show that the NAT networks had a significantly lower MSE (.185; SD: .021) than the UNNAT networks (.206; SD: .023) on the grammatical items ($t(38)=2.85$, $p<.01$). On the ungrammatical items, the NAT nets had a slightly higher error (.258; SD: .036) compared with the UNNAT nets (.246; SD: .034), but this difference was not significant ($t<1$). This pattern resembles the performance of the human subjects where the NAT group was 11% better than the UNNAT group at classifying the grammatical items, though this difference only approached significance (*t(38)=1.10, p=.279*). The difference was only <3% in favor of the NAT group for the ungrammatical items (*t=1*). Also similarly to the human subjects, there was a significant difference between the MSE on the grammatical and the ungrammatical items for both the NAT nets (*t(38)=7.69, p<.001*) and the UNNAT nets (*t(38)=4.33, p<.001*). If one assumes that the greater the difference between the MSE on the grammatical (low error) and the ungrammatical (higher error) items, the easier it should be to distinguish between the two types of items. As illustrated in Figure 5, this provides the NAT networks with a significantly better basis for making such decisions than the UNNAT networks (.072 vs. .040; *t(38)=4.31, p<.001*). Thus, the simulation results closely mimic the behavioral results, corroborating our suggestion that constraints on the learning and processing of sequential structure can explain why subjacency violations tend to be avoided: they were weeded out because they made the sequential structure of language too difficult to learn.

Conclusion

In this paper, we have provided evidence in favor of an alternative account of the universal constraints on complex question formation. The artificial language learning results show that not only are constructions involving subjacency violations hard to learn in and by themselves, but their presence also makes the language as a whole harder to learn. The connectionist simulations further corroborated these results, emphasizing that the observed learning difficulties in relation to the unnatural language arise from non-linguistic constraints on sequential learning. These results, together with the results on word order universals (Christiansen, 2000; Christiansen & Devlin, 1997), suggest that constraints arising from general cognitive processes, such as sequential learning and processing, are likely to play a larger role in sentence processing than has traditionally been assumed. This means that what we observe today as linguistic universals may be stable states that have emerged through an extended process of linguistic evolution. When language itself is viewed as a dynamic 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 our 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.

Acknowledgments

We would like to thank Takashi Furuhata, Lori Smorynski, and Brad Appelhans for their help with data collection.

References

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 order universals*. Paper to be presented at the Third Conference on the Evolution of Language, Paris, France.

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 Associates.

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

Elman, J.L. (1991). Distributed representation, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*, 195–225.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109-135.

Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition*. *25*(2), 524–531.

Kirby, S. (1998). *Language evolution without natural selection: From vocabulary to syntax in a population of learners*. Edinburgh Occasional Paper in Linguistics, EOPL-98-1.

Knowlton, B. J, & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. *20*(1), 79-91.

Newmeyer, F. (1991). Functional explanation in linguistics and the origins of language. *Language and Communication*, *11*(1/2), 3–28.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Brain and Behavioral Sciences*, *13*(4), 707–727.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General*. *125*(2), 123-138.

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