Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pecp21

How “small” is “starting small” for learning hierarchical centre-embedded structures?

Jun Lai [a b] & Fenna H. Poletiek [a b]

[a] Leiden Institute for Brain and Cognition , Leiden University , Leiden , The Netherlands

[b] Institute of Psychology , Leiden University , Leiden , The Netherlands

Published online: 03 Apr 2013.

Journal of Cognitive Psychology, 2013
Vol. 25, No. 4, 423–435, http://dx.doi.org/10.1080/20445911.2013.779247

Routledge
Taylor & Francis Group

How "small" is "starting small" for learning hierarchical centre-embedded structures?

Jun Lai[1,2] **and Fenna H. Poletiek**[1,2]

[1]Leiden Institute for Brain and Cognition, Leiden University, Leiden, The Netherlands
[2]Institute of Psychology, Leiden University, Leiden, The Netherlands

Hierarchical centre-embedded structures pose a large difficulty for language learners due to their complexity. A recent artificial grammar learning study (Lai & Poletiek, 2011) demonstrated a starting-small (SS) effect, i.e., staged-input and sufficient exposure to 0-level-of-embedding exemplars were the critical conditions in learning A^nB^n structures. The current study aims to test: (1) a more sophisticated type of SS (a gradually rather than discretely growing input), and (2) the frequency distribution of the input. The results indicate that SS optimally works under other conditional cues, such as a skewed frequency distribution with simple stimuli being more numerous than complex ones.

Keywords: Artificial grammar learning; Centre-embedding; Frequency distribution; Hierarchical structure; Starting small.

To the great interest of linguists and psychologists, children display an amazing ability in extracting rules from language and producing new sentences which obey the rules. Especially, how humans process complex recursive centre-embedded structures with long-distance dependencies, such as "the rat that the dog that the man walked chased ran" is still poorly explained (Corballis, 2007). Moreover, the learnability of this type of structures has become a major issue in language learning research, since recursion has been proposed to be the crucial feature of the human language faculty (Hauser, Chomsky, & Fitch, 2002). One implication of this position is that such structures cannot be learned from environmental stimuli only and by using general cognitive learning mechanisms. The environment contains too little information to induce rules of recursive complexity, and general learning mechanisms are linear, whilst the system to be learned is hierarchical. This point of view is in line with the *poverty of stimulus* hypothesis (Chomsky, 1980; Perfors, Tenenbaum, & Regier, 2011), which proposes that the accessible data are so impoverished that children are unable to induce and generalise structures from these data to acquire full knowledge of the language system. Therefore, natural language grammar learning must be assisted by an inborn device, according to this reasoning. Indeed, the intrinsic properties of recursion, especially centre-embeddings and the corresponding long-distance dependencies, actually pose difficulties for language learners, both in perception and production (Christiansen & Chater, 1999; Gibson, 1998).

A growing body of work attempts to probe into the fundamental cognitive mechanism of learning hierarchical centre-embedded structures (Friederici, 2004; Hochmann, Azadpour, & Mehler, 2008). Except for the *starting-small* (henceforth SS) effect (Elman, 1991), however,

Correspondence should be addressed to Jun Lai, Tilburg Centre for Logic and Philosophy of Science (TiLPS), Dante Building, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: J.Lai@uvt.nl

This research has been funded by the Netherlands Organization for Scientific Research (NWO).

the influence of facilitative factors in learning centre-embedding has hardly been investigated experimentally with artificial grammar in the laboratory environment. Elman (1993) trained a simple recurrent network in a word prediction task to learn the underlying rule of the given grammar. The network first failed to learn when it was exposed to the whole set of input, but then succeeded when being presented with an incremental input. This study showed an advantage of limitation of the input resources. Elman (1993) pioneered the concept that a simple recurrent network could learn sentences containing multiple hierarchical embeddings if it was first confronted with simple structure before stepping further into more complex compound sentences with subclauses. In line with Elman, Lai and Poletiek (2011) also observed this SS effect by manipulating the organisation of the input. Only when the input was arranged incrementally, did participants show learning of a hierarchical centre embedded grammar. In addition, we found that early presentation of a cluster of simple exemplars without embeddings was a prerequisite for learning the centre embedded structure from the embedded sentences presented later on.

A similar hypothesis to the SS effect was proposed by Newport (1990), who showed that early learners of American Sign Language were able to achieve higher competence because they started processing limited individual parts first; late learners who began with complete signs as wholes had more difficulties. However, in contrast to Elman who focused on the structure of staged input, Newport emphasised the internal limitation of cognitive capacity, which actually aided children in successful learning, reducing the units to short sequences in the earliest stage. Empirical evidence also came from Kersten and Earles (2001), who found that adults learned a miniature artificial language better, when they were exposed to an initial training of small constituents, instead of complete sentences. A more general argument was made by Kareev, Lieberman, and Lev (1997), who proposed that due to limitation of working memory capacity, people concentrated on small samples of information, which enlarged the possibility of early detection of correlations in the sample (Hertwig & Todd, 2003).

Some other studies have obtained results contradicting the SS facilitation, however. In two simulation studies, Rohde and Plaut (1999) found no facilitation by SS, but instead an advantage of "starting big" in the presence of semantic constraints. With a third simulation, they excluded the possibility that the constrained memory of the network facilitated learning. Therefore, Rohde and Plaut (1999, 2003) suggested that neither staged input nor restriction of memory was a necessary prerequisite for learning complex statistical regulations. Looking further into the role of cognitive capacity, Ludden and Gupta (2000) stated that the more cognitive resources were provided, the better performance that learners could achieve. Also, older children, and intellectually gifted children showed better learning in an implicit learning task, compared to younger, or intellectually delayed children with limited cognitive capacity (Fletcher, Maybery, & Bennett, 2000). As a final example of SS tests, Conway, Ellefson, and Christiansen (2003) found that participants were assisted in learning both nested and right-branching recursion by the SS input only under the visual modality, but not the auditory modality. Hence, not all learners, and not under all conditions do learners benefit from a growing input.

The purpose of the present research is to explore under what additional conditions of the environmental input does SS facilitate learning. In particular, we suggest that the frequency distribution of the input exemplars may moderate the influence of the SS effect: A starting-small ordering might be most helpful if the simplest exemplars of the grammar not only occur in the earliest stage, but also in higher frequency than the more complex exemplars. In the present work, we aim to test the effects of different types of SS ordering, frequency distribution, and their combination. As a variation of the traditional SS organisation of the input, we let the input grow smoothly, by inserting more complex stimuli gradually, rather than in clusters. By manipulating the frequency distribution, we further evaluate how much preliminary exposure to the simple structures with zero-level-of-embedding (0-LoE) is needed to enhance complex structure learning.

Frequency distribution of the input has been suggested to play a role in inducing structure from that input. For instance, in a categorisation task, adults showed better performance in speech perception by the use of frequency distribution cues of acoustic-phonetic information (Clayards, Tanenhaus, Aslin, & Jacobs, 2008). Moreover, previous studies with children indicated that a skewed distribution facilitated learning new constructions. For instance, Casenhiser and Goldberg (2005) showed that the more frequent a particular single verb was, the better that children learned and generalised the mapping between its form and meaning. Similarly, Kidd, Lieven, and Tomasello (2010) found that high lexical frequency largely boosted children's learning on sentences which contained verbs in high frequency.

In addition, frequency distribution has also been shown to enable adults to learn non-recursive grammatical features. Poletiek and Chater (2006) presented two groups of participants with the same unique exemplars of an artificial finite state grammar, but in two different frequency distributions. One followed the distribution of a natural random output of the grammar. For instance, short and simple exemplars were presented more frequently than long ones, as they were also more frequently repeated in a random output sample. The other distribution was even, i.e., each unique exemplar was presented an equal number of times, disregarding its length. The group exposed to the "natural" random output of the grammar performed better on a grammaticality judgement task than the group exposed to the equally distributed input.

Poletiek and Chater (2006), however, used a nonrecursive finite state grammar. The role of frequency distribution as a cue for inducing structure might also apply to complex recursive grammar learning. Moreover, if the skewness of the input effectively influences the learnability of complex structures, this might explain the twofold findings by Lai and Poletiek (2011): Centre-embedded structure learning requires a combination of both a SS regimen and early exposure to a relatively large cluster of short sequences *without* embeddings. Indeed, successful grammar induction might involve two separate and consecutive learning procedures, requiring (1) early massive exposure to short and simple 0-LoE sentences for grasping the basic pattern of language, and, after that, (2) a smaller number of 1- and 2-LoE items suffice for learning the recursive operation. In such a two-staged learning process, the familiarity of 0-LoE assists human parsers in detecting related elements in more complex items with embedded clauses, showing up in the stimulus set later on. Furthermore, we hypothesise that as exposure to 0-LoE items is more extensive, the detection of corresponding components in later materials is easier.

Hence, we suggest that learners would be helped in grasping a structure by being exposed to more frequent occurrences of simple items it generates, and less frequent complex ones (Casenhiser & Goldberg, 2005; Clayards et al., 2008; Poletiek & Chater, 2006). Notice that this skewed distribution resembles the Zipfian distribution reflected in natural languages (Kurumada, Meylan, & Frank, 2011), in which short and simple constructions occur extremely more often than long and complex occurrences of the grammar.

In the present experiments, we manipulate frequency distribution, i.e., equal versus unequal, and ordering in three ways: (1) the clustered SS set up as in Lai and Poletiek (2011); (2) a gradual SS regimen, i.e., inserting gradually more complex items over time—this gradual SS condition might be more similar to natural learning situations with increasingly complex input; and (3) a random ordering. These manipulations make it possible to evaluate, first, two different types of SS procedures, second, the effect of early exposure to a cluster of simple sentences, and third, the overall effect of frequency distribution of the input.

EXPERIMENT 1

In Experiment 1, we compare three input orderings: first, a discrete SS regimen with items clustered by the number of LoE, second, an incremental SS ordering, and third, a random ordering. Participants were randomly assigned to one of the groups.

Method

Participants. Forty-five students from Leiden University participated. All were native Dutch speakers.

Materials and design. Grammar $\underline{G}$ with an A^nB^n centre-embedded structure in Lai and Poletiek (2011) was used. A novel set of 120 learning strings was generated (Appendix A). Strings were composed of syllables from Category A, i.e., {be, bi, de, di, ge, gi}, and Category B, i.e., {po, pu, to, tu, ko, ku}. Pairs were specified by the consonants, i.e., {be/bi-po/pu}, {de/di-to/tu}, and {ge/gi-ko/ku}. Strings with three different lengths (two, four, or six paired syllables) were applied. Syllable occurrences were balanced in frequencies. The same number of test items was also produced, half grammatical and half ungrammatical (Appendix C). The violations were constructed by mismatching the specific pairing between A- and B-syllables (e.g., $A_1\underline{\mathbf{B_3}}$, $A_1A_2B_2\underline{\mathbf{B_3}}$,

$A_1A_2A_3B_3\underline{\mathbf{B_4}}B_1$, or $A_1A_2A_3B_3B_2\underline{\mathbf{B_4}}$).[1] Violations were not allowed in the middle AB position (except for 0-LoE, in which they were the only possible violation), since an ungrammatical A**B** bigram would be too salient and be easily recognised just by monitoring the superficial characteristics of test items.[2]

Each group was presented with 40 learning items for each LoE. In total, there were 12 blocks, with a learning phase (10 items) and a testing phase (10 items) in each. In the learning phase, the ordering of items was manipulated (Figure 1): For the clustered SS group, participants would first see 0-LoE learning items only in the first four blocks, then only 1-LoE in Blocks 5–8 and 2-LoE in Blocks 9–12. For the incremental SS group, participants would first see only 0-LoE in the first block; From Block 2 on, a few 1-LoE items were introduced gradually and in Block 6, 2-LoE items were introduced. As more complex items were displayed, the number of lower level ones decreased. For the random group, the same material was presented in a randomised order.

All groups were tested with the same items.

Procedure. In the learning phase, participants were instructed that the syllable strings presented were governed by an underlying rule. In each trial, after a fixation cross (500 ms), a learning item was presented syllable-by-syllable visually (800 ms per syllable, with no interval in-between). Participants would see 10 learning items consecutively. Next, 10 novel items were presented in the same way in the test phase, for which grammaticality judgements were required. Feedback was given (500 ms).

Results and discussion

We compared performance over the entire set of 12 blocks for different groups. An ANOVA showed a main effect of condition, $F(2, 42) = 3.23$, $p < .05$, $\eta_p^2 = .13$. As displayed by Figure 2a, only the clustered SS group ($M = 0.60$, $SE = 0.04$) performed significantly above chance, $t(14) = 2.64$, $p < .05$, $r = .58$. *T*-tests showed that the clustered SS group performed significantly better than the random group ($M = 0.50$, $SE = 0.01$), $t(28) = 2.48$, $p < .05$, $r = .42$; yet, there was no significant difference between the clustered SS group and the incremental SS group ($M = 0.54$, $SE = 0.03$), $t(28) = 1.26$, *ns*, nor between the incremental SS group and the random group, $t(28) = 1.40$, *ns*.

To further exclude the possibility that participants might have only concentrated on the outer AB pairs, we compared the performance on two types of violation in 2-LoE test items (i.e., violations at the last position and violations at the second-to-the-last position) for the clustered SS group. A paired *t*-test showed that there was no significant difference between the violations at the last position ($M = 0.55$, $SE = 0.06$), and the violations at the second-to-the-last position ($M = 0.51$, $SE = 0.05$), $t(14) = 0.84$, *ns*, indicating no particular focus on the first-last positions.

We conducted an additional analysis over performance by block. There was no main effect of block, $F(11, 462) = 1.29$, *ns*, nor significant interaction between block and condition, $F(22, 462) = 1.20$, *ns*. As shown in Figure 3a, the clustered SS group showed a gradual learning curve.

The higher performance in the clustered SS regimen replicated the SS effect in Lai and Poletiek (2011). However, the data regarding the incremental SS group suggested that participants were not assisted by the SS input when it increased gradually rather than discretely in complexity. One possible explanation is that as a consequence of the incrementally growing SS presentation, participants lacked sufficient preliminary training with 0-LoE exemplars only. Indeed, under the incremental SS condition, 1-LoE exemplars were introduced in the second block already, which was before all possible unique 0-LoE items could have been learned.

In Experiment 2, we therefore reconducted Experiment 1 with a skewed frequency distribution of the input items. The frequency distribution was determined according to the probabilities of the unique sequences in a random output generated by the grammar. This output typically produces short items with high probability; long and complex items with low probability (see also Poletiek & Wolters, 2009). Item probabilities were calculated by "running" a statistical version of the grammar (Charniak, 1993). In accordance with this distribution, more 0- than 1- and 2-LoE items would be presented during training (Appendix B).

[1] In order to avoid easy detection with the hint of surface heuristics, no repetition of exactly the same syllable was allowed in the same string. In the test string, the number of As and Bs is equal.

[2] This criterion results in: (1) for 1-LoE, the violations would always appear at the last position (e.g., $A_1A_2B_2\underline{\mathbf{B_3}}$); (2) for 2-LoE, we equally divided ungrammatical items into two types of violations: one type with violations at the last position (e.g., $A_1A_2A_3B_3B_2\underline{\mathbf{B_4}}$), and the other with violations at the second-to-the-last position (e.g., $A_1A_2A_3B_3\underline{\mathbf{B_4}}B_1$).

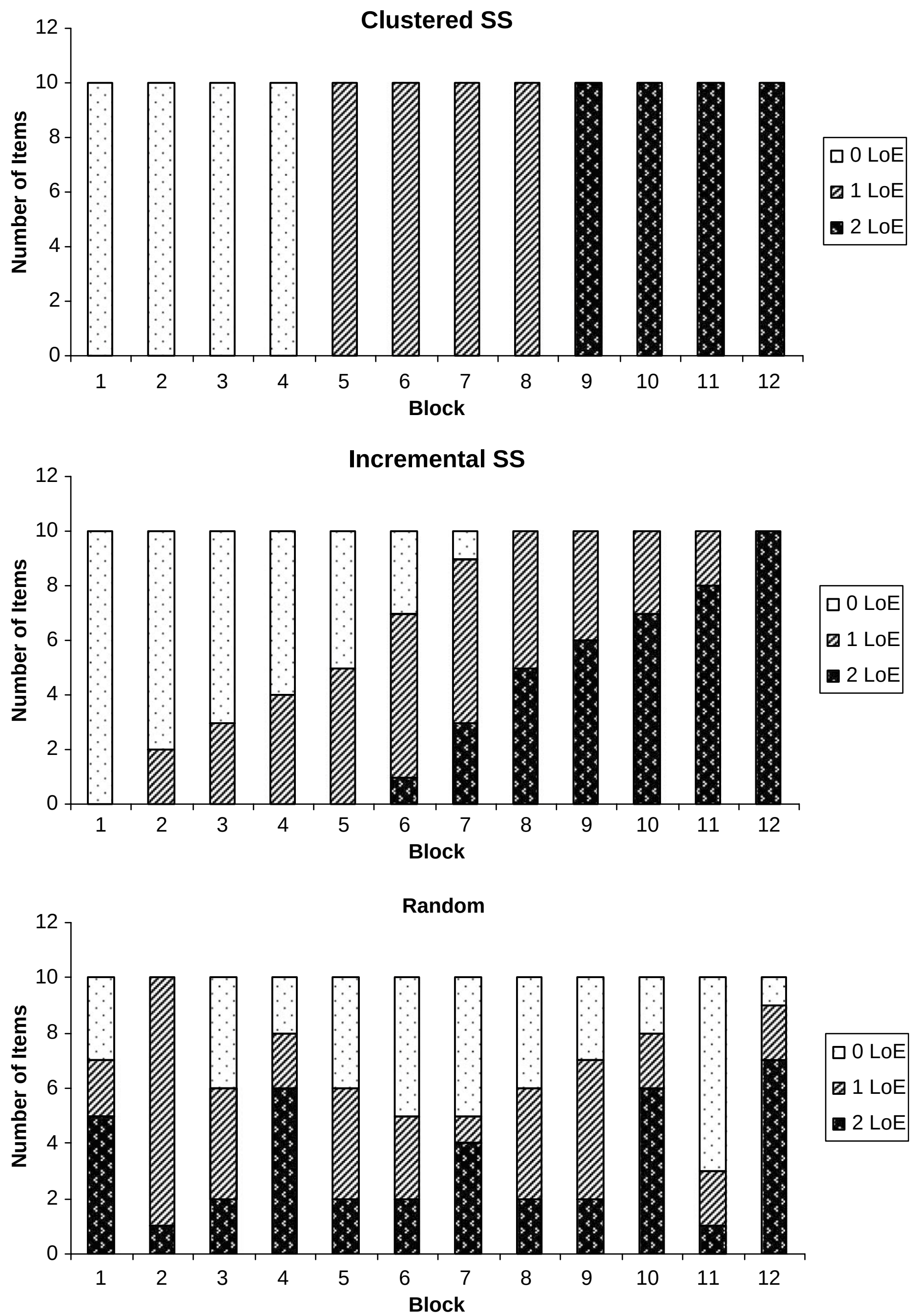

Figure 1. Experiment 1. The ordering of exemplars with 0-, 1-, and 2-LoE in the input under the clustered SS, the incremental SS, and the random condition.

EXPERIMENT 2

Method

Participants. Forty-five students from Leiden University participated. None had participated in Experiment 1.

Materials and design. Three experimental groups were presented with 60 items with 0-LoE, 40 items with 1-LoE and 20 items with 2-LoE (Figure 4): The clustered SS group would see 0-LoE only in the first six blocks, 1-LoE items in the next four blocks, and 2-LoE items in the last two blocks. For the incremental SS group, in the first three blocks participants would see 0-LoE

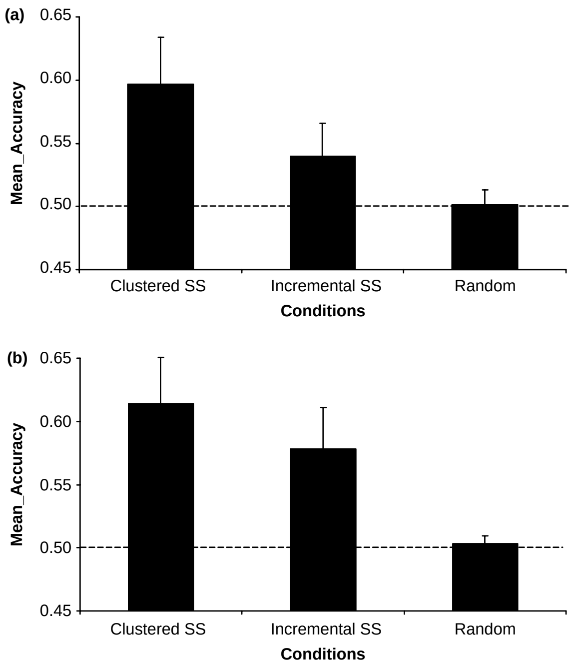

Figure 2. Performance in three groups for (a) Experiment 1, and (b) Experiment 2. The dotted line represents chance level ($M = 0.50$). Error bars indicate standard error of the mean.

items only; In Block 4, two items with 1-LoE were introduced, and, gradually, the input would contain more items with higher LoE. For the random group, the same materials were presented randomly.

Importantly, the same test items as in Experiment 1 were used.

Procedure. Identical to Experiment 1.

Results and discussion

An ANOVA showed a main effect of condition, $F(2, 42) = 3.90$, $p < .05$, $\eta_p^2 = .16$. As displayed in Figure 2b, performance was significantly better than chance for both the clustered SS group ($M = 0.61$, $SE = 0.04$), $t(14) = 3.11$, $p < .01$, $r = .64$, and the incremental SS group ($M = 0.58$, $SE = 0.03$), $t(14) = 2.39$, $p < .05$, $r = .54$. The random group ($M = 0.50$, $SE = 0.01$) did not differ significantly from chance, $t(14) = 0.47$, *ns*. *T*-tests indicated significant differences between the clustered SS group and the random group, $t(28) = 2.97$, $p < .01$, $r = .49$, and also between the incremental SS group and the random group, $t(28) = 2.25$, $p < .05$, $r = .39$, but not between the clustered SS group and the incremental SS group, $t(28) = 0.73$, *ns*.

The higher than chance accuracy of grammaticality judgement in the clustered SS group once again verified the original SS effect in Lai and Poletiek's (2011) study. In addition, in contrast to the results in Experiment 1, the incremental SS group, with a preliminary exposure to three blocks with 0-LoE only, now outscores chance level.

We also compared performance on different types of 2-LoE ungrammatical items. We found no difference between the violations at the last position and at the second-to-the-last position, for the clustered SS group. $M_{\text{Last}} = 0.64$, $SE_{\text{Last}} = 0.05$, $M_{\text{Second-to-the-last}} = 0.62$, $SE_{\text{Second-to-the-last}} = 0.05$, $t(14) = 0.32$, *ns*, as well as for the incremental SS group, $M_{\text{Last}} = 0.53$, $SE_{\text{Last}} = 0.05$, $M_{\text{Second-to-the-last}} = 0.51$, $SE_{\text{Second-to-the-last}} = 0.05$, $t(14) = 0.48$, *ns*.

A repeated-measure analysis showed that there was a main effect of block, $F(11, 462) = 2.80$, $p < .005$, $\eta_p^2 = .06$, and a significant interaction between block and condition, $F(22, 462) = 2.10$, $p < .005$, $\eta_p^2 = .09$ (Figure 3b).

Combined analysis

We probed into accuracy after exposure to various numbers of blocks with only 0-LoE learning items in both experiments (Figure 5). In line with our proposal, mean performance shows an increasing trend correlating with the number of training items with only 0-LoE items presented at the beginning of exposure to the input. When participants were trained with only one block of 0-LoE learning items in the beginning (i.e., the incremental SS group with equal distribution), their performance did not differ from chance ($M = 0.54$, $SE = 0.03$), $t(14) = 1.60$, *ns*. However, when they were exposed to three (i.e., the incremental SS group with unequal distribution), four (i.e., the clustered SS group with equal distribution), or six blocks (i.e., the clustered SS group with unequal distribution) with only 0-LoE learning items, they performed significantly above chance level ($M = 0.58$, $SE = 0.03$), $t(14) = 2.39$, $p < .05$, $r = .54$; ($M = 0.60$, $SE = 0.04$), $t(14) = 2.64$, $p < .05$, $r = .58$; ($M = 0.61$, $SE = 0.04$), $t(14) = 3.11$, $p < .01$, $r = .64$, respectively.

Interestingly, participants, who were exposed to only 0-LoE learning items during the first three blocks, already performed above chance level on 0-LoE test items during Blocks 1–3 ($M = 0.61$, $SE = 0.03$), $t(44) = 3.96$, $p < .001$, $r = .26$, and also

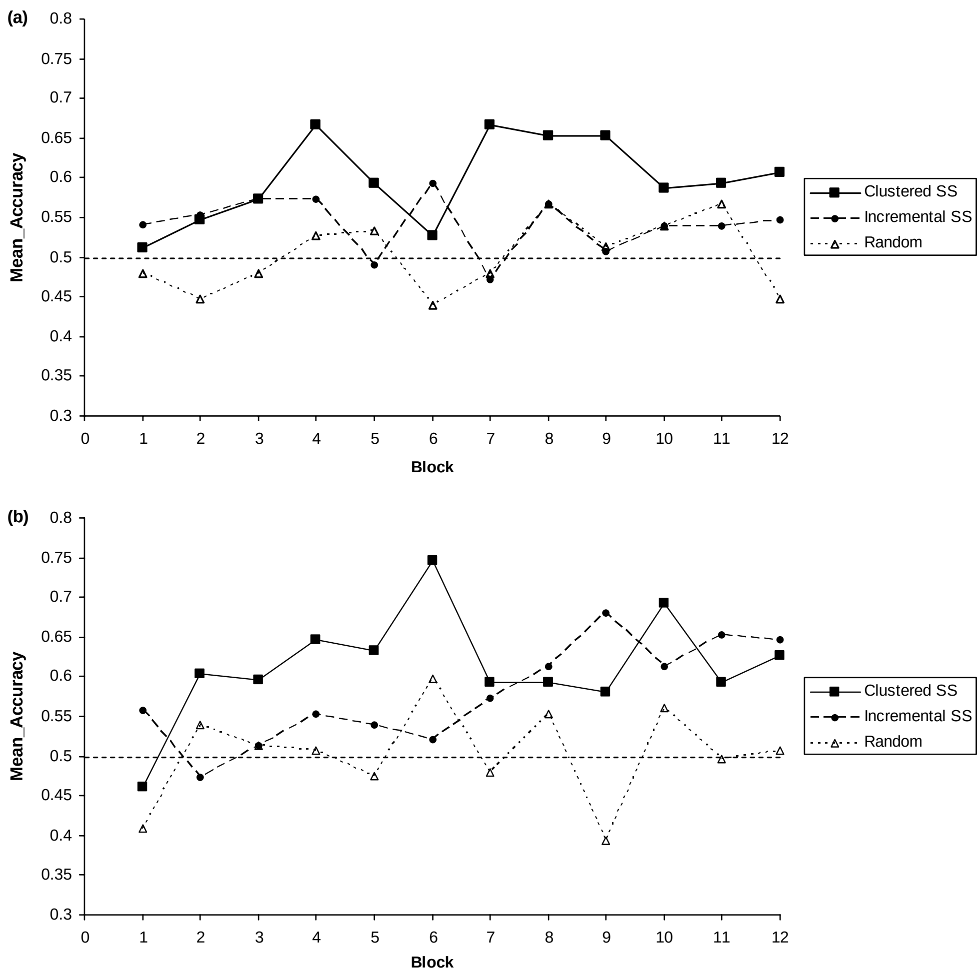

Figure 3. Performance on 12 blocks for three groups for (a) Experiment 1, and (b) Experiment 2. The dotted line represents chance level ($M = 0.50$).

during Blocks 4–12 ($M = 0.67$, $SE = 0.03$), $t(44) = 6.32$, $p < .001$, $r = .69$. Performance was at chance level for 1-LoE test items ($M = 0.52$, $SE = 0.02$), $t(44) = 0.82$, *ns*, and 2-LoE test items ($M = 0.52$, $SE = 0.02$), $t(44) = 0.74$, *ns* during Blocks 1–3, whereas performance improved to be above chance during Blocks 4–12, both for 1-LoE items ($M = 0.61$, $SE = 0.03$), $t(44) = 3.97$, $p < .001$, $r = .51$, and 2-LoE items ($M = 0.60$, $SE = 0.02$), $t(44) = 4.00$, $p < .001$, $r = .52$. The results indicate that three consecutive blocks with only simple 0-LoE learning items are crucial to grasp the embedding structure displayed in the more complex stimuli that follow.

GENERAL DISCUSSION

In Experiment 1, we compared the effect of two types of SS training regimens on learning a centre-embedded grammar: a discrete ordering with consecutive clusters with increasing LoE for each cluster, and a continuous ordering, in which exemplars with more embedded clauses are gradually inserted in the training input. Only the discrete SS group outperformed the randomly ordered control group significantly. This result replicates the facilitation effect of a discrete SS training regimen in Lai and Poletiek (2011).

The absence of the beneficial effect in the continuous version of the SS training was explained by the absence of sufficient preliminary training on exemplars of the grammar without applications of the centre-embedded rule. In Experiment 2, we tested this possibility with the same ordering conditions as Experiment 1, but with exemplars' frequencies inversely related to their complexity (50% of the learning set is simple 0-LoE items, 33% 1-LoE items, and 17% 2-LoE items). The skewed frequency distribution

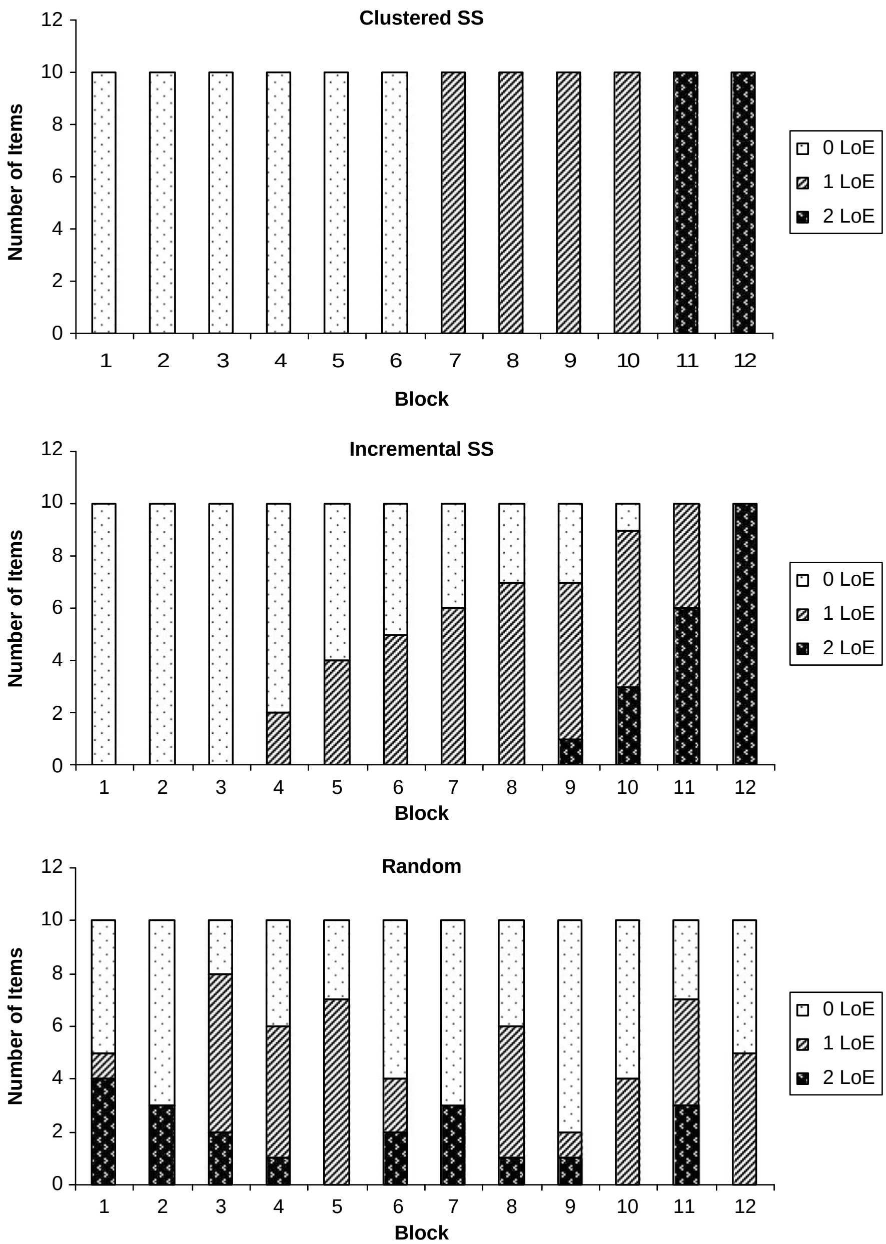

Figure 4. Experiment 2. The ordering of exemplars with 0-, 1-, and 2-LoE in the input under the clustered SS, the incremental SS, and the random condition.

formally corresponds to a random output of the grammar, and resembles the distribution of natural language input, in which short and simple sentences occur more often than long and complex ones. Testing two SS regimens with the skewed distribution provided the possibility to disentangle the contributions of the input ordering from the influence of early exposure to simple items only, on learning. Indeed, when the input distribution was skewed, the learner in the continuous SS condition would still be exposed to a substantial cluster with the simplest exemplars only in the beginning of training.

As we observed, learning was also enhanced by a continuous SS training, when the distribution of the input was skewed to favour highly frequent 0-LoE items. The contribution of the skewed frequency distribution might originate mostly from the massive exposure to 0-LoE items, instead of the decrease of multiple LoE items. Moreover, our proposal that this combined facilitation was accounted for by early intensive

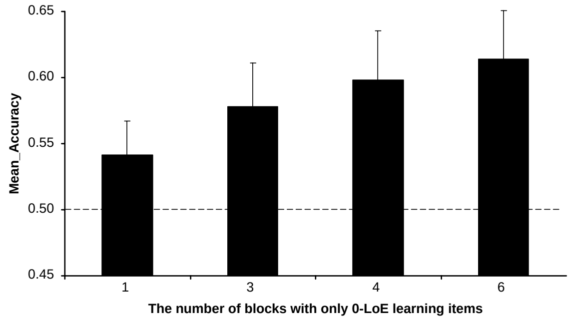

Figure 5. Performance after exposure to various numbers of blocks with only 0-LoE learning items in Experiments 1 and 2. The dotted line represents chance level (M =0.50). Error bars indicate standard error of the mean.

exposure to a cluster of 0-LoE only was supported by a finding emerging from both experiments: Only participants who were exposed to at least three blocks of simple structures without embeddings showed any learning of subsequently presented centre embedded structures. Those presented with embedded items right away did not improve after the first three blocks.

In sum, our data replicate the SS effect also when the input grows continuously rather than discretely from simple towards more complex, but only when the frequency distribution of the exemplars at training favours high numbers of simple exemplars. In this manner, by adapting two characteristics of the input to make it more representative for the natural linguistic input—continuous SS and skewed frequency distribution—we could show how these characteristics of the environment form an optimal setting for learning to emerge.

It seems that the earliest stage of training serves as an essential stepping stone for eventual acquisition of the complex centre-embedded grammar. A possible cognitive explanation of this facilitation process is that frequent and early exposure to the basic pattern of the grammar splits up the learning in different consecutive parts with separate learning goals: first, the solid acquisition of a basic pattern, and second, detecting the recursive operation that operates on that basic pattern. An environment that separates the steps and organises their time course accurately fits the needs of the learner. As in natural language, the child-directed speech contains mostly shorter and simpler phrases than adult speech (Pine, 1994).

Another aspect of this fit between environment and learner might be the constrained cognitive capacity of the learner in the first stage of exposure. As Kersten and Earles (2001) indicate, the limitations of children's processing abilities might make focus on small constituents first and enabled better ultimate language performance than adults. Young children learning language naturally start to process linguistic stimuli using the simplest ''model'' that accounts for the input (Chater & Vitányi, 2002). The *less is more* (Newport, 1990) hypothesis reflects this idea. First, linguistic sequences may be processed by an associative linear learning mechanism. As the input grows in complexity, along with cognitive capacities, processing might become more complex and hierarchical. Since our study was based on adult participants, future research is needed to investigate how developmental cognitive factors interact with the environmental characteristics investigated here.

Although the current results reveal some crucial properties of the learning process, there are of course limitations of this type of artificial grammar learning studies. For instance, our work mimicked some ideal ''error free'' learning environment, and used visual materials (Conway et al., 2003; Lai & Poletiek, 2011). Also, we tried to simulate the development of children's learning by observing adults' behaviour in a laboratory task. And we used a fixed artificial meaningless vocabulary, which differs largely much from the rich natural language vocabulary (Fedor, Varga, & Szathmary, 2012). However, there are also undeniable strengths of the artificial grammar learning approach, such as the possibility to investigate the hypothesised factors in isolation, disregarding temporarily the richness of natural language, e.g., semantics (Gomez & Gerken, 1999).

Our experiments indicate that in the lab and possibly in natural learning situations, learners can utilise complexity-based ordering and frequency variations of stimuli over time, as cues to abstract complex pattern information, avoiding in this manner the difficulty of inducing these complex structures by computation.

Original manuscript received April 2012
Revised manuscript received January 2013
First published online April 2013

REFERENCES

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, *8*(6), 500–508.

Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: MIT Press.
Chater, N., & Vitányi, P. (2002). Simplicity: A unifying principle in cognitive science? *Trends in Cognitive Sciences*, *7*, 19–22.
Chomsky, N. (1980). Discussion of Putnam's comments. In M. Piatelli-Palmarini (Ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky* (pp. 310–324). Cambridge, MA: Harvard University Press.
Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23*(2), 157–205.
Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804–809.
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: Lawrence Erlbaum Associates.
Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science*, *31*, 697–704.
Elman, J. L. (1991). *Incremental learning, or the importance of starting small* (Tech. Report No. 9101). Center for Research in Language, University of California at San Diego.
Elman, J. L. (1993). Learning and developing in neural networks: The importance of starting small. *Cognition*, *48*, 71–99.
Fedor, A., Varga, M., & Szathmary, E. (2012). Semantics boosts syntax in artificial grammar learning tasks with recursion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 776–782.
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. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences*, *8*, 245–247.
Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76.
Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.
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.
Hertwig, R., & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making* (pp. 213–231). Chichester: Wiley.
Hochmann, J.-R., Azadpour, M., & Mehler, J. (2008). Do humans really learn AnBn artificial grammars from exemplars? *Cognitive Science*, *32*, 1021–1036.
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*, 250–273.
Kidd, E., Lieven, E., & Tomasello, M. (2010). Lexical frequency and exemplar-based learning effects in language acquisition: Evidence from sentential complements. *Language Sciences*, *32*, 132–142.
Kurumada, C., Meylan, S., & Frank, M. C. (2011). Zipfian word frequencies support statistical word segmentation. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 2667–2672). Austin, TX: Cognitive Science Society.
Lai, J., & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition*, *118*(2), 265–273.
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 Associates.
Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11–28.
Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, *118*(3), 306–338.
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: 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* (pp. 1968–1973). Mahwah, NJ: Lawrence Erlbaum Associates.
Poletiek, F. H., & Wolters, G. (2009). What is learned about fragments in artificial grammar learning? A transitional probabilities approach. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *62*, 868–876.
Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*(1), 67–109.
Rohde, D. L. T., & Plaut, D. C. (2003). Connectionist models of language processing. *Cognitive Studies*, *10*, 10–28.

APPENDIX A

Learning stimuli in the clustered starting small (SS) and the incremental SS condition of Experiment 1

	Condition	
Blocks	*Clustered SS*	*Incremental SS*
Block1	detu	detu
	deto	deto
	ditu	ditu
	dito	dito
	bepu	bepu
	bipu	bepo
	bipo	geku
	geku	geko
	geko	giku
	giku	giko
Block 2	detu	detu
	deto	ditu
	ditu	bepu
	bepu	bepo
	bepo	bipo
	bipu	geko
	bipo	giku
	geko	giko
	giku	degekutu
	giko	didetotu
Block 3	detu	detu
	deto	deto
	dito	ditu
	bepu	dito
	bepo	bepu
	bipu	bipu
	geku	giku
	geko	digekotu
	giku	bibepopu
	giko	gebipoku
Block 4	deto	deto
	ditu	bipo
	dito	geku
	bepu	geko
	bepo	giku
	bipo	giko
	geku	dedituto
	geko	bebipupo
	giku	gebepuku
	giko	giditoko
Block 5	dedituto	deto
	degekutu	bepu
	didetotu	bepo
	digekotu	bipu
	bebipupo	geku
	bidetupu	degekoto
	bibepopu	digikotu
	gebepuku	bebipopu
	gebipoku	bidetupu
	giditoko	gedituku
Block 6	debepotu	dito
	degekoto	bipu
	dibeputo	geko

APPENDIX A (*Continued*)

	Condition	
Blocks	*Clustered SS*	*Incremental SS*
	digikotu	debepotu
	bebipopu	dibeputo
	biditupu	biditupu
	bigekupu	bigekupu
	gedituku	gegikuko
	gegikuko	gidetuko
	gidetuko	didegekutotu
Block 7	debiputu	bipo
	degikutu	degikutu
	digekuto	digekuto
	bedetopo	bedetopo
	begikopu	biditupo
	biditupo	gedetuko
	gedetuko	gibepuko
	gebepoko	debegekoputo
	gibepuko	bidibepotopu
	gigekoku	gibigekopoku
Block 8	debipoto	debiputu
	didetuto	begikopu
	dibiputo	gebepoko
	beditopu	gigekoku
	begikupo	biditopo
	biditopo	dedibepututo
	bigikopo	digibepokotu
	gidetoku	begedetukupo
	gibipuku	bidegikotupo
	gigekuko	gidebepotoku
Block 9	dedibepututo	dibiputo
	degebipukotu	beditopu
	dibigikuputo	gidetoku
	digibepokotu	bigikopo
	bedigikutopo	degebipukotu
	bidegikotupo	dibigikuputo
	bigeditokupo	bedigikutopo
	gebedetopuko	bigeditokupo
	gibegekupoko	gebedetopuko
	gibigekopoku	gibegekupoko
Block 10	debegekoputo	didetuto
	degibipokuto	begikupo
	dibigikoputu	gigekuko
	bededitotupu	degibipokuto
	begedetukupo	dibigikoputu
	bidibepotopu	bededitotupu
	bigiditukopu	bigiditukopu
	gebeditopoko	gebeditopoko
	gidibiputuku	gidibiputuku
	gigebipokoku	gigebipokoku
Block 11	debegikopotu	debipoto
	didegekutotu	gibipuku
	digedetokutu	debegikopotu
	bedegikotopu	digedetokutu
	begeditukopu	bedegikotopu
	bibegekupupo	begeditukopu
	gedebeputuku	bibegekupupo
	gebidetupoko	gedebeputuku
	gidebepotoku	gebidetupoko
	gibiditopuko	gibiditopuko

APPENDIX A (*Continued*)

	Condition	
Blocks	*Clustered SS*	*Incremental SS*
Block 12	debigikupotu	debigikupotu
	dibegekopoto	dibegekopoto
	digidetukuto	digidetukuto
	bedibipotupu	bedibipotupu
	begibipukopo	begibipukopo
	bigedetokupu	bigedetokupu
	gedibipotoku	gedibipotoku
	gebibepupoku	gebibepupoku
	gidebepotoko	gidebepotoko
	gigeditukoku	gigeditukoku

APPENDIX B

Learning stimuli in the clustered starting small (SS) and the incremental SS condition of Experiment 2

	Condition	
Blocks	*Clustered SS*	*Incremental SS*
Block1	detu	detu
	deto	ditu
	ditu	dito
	dito	bepu
	bepu	bipu
	bipu	bipo
	bipo	geku
	geku	geko
	geko	giku
	giku	giko
Block 2	detu	deto
	deto	ditu
	ditu	bepu
	dito	bepo
	bepo	bipu
	bipu	bipo
	bipo	geku
	geku	geko
	geko	giku
	giko	giko
Block 3	detu	detu
	deto	deto
	ditu	dito
	bepu	bepu
	bepo	bipu
	bipu	bipo
	bipo	geku
	geku	geko
	giku	giku
	giko	giko
Block 4	detu	detu
	deto	ditu
	dito	dito

APPENDIX B (*Continued*)

	Condition	
Blocks	*Clustered SS*	*Incremental SS*
	bepu	bepo
	bepo	bipu
	bipu	bipo
	bipo	geku
	geko	giko
	giku	gigekoku
	giko	gebepuku
Block 5	detu	detu
	ditu	deto
	dito	bepu
	bepu	bipu
	bepo	geko
	bipu	giku
	geku	dedituto
	geko	digikotu
	giku	biditupu
	giko	gigekuko
Block 6	deto	ditu
	ditu	dito
	dito	bepo
	bepu	bipo
	bepo	geku
	bipo	degikutu
	geku	didetotu
	geko	bidetupu
	giku	gibipuku
	giko	begikopu
Block 7	dedituto	deto
	degekutu	ditu
	didetotu	bepu
	digekotu	giko
	bebipupo	degekutu
	bidetupu	digekotu
	bibepopu	bebipupo
	gebepuku	bibepopu
	gebipoku	gebipoku
	giditoko	giditoko
Block 8	debepotu	deto
	degekoto	dito
	dibeputo	giku
	digikotu	debepotu
	bebipopu	degekoto
	biditupu	dibeputo
	bigekupu	bebipopu
	gedituku	bigekupu
	gegikuko	gedituku
	gidetuko	gidetuko
Block 9	debiputu	detu
	degikutu	bepo
	digekuto	geko
	bedetopo	debiputu
	begikopu	digekuto
	biditupo	bedetopo
	gedetuko	biditupo
	gebepoko	gegikuko
	gibepuko	gibepuko
	gigekoku	gebeditopoko

APPENDIX B (*Continued*)

	Condition	
Blocks	*Clustered SS*	*Incremental SS*
Block 10	debipoto	bepo
	didetuto	gedetuko
	dibiputo	dibiputo
	beditopu	gidetoku
	begikupo	bigikopo
	biditopo	beditopu
	bigikopo	debipoto
	gidetoku	didegekutotu
	gibipuku	dedibepututo
	gigekuko	bedegikotopu
Block 11	dedibepututo	didetuto
	degibipokuto	begikupo
	didegekutotu	biditopo
	digibepokotu	gebepoko
	bedegikotopu	degibipokuto
	bidegikotupo	digibepokotu
	bigeditokupo	bidegikotupo
	gedebeputuku	bigeditokupo
	gebeditopoko	gedebeputuku
	gidebepotoko	gidebepotoko
Block 12	debigikupotu	debigikupotu
	dibegekopoto	dibegekopoto
	digedetokutu	digedetokutu
	bededitotupu	bededitotupu
	begedetukupo	begedetukupo
	bigiditukopu	bigiditukopu
	gebedetopuko	gebedetopuko
	gigebipokoku	gigebipokoku
	dibigikuputo	dibigikuputo
	gidibiputuku	gidibiputuku

APPENDIX C

Testing stimuli in all conditions of Experiments 1 and 2

	Grammaticality	
Blocks	*Grammatical*	*Ungrammatical*
Block1	detu	be**to**
	dibiputu	gedeto**tu**
	bigekupo	gidito**pu**
	degebipokotu	dibigiko**tu**to
	gibidetopuku	bibeditupu**to**
Block 2	ditu	bi**tu**
	debeputo	geditu**po**
	bedetopu	gibepu**to**
	didebepututo	begedetu**to**pu
	gedegikotoku	bigidituko**ku**
Block 3	bepu	ge**tu**
	digikutu	dedito**pu**
	bigekopo	gibepo**tu**
	degibipokuto	digibipo**to**tu
	bigedetukupu	gedibepoto**pu**

APPENDIX C (*Continued*)

	Grammaticality	
Blocks	*Grammatical*	*Ungrammatical*
Block 4	bipu	gi**to**
	degekotu	bedetu**ku**
	dibipotu	gegiko**po**
	didegikototu	dedibepoto**ku**
	bidegikotopu	gegibipu**to**ku
Block 5	geku	de**pu**
	dibipoto	bidetu**ko**
	beditopo	gibipo**pu**
	dibegekupoto	debegeku**to**tu
	gebedetupoko	gibedetupo**to**
Block 6	giku	di**po**
	degikotu	begiko**to**
	gedetoko	bideto**tu**
	dibidetoputu	debegekopo**ku**
	bedigekotopo	bidegiku**po**pu
Block 7	deto	be**ku**
	gebipuku	digiko**pu**
	gibepoku	bigiku**tu**
	dibibepupotu	debegeko**ku**tu
	gibeditupuko	gebidetopu**tu**
Block 8	dito	bi**ko**
	gebepuko	degeku**po**
	gidituku	begeko**tu**
	digebepokutu	degebepoku**ko**
	begiditukopu	bidigeko**po**pu
Block 9	bepo	ge**po**
	bigikupu	debipo**ko**
	gidetoko	digeku**po**
	digebipokoto	degibipu**tu**to
	bigeditokupu	gibibepupo**to**
Block 10	bipo	gi**pu**
	begikupu	debipu**ko**
	geditoko	dibepo**ku**
	bidegikutopo	degibipuko**ku**
	gibegekupoko	gedigiku**po**ko
Block 11	geko	de**ko**
	beditupo	dibepu**ku**
	gidetuku	bidito**ku**
	bibegekupopu	dibedetupo**ko**
	gidebeputuku	gebeditu**to**ko
Block 12	giko	di**ku**
	debepoto	begeku**ko**
	bigekopu	gebipu**to**
	begibipukopo	bidigiku**ko**pu
	gegiditukoku	gigeditoku**po**