The Impact of Starting Small on the Learnability of Recursion

Jun Lai (laij@fsw.leidenuniv.nl)
Cognitive Psychology Department, Leiden University, Pieter de la Court building,
P.O. Box 9555, 2300 RB Leiden, the Netherlands

Fenna H. Poletiek (poletiek@fsw.leidenuniv.nl)
Cognitive Psychology Department, Leiden University, Pieter de la Court building,
P.O. Box 9555, 2300 RB Leiden, the Netherlands

Abstract

Recursion is argued to be the crucial property distinguishing human and non-human primates language learning faculty (Hauser, Chomsky, & Fitch, 2002). Recently, 2 studies (Bahlmann & Friederici, 2006; de Vries, Monaghan, Knecht, & Zwitserlood, 2008), which investigated the learnability of a recursive artificial grammar of the type of A^nB^n, used the same material but reported divergent results. We propose that the organization of the linguistic environment crucially determines learnability of the recursive structure, and that this factor might offer some explanation to the incompatible findings. In a grammaticality judgment task using the same materials as in Bahlmann and Friederici (2006) and de Vries et al.'s (2008), we found significantly better performance when the training input was arranged in a starting small fashion, than when it was organized randomly.

Keywords: Starting small; Recursion; Artificial grammar learning; Statistical learning.

Introduction

Exploring the mechanism behind language learning has been the focus of an enormous body of research in linguistics, psychology and education. The question is how children can possibly acquire such an astonishing complex system so rapidly, while the linguistic environment input is noisy and limited. Sentences like *The rat the cat the dog chased killed ate the malt.* (Chomsky & Miller, 1963) with two recursive center embedding clauses are nearly unintelligible, even for native English speakers (Bach, Brown, & Marslen-Wilson, 1986; Hudson, 1996; Newmeyer, 1988; Vasishth, 2001), due to the associated elements in the sentence being distant from one another (e.g. "the rat" and "ate"). Moreover, recursion is a self-referential principle that can be applied an infinite number of times, producing sentences with numerous embeddings being cognitively very hard to process. Among all syntactical characteristics of natural language, recursion has therefore been argued to be the most fundamental and challenging to acquire (Hauser, Chomsky, & Fitch, 2002).

A recent experimental study (Fitch & Hauser, 2004) using an artificial language has reported that cotton-top tamarins could master the *finite state grammar* (FSG) with the $(AB)^n$ type, but not a higher-level recursive *phrase structure grammar* (PSG) with the A^nB^n type, which could be learned by human participants. Using a familiarization-discrimination paradigm, Fitch and Hauser (2004) first presented the animal participants two auditory sets of consecutive consonant-vowel nonsense syllables (e.g. *la, pa, ba*). Category A syllables were spoken by a female speaker, while Category B syllables by a male. The two sets were identical except for the underlying structure, as well as the pitch. The $(AB)^n$ set in FSG was formed by local transitions between A and B, while the A^nB^n sentences were made according to a center embedding recursive rule (see Figure 1). After this training phase, a discrimination task was performed by the tamarins using the familiarization paradigm. It showed that tamarins could detect the ungrammatical sequences from the grammatical ones in FSG, but not in PSG. Contrastively, humans demonstrated clear discrimination in judging grammaticality of both grammars. This study has raised a renewed interest concerning the inductive learnability of recursive structures, using *artificial grammar learning* (AGL) paradigm (Bahlmann & Friederici, 2006; Bahlmann, J., Schubotz, R.I., & Friederici, A.D., 2008; de Vries, Monaghan, Knecht, & Zwitserlood, 2008; Kersten & Earles, 2001; Perruchet & Rey, 2005). Nevertheless, a study (Gentner, Fenn, Margoliash, & Nusbaum, 2006) concerning song birds' capability of processing A^nB^n structure posed a challenge to this "uniquely human" claim.

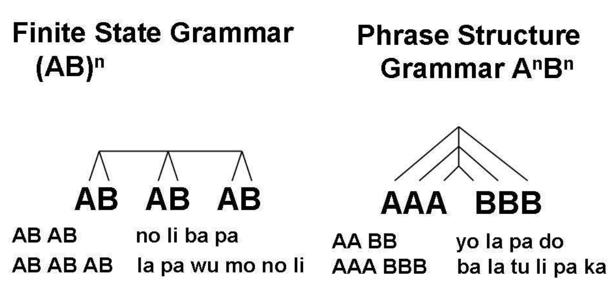

Figure 1. Structures of Finite State Grammar $(AB)^n$ and Phrase Structure Grammar A^nB^n used by Fitch and Hauser (2004). The phrase structure grammar is recursive, center-embedded, and generates long-distance dependencies.

Bahlmann and Friederici (2006, henceforth B&F) and Bahlmann et al. (2008) carried out an fMRI study to probe into the neural basis of processing center-embedding structures in AGL. Significantly stronger activation in Broca's area, involved in natural language processing, was observed in processing of hierarchically recursive structure A^nB^n, than for the $(AB)^n$ grammar. By contrast, de Vries et al. (2008) replicated this study by B&F but reported no learning of center-embedding structures. De Vries et al. (2008) first trained all participants on the same stimuli as B&F, and required participants to judge the grammaticality of new items violating the center-embedding rule. However, participants were tested with different types of violations, namely: *scrambled* (e.g. $A_xA_yA_z\underline{\mathbf{B_xB_zB_y}}$)[1] sequences and *scrambled + repetition* sequences ($A_xA_yA_z\underline{\mathbf{B_xB_yB_x}}$). As they predicted, their participants could detect the scrambled + repetition violations, but not the scrambled ones. Therefore, de Vries et al. (2008) argued that successful performance in the study of B&F was due to alternative heuristics, such as counting or repetition-monitoring, instead of learning the abstract center-embedded principle. Indeed, B&F applied *replacement violations* (e.g. $A_xA_yA_zB_z\underline{\mathbf{A_y}}B_x$) and *concatenation violations* (e.g. $A_xA_yB_y\underline{\mathbf{B_z}}$) in their testing materials, which could possibly also be detected without any knowledge of the center-embedding rule, by merely counting the A's and the B's, or by simply detecting a B that was unrelated to any of the A's in a sequence. De Vries et al. (2008) concluded that surface features of A^nB^n sequences were learned by humans, such as repetition patterns and the match between the number of A's and the number of B's, but not the abstract recursive principle determining the long-distance dependencies between each A and each B in such a sequence. In sum, the learnability of center-embedded structures by mere exposure to input exemplars could not unambiguously be established in research using artificial materials, thus far. It seems still inconclusive to which extent AGL studies could help us understand the mechanism of learning recursion.

Here we propose that two fundamental properties of the training set might point at an alternative account of the inconclusive findings. One crucial property is *starting small*, which is the way learning input is ordered. The notion of starting small was first raised by Elman (1991, 1993). He trained a connectionist network to parse complex structures which contained embedded subordinates. The network succeeded in learning only if it was provided with a staged training input (starting small), but not after exposure to the entire random input as a whole. A number of empirical researches showed supporting evidence for this study (Cochran, McDonald, & Parault, 1999; Kareev, Lieberman, & Lev, 1997; Kersten & Earles, 2001), while some other findings yielded contradictory results (Rohde & Plaut, 1999). Possibly the diverging findings might be explained by the highly different methodologies, such as type of study (experimental designs versus simulation studies), stimulus set, input size, training and testing procedures, or the type of grammar used. An input 'growing' gradually, might be especially efficient for learning a complex recursive structure, when the input contains sequences with long distance dependencies, as in the study of B&F.

The second property is *frequency distribution* of the input. In natural language, simple phrases or sentences with zero-level-of-embedding (0-LoE) appear much more frequently than those with several levels of embeddings (Poletiek & Chater, 2006). In real life, this type of short and typical sentences with only adjacent-dependencies, is encountered much more often than more complex compound sentences with several sub-clauses. Sentences with simple structures occur frequently (Philips, 1973; Pine, 1994; Poletiek & Chater, 2006; Snow, 1972). We propose that the distribution of simple and complex sentences in the input set might play a role in rule induction. In our experiment, we presented the input stimuli of our artificial grammar in a distribution that reflected the unequal occurrence of simple and complex sentences in natural language.

To a large extent, both properties of the input we hypothesize to help learners, also occur in the natural linguistic environment of children. Compared to adult-directed speech, child-directed speech has shorter linguistic constituents, simpler structures, and mainly adjacent-dependencies (Pine, 1994). A large amount of repetitions of syntactically short utterances help children learn the basic structure of language. As children grow, child-directed speech develops into more mature speech types (Bellinger, 1980; Garnica, 1977) because more complex constructions are gradually introduced. Therefore, if we can demonstrate experimentally successful grammar learning with a growing environmental input and unequal frequencies for simple and complex exemplars, this might help understanding environmental factors involved in the mechanism of natural language learning.

In the present study, we tested whether participants could learn the hierarchical recursive rule when the learning set was organized 'starting small' rather than randomly, and when unique simple exemplars were repeated, whilst the complex ones were not. We predict that participants will show learning under these conditions.

Experiment

Method

Participants. Twenty-eight students (20 female), from Leiden University participated in the experiment for course credit or payment. All were native Dutch speakers. All had normal or corrected to normal vision.

Materials and design. The same stimuli were used as in B&F and de Vries et al. (2008). There were two sets of syllables, categorized by their vowels. Syllables in Category A contained vowels -e/-i, i.e. {be, bi, de, di, ge, gi}, whereas syllables in Category B contained vowels -o/-u, i.e. {po, pu, to, tu, ko, ku}.

[1] In the figure of Fitch and Hauser (2004), there were no indices for $(AB)^n$ or A^nB^n, because any A could be related with any B. Contrarily, in B&F, de Vries et al. (2008) and the current study, indices were used to indicate dependencies between specific A's and B's.

Each syllable in Category A was associated with its counterpart in Category B according to the onset consonants. For instance, any A_x could be related with any B_x. There were two possible syllables for A_x, i.e. “be” or “bi” and two for B_x, “po” and “pu”. Therefore, the associated pairs were {be/bi-po/pu}, {de/di-to/tu} and {ge/gi-ko/ku}. Syllable strings were made out of two, four, and six paired-syllables following the hierarchical center-embedded rule A^nB^n. The resulting grammar G is schematically displayed in Figure 2. Frequencies of syllable occurrence were controlled for.

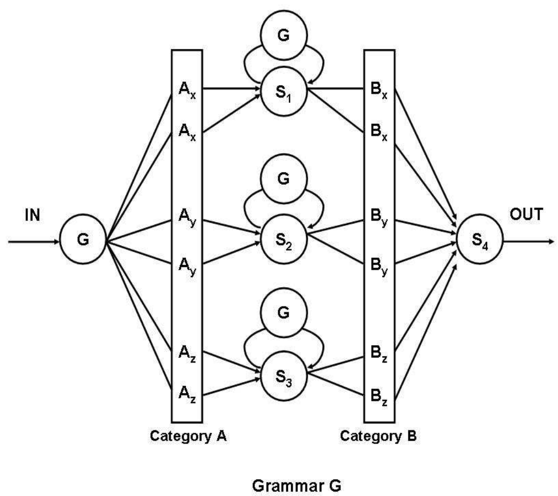

Figure 2. Grammar G, a recursive A^nB^n center-embedded structure. A_x={be, bi}; A_y={de, di}; A_z={ge, gi}; B_x={po, pu}; B_y={to, tu}; B_z={ko, ku}. Examples of strings generated by G are: bi pu (0-LoE), de ge ko tu (1-LoE), be di ge ku to po (2-LoE). “G” in the loops at states S_1, S_2 and S_3 refer to Grammar G, indicating that a center-embedded clause can legally be inserted at that state.

There were 12 blocks in total. Each block consisted of two phases, i.e. learning and testing. All learning and testing blocks together contained 144 strings respectively. Each learning phase was made of 12 syllable strings. After each learning phase, a testing phase followed with 12 novel syllable strings, of which six syllable strings were grammatical and six were ungrammatical.

Note that grammar G generates 12 unique 0-LoE items, 12^2 = 144 unique 1-LoE items, and 144 x 12 =1728 unique 2-LoE items. The 12 unique 0-LoE items were presented four times each (48 in total). Forty-eight 1-LoE items were sampled from the 144 possible ones and presented each once, without repetition. Finally, 48 2-LoE items were sampled from the 1728 unique exemplars of G, and not repeated. In this manner, the differential frequencies of repetitions of ‘simple’ vs. ‘complex’ exemplars of a grammar were represented in the input.

Participants were randomly assigned to one of the two experimental groups: the starting small (henceforth SS) group or the random group. All participants were exposed to the same items, i.e., syllable strings, generated by the grammar G in Figure 2. The learning items for the SS group were ordered by their levels of embedding (LoE). In the first four blocks of the SS group, only 0-LoE items were presented during learning. The following four blocks displayed 1-LoE items only. In the last four blocks, 2-LoE items were presented. In this manner, the learning phase was comprised of three consecutive stages, each of which contained four blocks. The ordering of syllable strings within one block was counterbalanced over participants. The random group would see exactly the same set of strings but in a random order. In the random group, each block and each stage contained an equal number of each LoE-category items.

Both groups were presented the same blocks of test items, in the same order. The grammatical test items were novel items with 0-, 1-, or 2-LoE. Ungrammatical items were made by mismatching syllables from Category A and their counterparts from Category B. To control for as many confounding surface cues as possible, the violations satisfied a number of demands. For two-syllable strings, violations appeared necessarily in the second position (e.g. $A_x\underline{\mathbf{B_y}}$); for four-syllable strings, violations appeared in the fourth position (e.g. $A_xA_yB_y\underline{\mathbf{B_z}}$, $A_xA_yB_y\underline{\mathbf{B_y}}$); and for six-syllable strings, violations appeared in the fifth or sixth position (e.g., $A_xA_yA_zB_z\underline{\mathbf{B_z}}B_x$, $A_xA_yA_zB_z\underline{\mathbf{B_x}}B_x$, $A_xA_yA_zB_zB_y\underline{\mathbf{B_z}}$, $A_xA_yA_zB_zB_y\underline{\mathbf{B_y}}$). In this way, no adjacent AB violations (illegal bigrams) were presented except for the two-syllable test items, in which violations were necessarily an illegal bigram, i.e. an illegal AB pair. Secondly, in contrast to B&F, no adjacent repetition of syllables appeared in the same sequence. All grammatical and ungrammatical test strings had an equal number of A’s and B’s. Hence, violations were not detectable by matching the number of A’s to the number of B’s. Thirdly, only one illegal pair was allowed in the same string to keep the global level of difficulty constant for each test item. As a result of these constrains, three types of violation were generated: first, violations of type $A_xA_yA_xB_xB_y\underline{\mathbf{B_y}}$ with A’s and B’s from the same subsets but not equally distributed; second, violations of type $A_yA_yB_y\underline{\mathbf{B_z}}$, or $A_xA_yA_yB_yB_y\underline{\mathbf{B_z}}$ with one B that could not be paired with any of the A’s; third, violations of type $A_xA_yB_y\underline{\mathbf{B_y}}$, or $A_xA_yA_zB_zB_y\underline{\mathbf{B_y}}$, with one A missing a B from the same subset. Constructing the violations in this manner, violations detection by superficial heuristics could be largely excluded and categorization performance could be reasonably attributed to knowledge of the hierarchical structure

Procedure. At the beginning of every learning trial, a fixation cross appeared in the center of the screen for 500 ms. Then, each syllable was presented separately for 800 ms, with no interval in-between. Participants were instructed that there was a rule underlying the sequences that they had seen. After presentation of 12 syllable strings, the testing phase followed, in which the sequences appeared in the same fashion. When the last syllable of each test item had disappeared, participants had to press the keyboard buttons indicating “YES” or “NO”. They were required to make a judgment whether the novel syllable string was grammatical or not, according to the rule underlying the sequences in the learning phase. After each judgment, appropriate feedback was given for 500 ms as B& F and de Vries et al. (2008) did. Approximately, the task took about 30 minutes.

Results and analysis

First, we estimated the mean proportion of “YES” responses to all test items. There was a small response bias favoring positive responses (M = .53, SE = .01, $p < .01$). Accordingly, d'-values were calculated and used as a measure for sensitivity to grammaticality of the responses, i.e. performance. We conducted an independent-samples t-test on mean d’-values for all test items, to compare performance between these two groups. Overall, the SS group (M = 1.51, SE = .36) highly outperformed the random group, M = .08, SE = .05, t (26) = 3.94, p = .001. Moreover, as indicated by a one-sample t-test comparing mean performance with chance level in both groups, only the SS group performed above chance, t (13) = 4.21, p = .001.

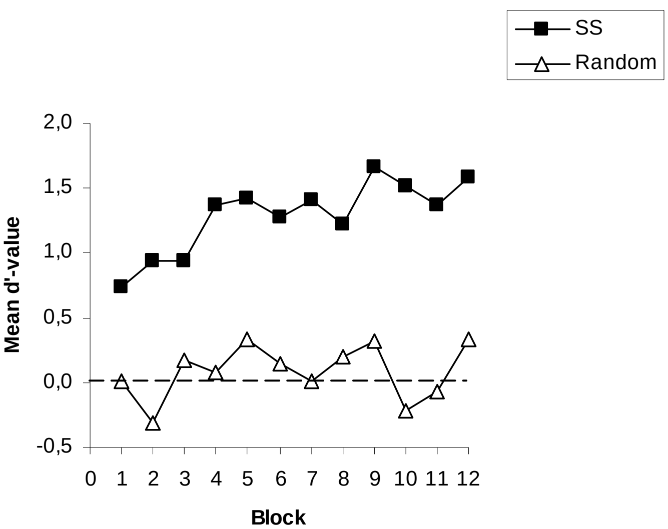

Figure 3. Experiment 1: Mean d’-values for all blocks in both conditions. Points represent mean d’-values per block. The dotted line represents chance level performance (d’= 0).

To evaluate the development over time, in both learning conditions, we compared performance on the first block (Block 1) with the last block (Block 12) for both groups. For the SS group, mean d’-values in Block 1 was M = .73 (SE = .30) and in Block 12, M = 1.59 (SE = .33). Performance had improved in the last block as compared to the first block as revealed by a t-test for means of paired samples, t (13) = 2.59, $p < .05$. In the random group, however, performance did not improve: in Block 1, M = .01 (SE = .21); in Block 12, M = .33, (SE = .29), t (13) = -.98, n.s.. Although in Block 1 the SS group performed slightly better than the random group in the same block, this difference was not significant, t (26) = 1.98, n.s.. However, in the last block, the SS group clearly outscored the random group, t (26) = 2.87, $p < .01$. In Figure 3, mean d’-values are displayed for all blocks in both conditions, showing learning in the SS group over time, but no learning for the random group.

To explore more in detail how the center-embedding recursive principle was learned, we looked into performance on test items with different LoEs. Performance on different types of test items (0-, 1-, and 2-LoE) was compared between conditions, at several stages of exposure. For this analysis, exposure was divided into three stages (Stage 1 consisted of Block 1-4, Stage 2 consisted of Block 5-8, and Stage 3 consisted of Block 9-12.). For the SS group, the stages of training reflected increasing LoE in the stimuli (Stage 1 comprised 0-LoE learning items only; Stage 2, 1-LoE items only; Stage 3, 2-LoE items only). In the random group, all LoEs were presented in the learning phases of every stage. To test the development of performance over time for test items with increasing LoEs, we carried out an ANOVA, with stage and LoE as within-subject factors, and condition as between-subject factor. The LoE × Stage × Condition interaction was significant, F (4, 104) = 2.94, $p < .05$, indicating that performance for different LoE test items developed differently in each learning condition.

Subsequently, an ANOVA was conducted with LoE as the within-subject factor and d’ performance as the dependent variable, for each group separately. For the SS group, a main effect of LoE was found, F (2, 26) = 10.86, $p < .001$. As can be seen in Figure 4, learning for test items with 0-LoE was quite high (M =1.89, SE =.39) and significantly better than learning for items with higher LoE in the SS group, M = 1.45, SE = .37, t (13) = 3.14, $p < .01$ and M = 1.29, SE = .33, t (13) = 4.19, p = .001 for 1-LoE and 2-LoE, respectively. This indicates that participants acquired fundamentally solid knowledge of the adjacent-dependencies of grammar G, under the SS learning condition. Violations of 0-LoE items were observed to be easier to detect than 1-LoE and 2-LoE ones because of their illegal adjacent-dependencies, i.e. bigrams. However, this advantage was only beneficial for the SS group, presented with all 0-LoE training items which clustered in the first stage of exposure. In the random group, participants did not perform differently for various LoE test items. No effect of LoE was found, F (2, 26) = 1.31, n.s. Chance level performance was observed in the random group for all types of test items.

Furthermore, our data revealed a main effect of stage in the SS group only: Performance on all types of test items improved along with exposure to increasing LoE items, F (2, 26) = 3.57, $p < .05$. The curves of 1-LoE and 2-LoE test items evolved equally (see Figure 4), suggesting that the center-embedding rule was learned and recognized equally well for items with one and two recursive loops. In contrast, no main effect of Stage was found for the random group, F (2, 26) = .87, n.s.: Performance was low at the beginning and did not increase significantly over time.

Finally, for the SS group, we compared participants' accuracy on all types of violations with an ANOVA, with Type of Violation as a within subjects factor, to test whether some surface characteristic of the test items (even after careful control for confounding surface cues) might have affected performance. No effect of Type of Violations on accuracy was found, F (2, 26) = .151, n.s.. This suggests that participants performed equally well over different types of violations, indicating knowledge of the hierarchical center-embedded structure learned in the SS procedure.

Hence, our findings indicate that center-embedded structures in an AGL could be learned through the SS procedure, but not in the random procedure, in accordance with our hypothesis. Moreover, an incremental exposure to the input in accordance with increasing applications of the recursive rule, correlated with a synchronic improvement in performance. Participants learned the center-embedding principle along with exposure to increasingly more complex exemplars. Robust knowledge of the 0-LoE exemplars could be shown in the SS group only, suggesting that this knowledge was a prerequisite for learning the embedding principle. Furthermore, the SS group did not judge less accurately test items with 2-LoE than items with 1-LoE, suggesting that the recursive rule was learned and recognized equally easily for 1- and 2-LoE strings.

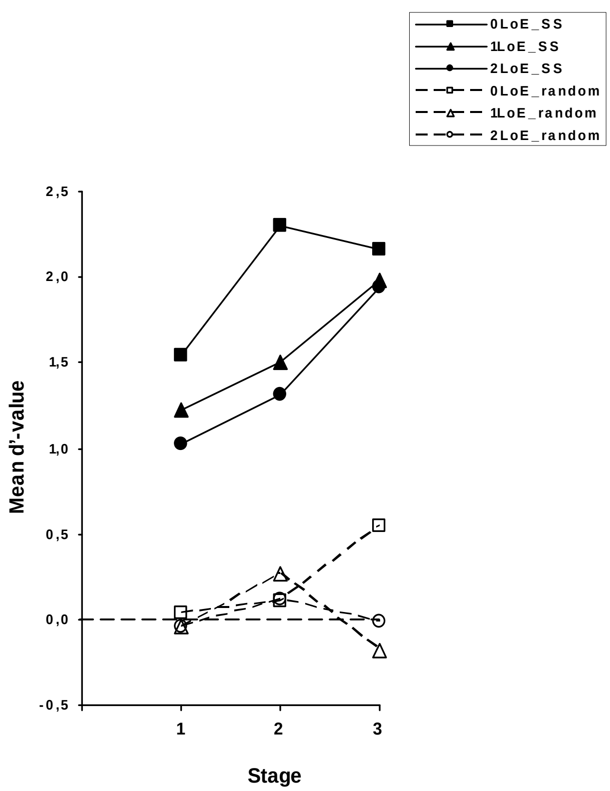

Figure 4. Experiment 1: Mean d'-values for 0-, 1-, and 2-LoE test items at different stages. Points represent mean d'-values of performance per stage. The dotted line represents chance level performance (d'= 0).

Discussion

We observed a 'starting small' effect highly facilitating learning a center-embedded recursive grammar. When participants were presented with a randomized input, there was no learning of the underlying hierarchical rule. Moreover, in our training materials as opposed to the materials presented in similar studies using the same unique training exemplars, simple stimuli were presented more frequently than complex ones, possibly contributing to the dramatic learning effect of the starting small ordering found in our study. In the AGL program, it is still under debate whether performance in learning reflects real knowledge of the abstract grammar, or local pattern learning, recognition of repetitions and other surface heuristics (Poletiek & Van Schijndel, 2009). In the present experimental set up, the violations inserted in the test materials were controlled as much as possible for surface cues that would make them easy to detect without knowledge of the structure. Though the use of cues can not be excluded definitely, our data make a strong case for the learnability of a center-embedded structure provided training with a staged input, and sufficient exposure to basic exemplars without embedded clauses.

Our training stimulus set may be regarded as a representation of the child's natural linguistic environment. The input contains not only a huge number of simple adjacent-dependencies (0-LoE items) produced by the grammar, but they were also presented repeatedly. From the complex items produced (1-, and 2-LoE items), a proportionally smaller sample was presented, and no repetitions occurred. This environment with both growing data and repetitions of basic patterns reflects, as we claim, the natural linguistic environment. In the SS group, due to an intensive training with only 0-LoE items, participants might become familiar with the most basic adjacent-dependencies, which might have provided them with a solid foundation for further induction of the recursive operation. Furthermore, the staged ordering helped participants gradually identify the recursive rule and the connections between long-distance dependencies. By contrast, previous studies failing to find recursion learning, trained participants with the whole corpus randomly presented as an entirety, and no 0-LoE items (de Vries et al., 2008). The two factors investigated here seem therefore to play a crucial role in learning complex recursive rules.

As Elman (1993) indicated humans' most amazing achievement in languages occurs in childhood. In this period, children are exposed to continuously repeated simple structures. Furthermore, the *less is more* proposal that the limited cognitive capacity of children is beneficial to language learning (Newport, 1988, 1990) is consistent with the starting small environmental factor found in our experiment.

In sum, the present study reveals crucial roles for staged input and for solid primary knowledge of the basically simple structures in learning a center-embedded recursive structure by induction. The picture raised is that preliminary simple associative learning mechanisms such as adjacent-dependencies learning might prepare learners for subsequent processing of gradually encountered more complex and more distant dependencies. Our research suggests that the old puzzle of the inductive learnability of recursive structures might benefit from a shift of focus from the formal characteristics of the structure to the stimulus environment and how this environment is nicely shaped to fulfill the needs of the language learner.

References

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes*, 249-262.

Bahlmann, J., & Friederici, A.D. (2006). FMRI investigation of the processing of simple linear and embedded hierarchical structures: An artificial grammar task. *Journal of Cognitive Neuroscience*, Annual Meeting Supplement, p.126.

Bahlmann, J., Schubotz, R.I., & Friederici, A.D. (2008). Hierarchical artificial grammar processing engages Broca's area. *NeuroImage, 42*, 525-534.

Bellinger, D. (1980). Consistency in the pattern of change in mothers' speech: Some discriminant analyses. *Journal of Child Language, 7*, 469-487.

Chomsky, N., & G. Miller. (1963). Introduction to the Formal Analysis of Natural Languages. *Handbook of mathematical Psychology*, R. Luce et al., eds. New York: John Wiley.

Cochran, B.P., McDonald, J.L., & Parault, S.J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language, 41*, 30-58.

De Vries, M.H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition, 107*, 763-774.

Elman, J.L. (1991). Incremental learning, or the importance of starting small. *Technical Report, 9101*. Center for Research in Language, University of California at San Diego.

Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71-99.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*, 377-380.

Garnica, O.K. (1977). Some prosodic and paralinguistic features of speech to young children. In C.E. Snow & C.A. Ferguson. *Talking to children: Language input and acquisition,* 63-68. Cambridge, England: Cambridge University Press.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440,* 1204-1207.

Hauser, M.D., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*, 1569-1579.

Hudson, R. (1996). The difficulty of (so-called) self-embedded structures. *UCL Working Papers in Linguistics 8*, 283-314.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General, 126*, 278-287.

Kersten, A.W., & Earles, J.L. (2001). Less really is more for adults learning a miniature artificial language. *Journal of Memory and Language, 44*, 25-273.

Newmeyer, F. (1988). Extensions and implications of linguistic theory: an overview. In F. Newmeyer, (ed.) *Linguistics: The Cambridge Survey 2. Linguistic Theory: Extensions and Implications*. Cambridge: Cambridge University Press. 1-14.

Newport, E.L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. *Language Sciences, 10*, 147-172.

Newport, E.L. (1990). Maturational constraints on language learning. *Cognitive Science, 14,* 11-28.

Perruchet, P., & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from non-human primates? *Psychonomic Bulletin and Review, 12 (2)*, 307-313.

Philips, J.R. (1973). Syntax and vocabulary of mothers' speech to young children: age and sex comparisons. *Child Development, 44*, 182-185.

Pine, J.M. (1994). The language of primary caregivers. In C. Gallaway & B.J. Richards (Eds.), *Input and interaction in language acquisition*. Cambridge, UK: Cambridge University Press.

Poletiek, F.H., & Chater, N. (2006). Grammar induction benefits from representative sampling. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Poletiek, F.H., & Van Schijndel, T.J.P. (2009). Stimulus set size and statistical coverage of the grammar in artificial grammar learning. *Psychonomic Bulletin & Review, 16 (6)*, 1058-1064.

Rohde, D.L.T., & Plaut, D.C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition, 72*, 67-109.

Snow, C.E. (1972). Mothers' speech to children learning language. *Child Development, 43*, 549-65.

Vasishth, S. (2001). An empirical evaluation of sentence processing models: Center embeddings in Hindi. In M. Daniels, D. Dowty, A. Feldman, & V. Metcalf (Eds.), *OSUWPL, Volume 56*, 159-181, Ohio State University.