

# Using Artificial Language Learning to Study Language Evolution: Exploring the Emergence of Word Order Universals

**Morten H. Christiansen**  
*Southern Illinois University*  
*morten@siu.edu*

The study of the origin and evolution of language must *necessarily* be an interdisciplinary endeavor. Only by amassing evidence from many different disciplines can theorizing about the evolution of language be sufficiently constrained to remove it from the realm of pure speculation and allow it to become an area of legitimate scientific inquiry. Fueled by theoretical constraints derived from recent advances in the brain and cognitive science, the last decade of the twentieth century has seen a resurgence of scientific interest in the origin and evolution of language. Nonetheless, direct experimentation is needed in order to go beyond existing data. Computational modeling has become the paradigm of choice for such experimentation as evidenced by the many computational papers presented at the two previous Evolution of Language conferences. Computational models provide an important tool with which to investigate how various types of constraints may affect the evolution of language. One of the advantages of this approach is that specific constraints and/or interactions between constraints can be studied under controlled circumstances.

In this paper, I point to *artificial language learning* (ALL) as an additional, complementary paradigm for exploring and testing hypotheses about language evolution. ALL involves training human subjects on artificial languages with particular structural constraints, and then testing their knowledge of the language. Because ALL permits researchers to investigate the language learning abilities of infants and children in a highly controlled environment, the paradigm is becoming increasingly popular as a method for studying language acquisition (e.g. Saffran, Aslin & Newport, 1996). I suggest that ALL can similarly be applied to the investigation of issues pertaining to the origin and evolution of language in much the same way as computational modeling is currently being used. In the remainder of this paper, I demonstrate the utility of ALL as a tool for studying the evolution of language by reporting on two ALL experiments that test predictions derived from previous computational work on the constraints governing the emergence of basic word order universals (Christiansen & Devlin, 1997).

## **Explaining the Emergence of Basic word Order Universals**

There is a statistical tendency across the languages of the world to conform to 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. Within the Chomskyan approach to language, head direction consistency has been explained in terms of an innate module (X-bar theory) that specifies constraints on the phrase structure of languages. Pinker (1994) has further suggested that this module emerged as a product of natural selection.

This paper presents an alternative explanation for head-order consistency based on the suggestion by Christiansen (1994) that 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 view, head direction consistency is a by-product of non-linguistic constraints on hierarchically organized temporal sequences. Christiansen & Devlin (1997) provided connectionist simulations in which simple recurrent networks were trained on corpora generated by 32 different grammars with differing amounts of head-order consistency. These networks did not have built-in linguistic biases; yet they were sensitive to the amount of head-order inconsistency found in the grammars. There was a strong correlation between the degree of head-order consistency of a given grammar and the degree to which the network had learned to master the grammatical regularities underlying that grammar: The higher the inconsistency, the more erroneous the network performance. This suggests that constraints on basic word order may derive from non-linguistic constraints on the learning and processing of complex sequential structure, thus obviating the need for an innate X-bar module for this purpose. Grammatical constructions incorporating a high degree of head-order inconsistency are difficult to learn and will therefore tend to disappear, whereas consistent constructions should proliferate in the evolution of language.

If this line of reasoning is correct, one would expect to be able to find evidence of sensitivity to head-order inconsistency in human sequential learning performance. Experiment 1 tests this prediction using an ALL task with normal adults. More generally, this account also predicts a strong association between language processing and the processing of sequential structure. Experiment 2 tests this prediction comparing the performance of agrammatic aphasics with matched controls in an ALL task.

### **Experiment 1: Testing for Sensitivity to Head-Order Consistency in Sequential Learning**

Two artificial languages were created based on two grammars taken from the Christiansen and Devlin (1997) simulations (see Table 1). Note that the consistent grammar is all head-final to avoid possible contamination from the head-initial nature of English. Both grammars encoded subject-noun/verb agreement. Pairs of strings were generated—one from the consistent grammar and one from the inconsistent grammar—using a vocabulary consisting of six consonants (X = plur. N; Z = prep/post; Q = plur. N; V = sing. N; S = sing. V; M = plur. V). Each string in a pair has the same lexical items and the same grammatical structure as the other, but may differ in the sequential ordering of the lexical items depending on the grammar (e.g., the pair VVQXQXS and VQQVXXS). Thirty pairs in which the sequential ordering differed were selected for *training*. Thirty pairs of identical strings differing from the training items were selected to serve as grammatical test items. Thirty ungrammatical test items were generated by changing a single letter in each grammatical item (first and last letters excluded) to produce an item that was ungrammatical according to both grammars.

**Table 1: The Two Grammars Used in Experiment 1**

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

In the consistent condition (CON), 20 subjects were trained on the consistent items. In the inconsistent condition (INCON), 20 subjects were trained on the inconsistent items. During training each string was presented briefly on a computer screen, and the subject prompted to type it in using the keyboard. Subjects in both conditions were trained on three blocks of 30 training items, before being tested on two blocks of the 60 test items. Subjects were informed about the rule-based nature of the stimuli only prior to the test phase, and asked to classify the novel strings according to whether or not they followed the same rules as the training items. In a third control condition, 20 subjects went directly to the test phase.

With a classification performance of 63.0%, the CON group was significantly better at classifying the test items than the INCON group with only 58.3% ( $t(38)=2.54$ ,  $p<.02$ ). The CON group (67.8%) was numerically better than the INCON (65.8%) at classifying grammatical items, though this difference was not significant ( $t < 1$ ). However, the INCON group was essentially at chance when it came to rejecting the ungrammatical items (51.7%), and performed significantly worse than the CON group (58.1%) on these items ( $t(38)=2.47$ ,  $p<.02$ ). With a correct classification of only 52.38%, the performance of the control subjects demonstrated that there was no bias in the test items. Thus, the results show that the subjects found the consistent grammar easier to learn than the inconsistent grammar as reflected in their ability to perform grammaticality judgments on the test items.

### **Experiment 2: Testing for an Association between Sequential Learning and Language**

The theory of language evolution presented here suggests that language evolved to fit sequential learning and processing mechanisms. A straightforward prediction from this is that breakdown of language should be associated with breakdown of sequential learning. Christiansen, Kelly, Shillcock & Greenfield (2000) provide evidence in support for this prediction. They compared the ALL performance of seven agrammatic aphasic patients with seven normal controls matched for age, gender, socio-economic status and reasoning ability. A simple finite-state grammar was used for this experiment, with letter strings replaced by combinations of symbols (using the zapf dingbats font; e.g., \*●■\*■). For training, subjects carried out a match/mismatch pairing task in which they had to determine whether the second of two consecutively presented symbol strings matched the first. Testing was similar to Experiment 1, in that the subjects were informed about the rule-based nature of the training items and asked to classify 40 new test items (half grammatical/half ungrammatical).

Both aphasics and normals were able to perform the match/mismatch training task without any problems. However, the aphasic patients performed at chance (51.1%) on the classification task, significantly worse than the normal controls (62.1%;  $t(12)=2.53$ ,  $p<.03$ ). The controls were significantly better at classifying the grammatical items (65.0%) compared with the aphasics (38.6%;  $t(12)=2.37$ ,  $p<.04$ ). However, both groups were equally good at classifying the ungrammatical items (aphasics: 63.7%; controls: 59.3%;  $t<1$ ). The results suggest that the aphasics had problems with sequential learning in this experiment, whereas amnesics have few problems on similar ALL tasks (Knowlton & Squire, 1994). Thus, as predicted, the language breakdown in aphasia appears to be associated with a breakdown of sequential learning.

### Conclusion

The two ALL experiments and the computational simulations presented by Christiansen & Devlin (1997) support the claim that basic word order universals (head-ordering) can be explained in terms of non-linguistic constraints on sequential learning and processing, rather than as a product of innate linguistic knowledge. Thus, rather than a biological adaptation of learning mechanisms to fit linguistic structure, the evidence points to the adaptation of linguistic structure to fit pre-existing sequential learning mechanisms.

More generally, the results of the two experiments reported above and similar ALL results pertaining to subjacency reported in Ellefson & Christiansen (2000) suggest that ALL is a useful tool for exploring issues relating to language evolution. It may be objected that the languages used in ALL experiments are simple and deviate significantly from natural language. However, the same objection can be raised against the computational models of language evolution, but this has not diminished their impact, nor their usefulness to the study of language evolution. Moreover, ALL also provides a new tool with which to study other aspects of language evolution, such as creolization (Hudson & Newport, 1998) and comparative aspects of language evolution (Hauser, Newport & Aslin, 2000). In this way, ALL promises to open up a whole new direction in the search for evidence to constrain scientific theories of language evolution.

### 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. & 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.
- Christiansen, M.H., Kelly, L., Shillcock, R.C. & Greenfield, K. (2000). *Artificial grammar learning in agrammatism*. Manuscript in preparation.
- Ellefson, M.R. & Christiansen, M.H. (2000). *The evolution of subjacency without Universal Grammar: Evidence from artificial language learning*. Paper to be presented at the 3rd Conference on the Evolution of Language, Paris, France.
- Hauser, M., Newport, E. L. & Aslin, R. N. (2000). *Statistical learning of serial order in cotton-top tamarins*. Submitted manuscript.
- Hudson, C.L. & Newport, E.L. (1998). Creolization: Could adults really have done it all? In A Greenhill, H. Littlefield & C. Tano (Eds.), *Proceedings of the 23rd annual Boston University Conference on Language Development* (pp. 265-276). Boston: Cascadilla Press.
- Knowlton, B.J. & Squire, L.R. (1994). The information acquired during artificial grammar learning. *JEP: Learning, Memory and Cognition*, 20, 79-91.
- Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: William Morrow.
- Saffran, J.R., Aslin, R.N. & Newport, E.L. (1996). Statistical learning by 8-month olds. *Science*, 274, 1926-1928.