

What in the World Makes Recursion so Easy to Learn? A Statistical Account of the Staged Input Effect on Learning a Center-Embedded Structure in Artificial Grammar Learning (AGL)

Fenna H. Poletiek

In an artificial grammar learning study, Lai & Poletiek (2011) found that human participants could learn a center-embedded recursive grammar only if the input during training was presented in a staged fashion. Previous studies on artificial grammar learning, with randomly ordered input, failed to demonstrate learning of such a center-embedded structure. In the account proposed here, the staged input effect is explained by a fine-tuned match between the statistical characteristics of the incrementally organized input and the development of human cognitive learning over time, from low level, linear associative, to hierarchical processing of long distance dependencies. Interestingly, staged input seems to be effective only for learning hierarchical structures, and unhelpful for learning linear grammars.

Keywords: artificial grammar learning; linguistic environment; recursion; staged input; statistical learning

1. Recursion Learning in the Artificial Grammar Learning Paradigm

Language acquisition is one of the most complex tasks imaginable. Young learners, from infancy on, are faced with a noisy, degraded, and small set of streams of sounds — linguistic stimuli — from which grammatical principles have to be induced. Though generalization from the stimuli is needed to learn the grammar, it is bound to complex constraints: It should not go too far, and not be simple and linear. It is one of the most persistent mysteries in cognitive science how humans achieve this goal. How does this learning proceed?

Infants have been observed to induce simple linear statistical structure in a stream of sounds (Saffran 2003). Older children, however, induce highly complex non-linear rules from what they hear. For example, children never erroneously transform a sentence like ‘The man who was here yesterday is Sam’ into the corresponding question ‘Was the man who here yesterday is Sam?’ by simply moving the first encountered subordinate clause verb *was* to the front rather than the main verb *is* (Gómez & Gerken 2000). Moreover, in a statement like ‘The man



the dog bites shouts', the first encountered noun (subject) is associated with the last verb rather than with the first encountered next verb, revealing the application of a hierarchical principle.

The non-linear process required for natural language seems hard to explain with statistical learning mechanisms. Recently (and less recently; see Gold 1967), it has been proposed that this type of hierarchical structures is unique for human language and therefore is a crucial characteristic of the human language faculty (Hauser *et al.* 2002, Fitch & Hauser 2004; see also Corballis 2007). Very little is known, however, about how these structures are actually learned and used, and to what extent general statistical learning mechanisms and the learner's environment factor into acquiring hierarchical structures like center embedding.

The purpose of the present paper is to propose an explanation for a recently found facilitation effect of the organization of the linguistic input on learning a center-embedded structure (Lai & Poletiek 2011). The effect is accounted for in terms of the match between the statistical characteristics of the input and the developmental pattern of the learning process. I propose that the changing organization of the linguistic environment over time narrowly matches the synchronic development of cognitive learning mechanisms, binding formal language complexity to learnability.

Because of the extremely high complexity of natural grammars, little information about fundamental mechanisms involved in natural grammar learning can be derived directly from the features of language. Therefore, a growing body of research on grammar learning uses artificial grammars for both simulation studies and empirical experimentation. A now classical paradigm is the Artificial Grammar Learning (AGL) procedure (Reber 1969, 1993). Reber (1969) exposed human participants to exemplars of a simple finite state grammar (see Figure 1a below) with a few 'words' (mostly letters). Next, participants are given a test task in which new strings are presented, half of which are grammatical and half are not. Participants give grammaticality judgments for each test string. Typically, participants perform significantly above chance level, indicating that the structure was induced during training and applied during the test phase, at least to some extent.

The artificial grammar learning paradigm can be used to perform a laboratory test of possible effects of environmental characteristics on the learnability of sequential structures, by simulating these characteristics in the experimental task and comparing learning behavior under experimental conditions with a matched control condition in which the investigated characteristics are not implemented.

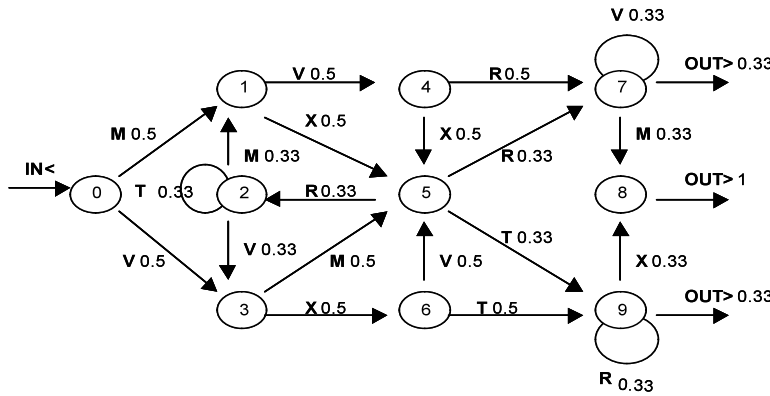
2. Staged Input Facilitates Hierarchical Processing

Artificial grammars with a center-embedded rule have been shown to be extremely hard to learn by induction. Participants failed to show any knowledge of the hierarchical center-embedded structure after exposure to a randomly ordered set of exemplars (de Vries *et al.* 2008). In Lai & Poletiek's (2011) study, Reber's (1969) procedure was slightly adapted. Rather than presenting one learning and one test phase, the task was divided in twelve blocks each with a set of twelve

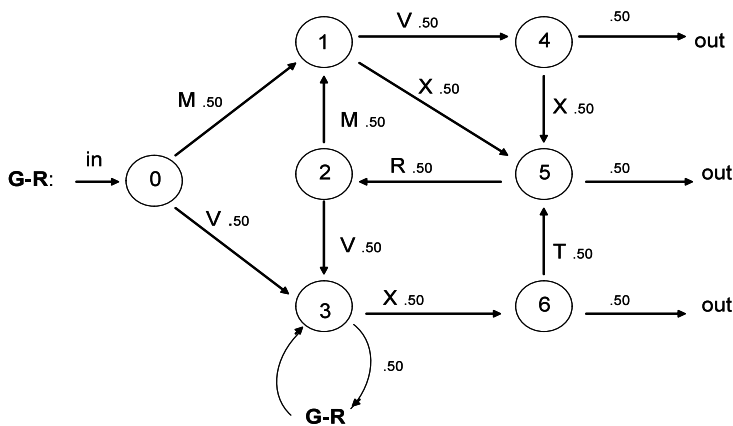
training strings followed by twelve test strings. This procedure allowed us to measure the development over time of grammaticality judgments performance as exposure increases.

In contrast to what de Vries *et al.* (2008) found, our participants performed well — but only if the input exemplars with which they were trained were presented in an incremental fashion, starting with the shortest and least complex exemplars without embeddings and ending with exemplars with multiple levels of embedding. This result suggests that the time course of exposure to hierarchically increasingly complex stimuli, allows cognitive learning of the hierarchical system. A statistical analysis of the input presented incrementally may provide an account of this staged input effect. Moreover, as I will argue below, this analysis also provides an explanation of why hierarchical structures rather than simple linear finite-state structures benefit from an incrementally organized input.

Consider a finite-state structure (G-FS) with five elements (letters M, V, R, X, T) and a hierarchical recursive center-embedded structure (G-R), with the same five elements:



a) Finite-state system G-FS



b) Center-embedded system G-R

Figure 1: Schematic probabilistic Markov-representation of a) artificial finite-state grammar (G-FS) and b) center-embedded recursive structure (G-R).

Both systems generate strings of elements (e.g. MXTRR for G-FS and VMXRMVX for G-R). For both systems, the probability of each unique element the system generates ($p(\text{exemplar} | G)$) can be calculated (van der Mude & Walker 1978, Charniak 1994, Poletiek & Wolters 2009). The sum of the probabilities of all unique strings generated by a system (i.e. the ‘full output’) is, or approximates, one (Poletiek & van Schijndel 2009). The probability distributions of the unique strings generated by both grammars differ. Indeed, it can be shown that the probability distribution of the exemplars generated by G-FS is more ‘even’ than the probability distribution of the strings generated by the recursive structure G-R. In the recursive system, the strings without any center-embedded clause are much more probable than strings with embeddings. Moreover, as the number of levels of embedding increases, the probability of production by the system drops quickly to approximate zero.

If the exemplars of both grammars are ordered in a staged fashion, according to their decreasing probabilities to be produced by G-R and G-FS, then let us assume that learning of the grammars at any point in time may be represented by the sum of the probabilities of the exemplars a learner has been exposed to up to that point in time (Lai & Poletiek 2011). For example, if this sum ($\Sigma(p(\text{exemplar} | G))$) is .50 after exposure to n exemplars presented in a growing fashion (staged input), the learner has been exposed and allowed to learn ‘half’ of the system. In Figure 2, the evolution over time of $\Sigma(p(G | \text{string}))$ is displayed for both G-FS and G-R, for an input presented over time according to decreasing probabilities of the exemplars. Assuming that the cumulative probabilities of the gradually increasing set of input stimuli reflect the proportion of the full language (i.e. 100% of the stimuli it generates) at each stage of exposure, Figure 2 displays the evolution of this cumulative value over time for a Finite State Grammar and for a Recursive Grammar.

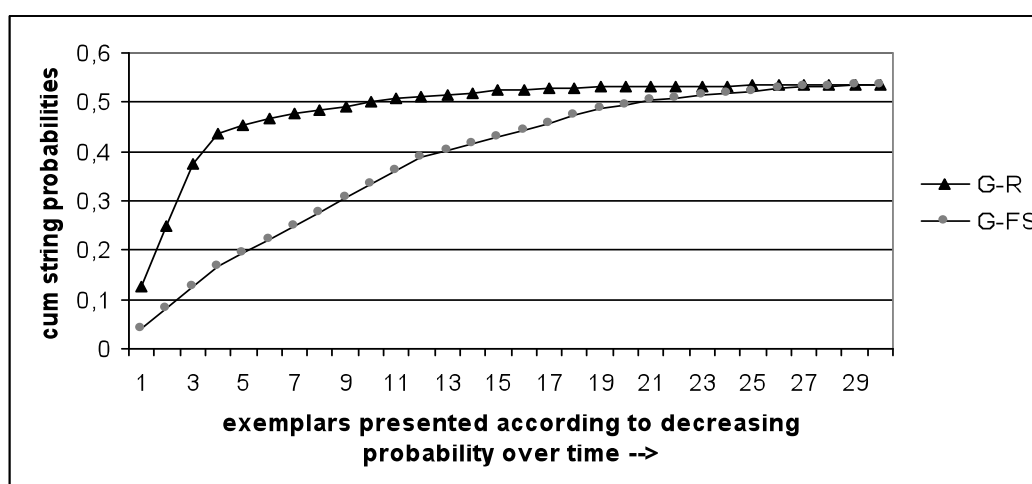


Figure 2: Cumulative exemplar probabilities for exemplars of the grammars G-FS and G-R ranked according to decreasing probabilities.

Consider two learners. After exposure to the 30 most probable exemplars of G-FS learner A has learned 50% of that grammar (Figure 2). Likewise, learner B, after having been exposed to the 30 most probable exemplars of G-R would also have learned 50% of G-R. Thus, the two learners have seen an equal number of exemplars of their ‘own’ grammar, covering an equal part of the grammar generating them. The difference between the two learning situations, however, is in the shapes of the curves.

As Figure 2 shows, in the earliest stage of exposure (e.g., after having seen five exemplars), the proportion (Poletiek & van Schijndel 2009) of the recursive system G-R, covered by the exemplars, exceeds by far the proportion of G-FS after exposure to five exemplars of G-FS. Assuming that the cumulative probabilities curve (y-values) reflects how much of the underlying system has been learned at each point of exposure (x-values), the lines might be considered to model learning curves of the two learners after a given amount of exposure. Figure 2 then reveals that presenting the input in an incremental fashion strongly boosts the learning curve of the *recursive* language in the *early stage* of learning (see also Elman 1993), according to this simulation.

This facilitation effect of staged input for recursive grammars was verified by Lai & Poletiek (2011) in their AGL study. Interestingly, the incremental presentation of the input does not help much for learning a linear finite-state grammar. As can be seen in Figure 2, the non-recursive grammar produces a more linear learning curve, implying a weaker effect of the organization of the input over time for non-recursive linear systems.

3. Artificial and Natural Grammar Learning

Translating this analysis to natural grammar learning requires a mapping of the artificial situation displayed in Figure 2 onto a natural developing language learner and natural linguistic input. A number of arguments can be advanced for the correspondence between the artificial data analysis and natural language learning. First, cognitive learning generally is time-course sensitive (Pine 1994). Not only language is acquired most effectively in the first years of life — also most skills and cognitive functions are learned best when we are young. Moreover, learning mechanisms are mainly statistical and associative during early childhood (Saffran 2003), becoming increasingly sophisticated and covering long-distance dependencies in later stages of learning. Accordingly, during the early stage of life, when the child is exposed to basic and short exemplars of the structure, cognitive processing is simple, linear, and associative, providing important information about the basic rules of the structure. Using this basic knowledge as a stepping stone, the learner’s growing cognitive capacity can process increasingly complex non-linear operations which allow the detection of recursive patterns.

Second, the staged environment assumed in the present artificial world may be argued to represent the linguistic natural environment of a young learner. Indeed, as studies into child-directed speech suggest (Pine 1994), linguistic utterances children are exposed to are simpler, shorter, and contain more frequent constructions than adult language. Only later on is the system to be in-

duced by the natural learner hierarchical and recursive — not linear.

In sum, the present theoretical explanation of the beneficial effect of a staged linguistic input on grammar induction, derived from artificial grammar study, suggests a well-tuned fit between the organization of the linguistic environment and the development of learning abilities. In addition, the model can explain why this facilitation occurs specifically for recursive grammars, and not for linear ones. Generally, as shown in this analysis, artificial grammar studies and statistical models of the effects they reveal are useful tools to our understanding of fundamental processes underlying natural grammar learning.

References

- Charniak, Eugene. 1994. *Statistical Language Learning*. Cambridge, MA: MIT Press.
- Corballis, Michael C. 2007. Recursion, language, and starlings. *Cognitive Science* 31, 697–704.
- Elman, Jeffrey L. 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48, 71–99.
- Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a non-human primate. *Science* 303, 377–380.
- Gold, E. Mark. 1967. Language identification in the limit. *Information and Control* 10, 447–474.
- Gómez, Rebecca L. & LouAnn Gerken. 2000. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences* 4, 178–186.
- Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
- Lai, Jun & Fenna H. Poletiek. 2011. The Impact of Adjacent-Dependencies and Staged-Input on the Learnability of Center-Embedded Hierarchical Structures. *Cognition*, 118(2), 265-273
- Pine, Julian M. 1994. The language of primary caregivers. In Clare Gallaway & Brian J. Richards (eds.), *Input and Interaction in Language Acquisition*, 15–37. Cambridge: Cambridge University Press.
- Poletiek, Fenna H. & Tessa J. P. van Schijndel. 2009. Stimulus set size and grammar coverage in artificial grammar learning. *Psychonomic Bulletin & Review* 16, 1058–1064.
- Poletiek, Fenna H. & Gezinus Wolters. 2009. What is learned about fragments in artificial grammar learning: A transitional probabilities approach. *Quarterly Journal of Experimental Psychology* 62, 868–876.
- Reber, Arthur S. 1969. Transfer of syntactic structures in synthetic languages. *Journal of Experimental Psychology* 81, 115–119.
- Reber, Arthur S. 1993. *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford: Oxford University Press.
- Saffran, Jenny R. 2003. Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science* 12, 110–114.

- van der Mude, Antony & Adrian Walker. 1978. On the inference of stochastic regular grammars. *Information and Control* 38, 310–329.
- de Vries, Meinou H., Padraic Monaghan, Stefan Knecht & Pienie Zwitserlood. 2008. Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition* 107, 763–774.

Fenna H. Poletiek
Leiden University
Cognitive Psychology Department
P.O. Box 9555
2300 RB Leiden
The Netherlands
poletiek@fsw.leidenuniv.nl