

The Evolution of Subjacency without Universal Grammar: Evidence from Artificial Language Learning

Michelle R. Ellefson and Morten H. Christiansen

*Southern Illinois University
{ellefson, morten}@siu.edu*

The acquisition and processing of language is governed by a number of universal constraints. Undoubtedly, many of these constraints derive from innate properties of the human brain. Theories of language evolution seek to explain how these constraints evolved in the hominid lineage. Some theories suggest that the evolution of a Chomskyan universal grammar (UG) underlies these universal constraints. 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). On this account, many language universals may reflect non-linguistic, cognitive constraints on learning and processing of sequential structure rather than innate UG. If this is correct, it should be possible to uncover the source of some linguistic universal in human performance on sequential learning tasks. This prediction has been borne out in previous work by Christiansen (2000) in terms of an explanation of basic word order universals. In this paper, we take a similar approach to one of the classic linguistic universals: subjacency.

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, "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 sentences in Table 1. According to the subjacency principle, sentences 3 and 6 are ungrammatical because too many boundary nodes are placed between the interrogative pronouns and their respective 'gaps'. In the remainder of this paper, we explore 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.

Table 1. Examples of Grammatical and Ungrammatical NP- and Wh-Complements

1. Sara asked why everyone likes cats. N V Wh N V N	4. Sara heard (the) news that everybody likes cats. N V N Comp N V N
2. Who (did) Sara ask why everyone likes cats? Wh N V Wh N V N	5. What (did) Sara hear that everybody likes? Wh N V Comp N V
3. *What (did) Sara ask why everyone likes? Wh N V Wh N V	6. *What (did) Sara hear (the) news that everybody likes? Wh N V N Comp N V

Artificial Language Experiment

We created two artificial languages, natural (NAT) and unnatural (UNNAT), consisting of letter strings, derived from a basis of 6 different constructions (see Table 2). Each training set consisted of 30 items. In NAT training, 10 items were grammatical complement structures involving complex extractions in accordance with subadjacency (SUB) (5 and 6 in Table 2). For UNNAT training, the 10 SUB items involved subadjacency violations (5* and 6*). The 20 remaining training items were general grammatical structures (GEN) that were the same for both groups (1–4 in Table 2). The test set contained 60 novel strings, 30 grammatical and 30 ungrammatical for each group. Twenty-eight novel SUB items, 14 each, grammatical and ungrammatical complex extraction structures were created. For UNNAT, ungrammatical SUB items were scored as grammatical and grammatical SUB items were scored as ungrammatical. The reverse was true for NAT. We created 16 novel grammatical GEN items. Sixteen ungrammatical GEN items were created by changing a single letter in each grammatical item, except for those letters in the first or last position. Both training and test items were controlled for length across conditions and balanced according to different types of frequency information.

Table 2. 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.

In total, 60 adults participated in this experiment, 20 in each of three conditions (NAT, UNNAT, and CONTROL). NAT and UNNAT learned the natural and unnatural languages, respectively. CONTROL 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, with 2 blocks of the 60 randomly presented items, participants decided if the test items were created by the same (grammatical) or different (ungrammatical) rules as the training items.

Results and Discussion

Controls. 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%) ($t(38)=3.27, p<.01$). 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 ($t(38)=2.80, p<.01$). Although marginal, NAT (52%) was also better than UNNAT (50%) at classifying SUB items ($t(38)=1.86, p=.071$). 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 lose out in competition with other languages such as NAT because the latter is easier to learn.

Computational 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.

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., 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 consonants 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 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.

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. Following training, the networks were tested separately on the 30 grammatical and 30 ungrammatical items (again, according to their respective grammar). Performance was measured in terms of how well the networks were able to approximate the correct probability distribution given previous context. The results are therefore 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).

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 assume that 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, then the NAT networks would have 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.

Conclusions

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 UNNAT language arise from non-linguistic constraints on sequential learning. 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 UNNAT 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.

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.
- 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.
- 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.
- 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.