Cognition 118 (2011) 265–273

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures

Jun Lai [*], Fenna H. Poletiek

Leiden University, Cognitive Psychology, 2300 RB Leiden, The Netherlands

ARTICLE INFO

Article history:
Received 25 January 2010
Revised 9 November 2010
Accepted 16 November 2010
Available online 9 December 2010

Keywords:
Starting small
Hierarchical structures
Center-embedding
Artificial grammar learning

ABSTRACT

A theoretical debate in artificial grammar learning (AGL) regards the learnability of hierarchical structures. Recent studies using an A^nB^n grammar draw conflicting conclusions (Bahlmann & Friederici, 2006; De Vries, Monaghan, Knecht, & Zwitserlood, 2008). We argue that 2 conditions crucially affect learning A^nB^n structures: sufficient exposure to zero-level-of-embedding (0-LoE) exemplars and a staged-input. In 2 AGL experiments, learning was observed only when the training set was staged and contained 0-LoE exemplars. Our results might help understanding how natural complex structures are learned from exemplars.

1. Introduction

Recursion, as in sentences with hierarchically built up center-embeddings, is regarded as a crucial property of human language (Hauser, Chomsky, & Fitch, 2002). However, sentences with several levels of embedding (LoE) are difficult to process, even for native speakers (Bach, Brown, & Marslen-Wilson, 1986; Hudson, 1996; Newmeyer, 1988; Vasishth, 2001). *The rat the cat the dog chased killed ate the malt* (Chomsky & Miller, 1963, pp. 286–287) is a typical center-embedded sentence incorporating two sub-clauses. The dependencies between related constituents become harder to associate as more clauses are inserted, not least since the counterparts get further away from each other.

Recursion refers to structures that are self-referential, and infinitely productive. In center-embedded structures, inserting a grammatical sentence within another generates a new grammatical sentence. This operation can be applied infinitely, generating numerous output sentences. Since Hauser et al. (2002) stressed the crucial importance of recursive rules in natural languages, a renewed interest has risen concerning the learnability of recursion. Most studies use the artificial grammar learning (AGL) paradigm (Corballis, 2007; Gentner, Fenn, Margoliash, & Nusbaum, 2006; Perruchet & Rey, 2005). In particular, Fitch and Hauser (2004) proposed that the ability of mastering hierarchical structures was critical to distinguish human and nonhuman primates. They argued that humans could grasp hierarchical structures generated by an A^nB^n grammar (see Fig. 1), while tamarins were incapable. Moreover, Bahlmann and Friederici (2006) (henceforth B&F) and Bahlmann, Schubotz, and Friederici (2008) carried out an fMRI study to probe into the neural basis of processing long-distance dependencies. Significantly greater blood flow was observed in Broca's area during processing of hierarchical-dependency A^nB^n compared to adjacent-dependency $(AB)^n$.

However, as indicated by Perruchet and Rey (2005), the mapping of A-to-B is the essential characteristic of hierarchical center-embedding recursion. At each LoE, this mapping has to be legal according to the grammar.[1] Therefore, Fitch and Hauser (2004), whose grammar did not specify

---

[*] Corresponding author. Tel.: +31 715276752; fax: +31 715273619.
E-mail address: laij@fsw.leidenuniv.nl (J. Lai).

[1] For instance, $A_1A_2A_3B_3B_2B_1$ is grammatical, whereas $A_1A_2A_3B_1B_2B_3$ is not.

0010-0277 © 2010 Elsevier B.V. Open access under the Elsevier OA license.
doi:10.1016/j.cognition.2010.11.011

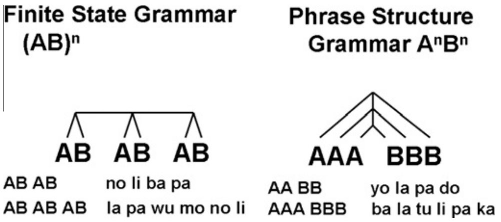

Fig. 1. Structures of finite state grammar $(AB)^n$ and phrase structure grammar A^nB^n used by Fitch and Hauser (2004). Examples of Category A words are: no, ba, la, wu and Category B words are: li, pa, ka, do.

such mapping, could not demonstrate knowledge of center-embeddings in their experiment. The same problem applies for B&F. Though B&F did use a grammar specifying a hierarchical A–B mapping, their test materials were incapable of detecting center-embedded structure learning. When the test materials were controlled, participants failed to learn, as showed by De Vries, Monaghan, Knecht, and Zwitserlood (2008), who argued that performance in B&F is based on superficial heuristics, like counting the A's and B's, or repetition-monitoring, instead of learning the center-embedded principle.[2]

Previous research has mainly focused on the cognitive learnability of center-embedded structures, rather than on features of the environmental input. Here, we propose two crucial but previously poorly attended environmental factors: One is the organization of the input by stages (*starting small*, henceforth SS) and the second is sufficient exposure to the grammar's basic adjacent-dependencies in the earliest stage of learning. The purpose of the present research is to explore the impact of these two closely-related conditions on learning center-embeddings.

Considering natural language learning, child-directed speech globally satisfies these conditions, as it has, in the earliest stage, short linguistic constituents, simple grammatical constructions, and little syntactical variability (Pine, 1994; Tomasello, 2003). As children grow, child-directed speech develops gradually into more mature speech types (Bellinger, 1980; Garnica, 1977). Hence, the input on which the learning process operates, does not come in a random order. Therefore, if we can demonstrate experimentally the facilitation effect of a growing environmental input, and early exposure to zero-level-of-embedding (0-LoE) exemplars, this result might help understanding the role of the environment in complex natural language learning.

The notion of SS was first raised by Elman (1991, 1993). He trained a connectionist network to parse complex structures which contained embedded subordinates. The network succeeded only if provided with a staged-input, but not after exposure to the entire input as a whole. Subsequent studies yielded mixed results, though. Some findings are consistent with Elman's effect (Conway, Ellefson, & Christiansen, 2003; Kersten & Earles, 2001; Krueger & Dayan, 2009; Newport, 1988, 1990; Plunkett & Marchman, 1990). However, other research reported no effect of staged-input (Fletcher, Maybery, & Bennett, 2000; Ludden & Gupta, 2000; Rohde & Plaut, 1999).

In the current study, two AGL experiments were carried out using similar materials as B&F and de Vries et al. (2008). In Experiment 1, we compared learning with a staged-input and a random input. Both learning sets contained 0-LoE exemplars. In Experiment 2, 0-LoE learning items were omitted.

2. Experiment 1

All participants were exposed to the same strings, generated by grammar $\underline{G}$ (Fig. 2). In the SS condition, syllable strings were presented progressively according to their LoE.[3] In the random condition, exactly the same set was presented randomly. We hypothesize that the SS group outperforms the random group.

2.1. Method

2.1.1. Participants

Twenty-eight students (20 female), from Leiden University participated. All were native Dutch speakers.

2.1.2. Materials and design

There were two sets of syllables, categorized by their vowels. Category A contained -e/-i, i.e. {be, bi, de, di, ge, gi}, whereas Category B contained -o/-u, i.e. {po, pu, to, tu, ko, ku} (see Appendix). Each A-syllable was connected with its counterparts in Category B according to another cue: their consonants, i.e. {be/bi-po/pu}, {de/di-to/tu} and {ge/gi-ko/ku}. Strings were constructed with two, four, or six paired-syllables following the A^nB^n rule. Frequencies of syllable occurrence were controlled for.

The experiment consisted of 12 blocks, with a learning phase and a testing phase each. Twelve strings were presented in each learning phase, and 12 novel strings in each testing phase, of which six were grammatical and six ungrammatical. Both groups were presented the same test strings with 0-, 1-, or 2-LoE. Ungrammatical strings were created by mismatching A-syllables with B-syllables. For two-syllable strings, violations appeared necessarily in the second position ($A_1\underline{\mathbf{B_2}}$); for four-syllable strings, in the fourth position ($A_1A_2B_2\underline{\mathbf{B_3}}$); and for six-syllable strings, in the fifth or sixth position ($A_1A_2A_3B_3\underline{\mathbf{B_4}}B_1$, $A_1A_2A_3B_3B_2\underline{\mathbf{B_4}}$). For instance, the violation B_4 in $A_1A_2A_3B_3B_2\underline{\mathbf{B_4}}$ means that the last B mismatches any A in this sequence. In this manner, no adjacent AB violations in the middle of a string could occur, except, necessarily, for two-syllable test strings. Moreover, in contrast to B&F, no repetition of exactly the same syllable appeared in the same sequence, and all test strings had an equal number of A's and B's.

---

[2] Indeed, in B&F, violations were *replacement violations* (e.g. $A_1A_2A_3B_3A_2B_1$) and *concatenation violations* (e.g.$A_1A_2B_2B_3$). Contrarily, de Vries et al. (2008) tested two other types: *scrambled* (e.g. $A_1A_2A_3B_1B_3B_2$) and *scrambled + repetition* ($A_1A_2A_3B_1B_3B_1$). Their participants could detect the scrambled + repetition violations, but not the scrambled ones.

[3] For the SS group, in the first four blocks, only 0-LoE learning items were presented. The following four blocks displayed 1-LoE items only. In the last four, 2-LoE items were presented. The ordering of strings within one block was counterbalanced over participants.

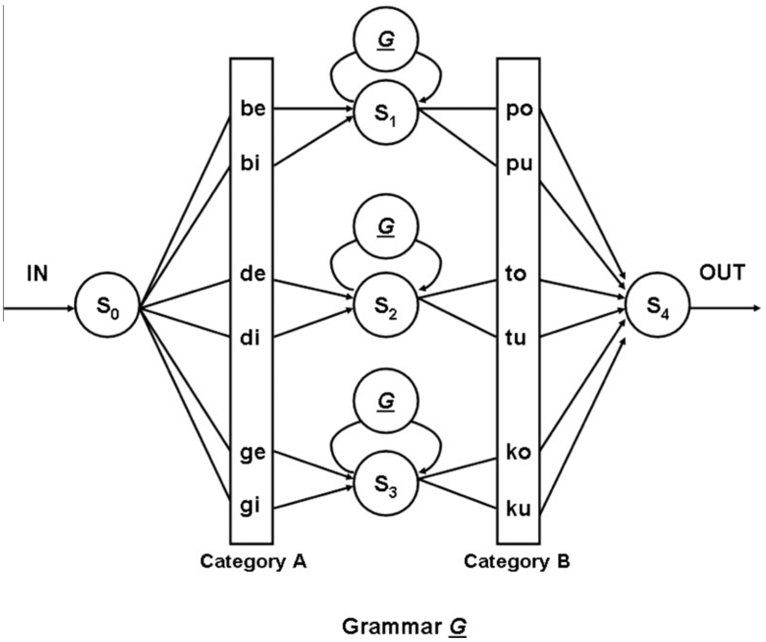

Fig. 2. Grammar G, an A^nB^n center-embedded structure. The grammar starts from S_0 and follows one of all possible paths until S_4. "G" in the loops at states S_1, S_2 and S_3 refer to the self-referential rule, indicating that a center-embedded clause can legally be inserted at that specific state. 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).

As a result, violations could not easily be detected on the basis of surface heuristics or bigram violations.

2.1.3. Procedure

Participants were informed that they would see strings satisfying a sequential rule. Each learning trial started with a fixation cross (500 ms). Then, each syllable was presented separately for 800 ms, with no interval in-between.[4] After presentation of 12 strings, a testing phase followed. When the last syllable of each test string disappeared, participants had to indicate "YES" or "NO" depending on whether they believed the string satisfied the rule also underlying the learning strings. Feedback was given (500 ms). For ease of comparison with findings by B&F and de Vries et al. (2008), their explicit procedure was also applied in the current study. The task took 30 min approximately.

2.2. Results and discussion

A t-test on mean d'-values[5] revealed that, overall, the SS group, $d' = 1.51$(73% correct), highly outperformed the random group, $d' = .08$ (52% correct), $t\,(26) = 3.94$, $p = .001$. Only the SS group performed above chance, $t\,(13) = 4.21$, $p = .001$.

Moreover, the SS group improved in Block 12, $d'_{12} = 1.59$ (78% correct), compared to Block 1, $d'_1 = .73$ (63% correct), $t\,(13) = 2.59$, $p < .05$. In the random group, however, performance did not improve over time: $d'_1 = .01$ (50% correct), $d'_{12} = .33$ (56% correct), $t\,(13) = -.98$, n.s.. Although in Block 1 the SS group performed slightly better than the random group, 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 Fig. 3, mean d'-values are displayed for all blocks, showing learning in the SS group over time, but not for the random group.

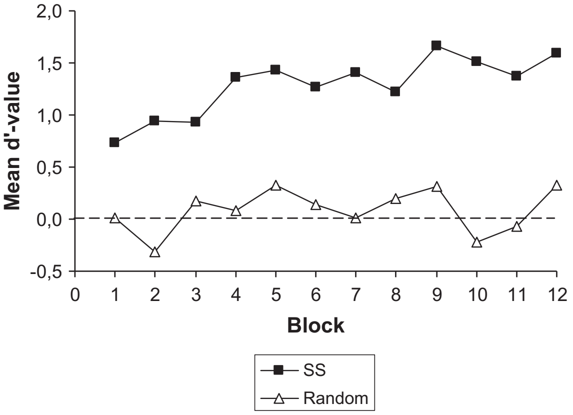

Fig. 3. Experiment 1: mean d'-values for all blocks in both conditions. Points represent mean d'-values per block. The dotted line represents chance level performance ($d' = 0$).

---

[4] With this manipulation, we tried to simulate the situation of natural language processing maximally, in the laboratory environment.

[5] Due to a small response bias favoring positive responses ($M = .53$, $SE = .01$, $p < .01$), d'-values were applied as a measure for sensitivity to grammaticality of the responses.

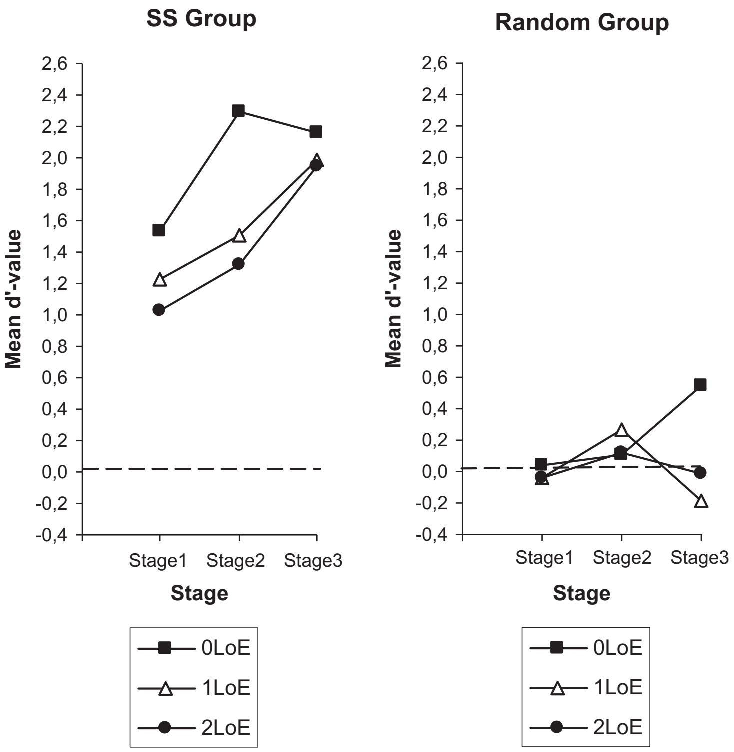

Fig. 4. Experiment 1: mean d'-values for 0-, 1-, and 2-LoE test items at different stages. Points represent mean d'-values of performance per stage. The dotted line represents chance level performance ($d' = 0$).

In addition, performance on different types of test items (0-, 1-, and 2-LoE) was compared at several stages of exposure.[6] An ANOVA, with LoE and stage as within-subject factors and condition as between-subject factor showed main effects of LoE, $F(2, 52) = 9.00$, $p < .001$; of stage, $F(2, 52) = 3.92$, $p = .04$; and of condition, $F(1, 26) = 17.30$, $p < .001$. The LoE × Stage × Condition interaction was significant, $F(4, 104) = 2.94$, $p = .02$, indicating, that performance on various LoE test items developed differently under each condition.

Subsequently, for each group we conducted an ANOVA with LoE and stage as within-subject factors. Under the SS condition, there were main effects of LoE, $F(2, 26) = 10.86$, $p < .001$, and of stage, $F(2, 26) = 3.57$, $p < .05$. Performance for 0-LoE items (see Fig. 4), $d' = 1.89$ (77% correct), was significantly better than 1-LoE, $d' = 1.45$ (72% correct), $t\ (13) = 3.14$, $p < .01$ and 2-LoE, $d' = 1.29$ (70% correct), $t\ (13) = 4.19$, $p = .001$, respectively. However, in the random group, chance level performance was observed for all types of test items. There was no effect of LoE, $F(2, 26) = 1.31$, n.s., neither of stage, $F(2, 26) = .87$, n.s..

In sum, our findings revealed learning of center-embedded structures in the SS procedure, but not in the random procedure. Moreover, gradual exposure to the staged-input, co-occurred with a synchronic improvement in performance. Strikingly, at the end of the first stage, when the SS group had been exposed to 0-LoE only, they performed better ($d' = 1.36$, 74% correct) than the random group ($d' = .08$, 52% correct), who did see higher-than-0-LoE learning items, $t\ (26) = 3.42$, $p < .005$.

To test further whether performance in the SS group could rely on other strategies, even after careful control for possible confounding surface cues (de Vries et al., 2008) in the test materials, we looked for complex surface calculations that might have underlain detection of particular violations. We subsequently classified these violations according to the surface rule that could possibly have been used to detect them.[7] We then could predict that if knowledge of the center-embedded principle was the basis of response, equal performance on all types of violations, should be found. If, alternately, participants relied on surface cues, different performance may be expected for types of violations detectable with different cues or calculations. In particular, lower performance can be expected as more

---

[6] Stage 1 consisted of Block 1-4, Stage 2 consisted of Block 5-8, and Stage 3 consisted of Block 9-12 (see Appendix). Especially for the SS group, Stage 1 comprised 0-LoE learning items only; Stage 2, 1-LoE items only; Stage 3, 2-LoE items only; whereas for the random group, various LoEs were presented in all learning stages.

[7] Three types of violations were distinguished: Type I ($A_1A_2A_1B_1B_2B_2$) violation with A's and B's from the same subsets but not equally distributed for the A's as for the B's; Type II ($A_1A_1B_1B_2$, or $A_1A_2A_2B_2B_2B_3$) with a B that could not be paired with any A; Type III ($A_1A_2B_2B_2$, or $A_1A_2A_3B_3B_2B_2$), with one A missing a B from the same subset. Indices here refer to subsets of syllables within A or B category. Each subset consists of two different syllables.

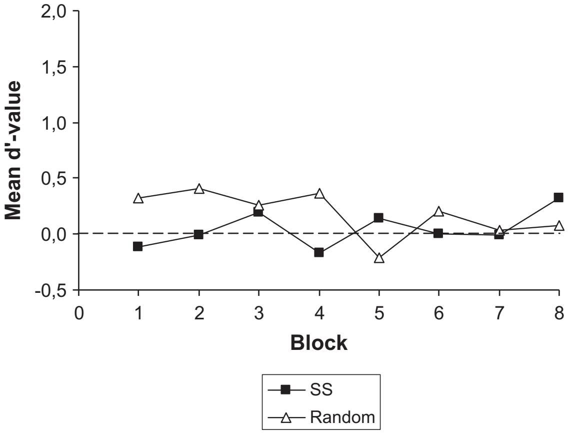

Fig. 5. Experiment 2: mean d'-values for all blocks in both conditions. Points represent mean d'-values of performance per block. The dotted line represents chance level performance ($d' = 0$).

complex calculations are needed to detect a violation. We found no effect of type of violation on performance, $F(2, 26) = .15$, n.s.. Participants' performance in the SS group was actually highly similar for all types of violations.[8]

A possible surface heuristic that de Vries et al. (2008) paid attention to, is 'monitoring repetitions'. In our materials, no exact repetitions could occur; though repetitions of syllables within the same A or B subcategory could (for example *bebi*- or -*totu* could occur as part of a sequence). However, this type of repetitions was independent of grammaticality of the sequence in our test materials: subset repetitions both occurred in grammatical (e.g., $A_1A_1B_1B_1$) and ungrammatical (e.g., $A_1A_1A_2B_2\underline{\mathbf{B_2}}B_1$) items. Thus, subset repetitions could not be used as a heuristic. Overall, our stimuli and data weaken the possibility that participants used surface rules to perform the grammaticality-judgment task.

Since robust knowledge of 0-LoE exemplars was shown in the SS group only, knowledge of two-syllable sequences might be necessary to grasp the embedding principle. Indeed, primary exposure to adjacent-dependencies was hypothesized to be another crucial factor facilitating learning. We conducted Experiment 2 to verify this hypothesis. We compared again a SS group with a random group, as in Experiment 1, removing all 0-LoE learning items in both conditions.

3. Experiment 2

3.1. Method

3.1.1. Participants

Eighteen students (13 female) from Leiden University participated. None had participated in Experiment 1.

3.1.2. Materials and design

The same materials except 0-LoE learning items were adopted from Experiment 1. Participants were trained with 96 items possessing 1- or 2-LoE (See Appendix). In the learning phase, the SS group was first presented with four blocks of 1-LoE items, and subsequently, with four blocks of 2-LoE items, whereas the random group was presented with the same input randomly.

3.1.3. Procedure

Identical to Experiment 1.

4. Results and discussion

Overall the SS group, $d' = .05$ (51% correct), did not differ from the random group, $d' = .18$ (53% correct), $t(16) = -1.11$, n.s. Both groups performed at chance level. Additionally, for both groups (see Fig. 5), performance did not change between the first and the last blocks, $d'_1 = -.12$ (48% correct), $d'_8 = .32$ (56% correct), $t(8) = 1.50$, n.s. for the SS group, and $d'_1 = .32$ (56% correct), $d'_8 = .08$ (51% correct), $t(8) = .72$, n.s., for the random group. These data indicate that participants could not distinguish grammatical items when no 0-LoE training items presented to them, even in an SS procedure.

5. General discussion

The present research provides insight into two crucial environmental conditions affecting the learnability of a hierarchical center-embedded grammar: first, the effect of an incrementally presented input; second, the importance of exposure to adjacent-structures in the earliest stage of training. Experiment 1 showed that participants performed better on a grammaticality-judgment task after training with an input organized incrementally, according to their LoE. Also, even basic adjacent-dependencies were better learned under SS conditions. The facilitation effect of SS disappeared, as Experiment 2 further revealed, when participants were deprived of exposure to the 0-LoE exemplars. The lack of 0-LoE resulted in an incapability to detect structure, no matter whether the stimuli were presented incrementally or randomly. Clustered exposure to basic adjacencies and a staged-input seem to play crucial roles in learning embedded hierarchical structures.

As previous studies (Christiansen & Dale, 2001; McDonald & Plauche, 1995; Perruchet & Rey, 2005; Poletiek, 2002; Poletiek & Chater, 2006) have suggested, SS may have a better impact when it is assisted by some other cues. The current data indicate that the SS effect can operate if and only if it is combined with sufficient primary exposure to basic adjacent-dependencies of the structure. Especially the striking effect that the SS group outperformed the random group after exposure to 0-LoE only, possibly indicates that once participants were familiarized with the basic associations, they could recognize the associated pairs, even if located in remote positions. Possibly, knowledge of the fundamental adjacent-dependencies serves as a crucial stepping stone in exploring complex hierarchical structures in subsequent stimuli.

The effects of staged-input and early adjacent-dependencies point at the close collaboration between cognition and environment, specifically between an incremental

---

[8] Mean accuracy for test items with violation Type I, II, III were .69, .69, and .67 respectively.

learning mechanism and an incrementally organized input. Thus far, research has mainly focused on the cognitive mechanisms underlying learning complex structures. For instance, a recent fMRI study demonstrated that the activation of the left pars opercularis in processing hierarchical center-embeddings (Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006), also occurs during processing of German (Makuuchi, Bahlmann, Anwander, & Friederici, 2009). And several studies with artificial materials have looked at how long-distance-dependencies are processed (Mintz, 2002, 2003; Onnis, Monaghan, Christiansen, & Chater, 2004).

Our study suggests the importance of a good match between cognition and the environment, in facilitating the learning process of hierarchical center-embeddings. This match may also be at work in natural language learning. Although the procedure used in the present lab study (explicit instructions and visual presentation of the stimuli), deviates from the natural language learning context, the facilitating factors we found may be operating in the natural situation as well. Indeed, the natural environment (child-directed speech) is incremental and the early learning strategy associative. Some other studies on language learning are in line with this analysis. Gómez and Maye (2005) argue that the ability to associate constituents is important in learning natural syntax, especially since center-embedded recursion is one of its main features. A study on American Sign Language (Newport, 1990) showed that early learners outperformed late learners because the former went through a stage in which they were highly familiarized with the simplest constituents. After that, they could become proficient at combining short constituents into more complex entireties.

Our results also generate new questions. For instance, are hierarchical center-embeddings only learnable after some critical level of prior knowledge on adjacent-dependencies has been obtained? Future work has to find out to what criterion learners have to acquire basic knowledge before increasing input complexity can be processed. Moreover, the frequencies of each LoE-category of training items are also interesting for investigation. A current study in our lab suggests that decreasing numbers of exemplars with increasing complexity are needed for learning the underlying system (Poletiek, Chater, & Van den Bos, submitted for publication). Another question is whether different modalities of exposure would affect performance (Conway & Christiansen, 2005). Finally, it is important to find out the limits of the generalizability of the present and similar data for explaining natural processes. A straightforward question is to what extent the huge complexity of natural grammars might invalidate generalization from the experimental noiseless artificial situation.

In sum, the present study reveals crucial roles for a staged-input and solid primary knowledge of the basic structures, in learning by induction a center-embedded structure. From a more general point of view, our research suggests that the old puzzle of the learnability of hierarchical structures might benefit from a shift of focus on the stimulus environment and its fitness to how human learning works and develops over time.

Acknowledgements

This research has been funded by the Netherlands Organization for Scientific Research (NWO). We thank Jeffrey Elman, Pierre Perruchet and two anonymous reviewers for their valuable comments on the present work. We also thank Niels Schiller and Rinus Verdonschot for their valuable suggestions. Correspondence concerning the article should be addressed to: Jun Lai, Cognitive Psychology Department, Leiden University, Pieter de la Court building, P.O. Box 9555, 2300 RB Leiden, the Netherlands. Electronic mail may be sent to: laij@fsw.leidenuniv.nl.

Appendix

Stimuli presented in Experiment 1 (with 0-LoE learning items) and in Experiment 2 (without 0-LoE learning items). In the starting small condition, learning stimuli were presented as displayed; in the random condition, learning stimuli were presented randomly.

	Stage 1			
Phase	Block 1	Block 2	Block 3	Block 4
Learning	bepu	bepu	bepu	bepu
	bepo	bepo	bepo	bepo
	ditu	ditu	ditu	ditu
	dito	dito	dito	dito
	giku	giku	giku	giku
	giko	giko	giko	giko
	bipo	bipo	bipo	bipo
	detu	detu	detu	detu
	bipu	bipu	bipu	bipu
	geko	geko	geko	geko
	deto	deto	deto	deto
	geku	geku	geku	geku

Appendix (*continued*)

	Stage 1			
Phase	Block 1	Block 2	Block 3	Block 4
	Grammatical			
Testing	deto	bipu	bepu	giku
	geku	geko	ditu	dito
	dibeputo	debiputu	debeputo	bigekupo
	biditupo	bedetopo	bebipupo	geditoku
	debigekopotu	degebepukotu	gebeditupuku	dibegikuputo
	gidibeputuko	bibeditupopu	gigebipukuko	bigidetukupu
	Ungrammatical			
	bi**ko**	de**ko**	be**tu**	ge**po**
	ge**pu**	ge**to**	gi**to**	de**pu**
	degiko**ku**	digiko**ku**	degeko**po**	dibepo**ko**
	gebepo**pu**	begiku**to**	bidito**ko**	gibipo**to**
	dibegikupu**po**	digebepo**tu**to	begiditukoku**ku**	dibibepopu**ku**
	bedibipu**ko**po	gedibiputu**po**	digideto**pu**tu	gigeditu**pu**ko

	Stage 2			
Phase	Block 5	Block 6	Block 7	Block 8
Learning	dedituto	debeputu	debepoto	debipotu
	degikoto	degekutu	degekotu	degikuto
	dibiputu	didetuto	didetotu	dibepotu
	digikuto	digikotu	dibiputo	digekuto
	beditupo	beditopo	bedetupo	bedetopu
	begekupo	begekopu	begikopu	bebipopu
	bidetopo	biditopu	bibepupo	bidetupo
	bigekupu	bigekopu	bigikupu	bigikopo
	gedetuku	gedetoko	gedituku	gebepuko
	gegikuko	gebepoku	gebipuko	gegekoku
	giditoku	gidetoku	gidetoko	gidituku
	gigekuko	gibepuko	gibipuku	gibipuko
	Grammatical			
Testing	dito	bipo	bipu	deto
	bepo	geko	detu	geku
	gegikoku	digekutu	gebipuku	degikutu
	debipoto	bigikopu	gidetuko	dibeputu
	gibegekupoko	bidigikotopu	degeditokotu	gedibipotuko
	dibedetupoto	gedegikutuko	begibipokupu	gidebepotuko
	Ungrammatical			
	be**ko**	gi**tu**	bi**tu**	di**po**
	de**ku**	de**po**	gi**pu**	be**to**
	digeko**po**	biditu**ku**	debepo**ku**	degeku**pu**
	giditu**pu**	gebipo**to**	begiku**ko**	gebepu**tu**
	bigidetoko**tu**	begedetu**pu**po	gegidetuko**to**	bibegekopo**ku**
	bigeditu**to**po	gibeditopu**tu**	gedebipu**po**ko	gedibipu**ko**ku

	Stage 3			
Phase	Block 9	Block 10	Block 11	Block 12
Learning	dedibepotuto	dedigikututo	debeditoputu	debegekoputu
	degigekokutu	debiditupoto	degebipukotu	degibepukuto
	dibibepupoto	didebepototu	dibegekoputo	dibidetopotu
	digebipukotu	digigekokutu	dibigikuputo	digedetukoto

(*continued on next page*)

Appendix (*continued*)

	Stage 1			
Phase	Block 1	Block 2	Block 3	Block 4
	bebegekupopu	bebiditupopu	bedegekotupu	bedidetutopo
	begeditokupo	begidetokopo	bebigikupupo	begebipokopu
	bidibeputupo	bibedetopopu	bidegikutopu	bidibepotopu
	bigeditokopo	bigedetukupo	bigigekukopo	bibiditupopu
	gedegikotoku	gedibeputuku	gedigikutoko	gedebipotuko
	gebebipopuku	gebebipupoko	gegebepokuko	gegibipokoku
	gibedetupoko	gidibipotoko	gidegekotuku	gidebipotoko
	gibebipupoko	gibidetopuku	gigeditokoku	gigidetukoku
	Grammatical			
Testing	bepu	ditu	detu	bipo
	giko	giko	bepo	giku
	begekopo	gibepuku	bedetupu	gedetoku
	gebipoko	digikutu	bidetopu	gibepoku
	bedidetutopu	gegibepukuko	bedigekotopu	debegekoputo
	bididetotupu	debibepoputu	bigiditukopu	digebepukutu
	Ungrammatical			
	ge**tu**	bi**to**	di**ko**	di**ku**
	di**pu**	be**ku**	bi**ku**	gi**po**
	geditu**po**	dibipo**pu**	beditu**to**	bibepo**tu**
	begiko**ku**	begeku**tu**	gibepo**tu**	gedetu**pu**
	bedegekotu**to**	bidibepo**ko**pu	gidideto**po**ko	debigeku**to**tu
	dibegeku**ko**tu	degibipuko**po**	gedegikuto**tu**	debiditupu**ku**

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. Annual meeting supplement. *Journal of Cognitive Neuroscience*, 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., & Miller, G. (1963). Introduction to the formal analysis of natural languages. In R. Luce et al., (Eds.), *Handbook of mathematical psychology* (pp. 286–287). New York: John Wiley.

Christiansen, M. H., & Dale, R. A. C. (2001). Integrating distributional, prosodic and phonological information in a connectionist model of language acquisition. In *Proceedings of the 23rd annual conference of the cognitive science society* (pp. 220–225). Mahwah, NJ: Lawrence Erlbaum.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 24–39.

Conway, C. M., Ellefson, M. R., & Christiansen, M. H. (2003). When less is less and when less is more: Starting small with staged input. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual conference of the cognitive science society* (pp. 270–275). Mahwah, NJ: Erlbaum.

Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science, 31*, 697–704.

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.

Fletcher, J., Maybery, M. T., & Bennett, S. (2000). Implicit learning differences: A question of development level? *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*, 246–252.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, 103*, 2458–2463.

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

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

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*(2), 183–206.

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.

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.

Krueger, K. A., & Dayan, P. (2009). Flexible shaping: How learning in small steps helps. *Cognition, 110*, 380–394.

Ludden, D., & Gupta, P. (2000). Zen in the art of language acquisition: Statistical learning and the less is more hypotheses. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd annual conference of the cognitive science society* (pp. 812–817). Mahwah, NJ: Lawrence Erlbaum.

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences of the USA, 106*, 8362–8367.

McDonald, J. L., & Plauche, M. (1995). Single and correlated cues in an artificial language learning paradigm. *Language and Speech, 38*(3), 223–236.

Mintz, T. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*, 678–686.

Mintz, T. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91–117.

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* (pp. 1–14). Cambridge: Cambridge University Press.

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.

Onnis, L., Monaghan, P., Christiansen, M. H., & Chater, N. (2004). Variability is the spice of learning, and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. In *Proceedings of the 26th annual conference of the cognitive science society* (pp. 1047–1052). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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.

Pine, J. M. (1994). The language of primary caregivers. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 109–149). Cambridge, UK: Cambridge University Press.

Plunkett, K., & Marchman, V. (1990). *From rote learning to system building*. Center for Research in Language. TR 9020. San Diego: University of California.

Poletiek, F. H. (2002). Learning recursion in an artificial grammar learning task. *Acta Psychologica, 111*, 323–335.

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* (pp. 1968–1973). Mahwah, NJ: Lawrence Erlbaum.

Poletiek, F. H., Chater, N., & Van den Bos, E. (submitted for publication). The influence of stimulus frequencies in artificial grammar learning. Evidence for deep and surface statistical learning.

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.

Tomasello, M. (2003). *Constructing a language: A usage based theory of language acquisition*. Cambridge, MA: Harvard University Press.

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* (Vol. 56, pp. 159–181). Ohio State University.