Studying Frequency Effects in Learning Center-embedded Recursion

Lai, J.; Krahmer, E.J.; Sprenger, J.M.

Published in:
Proceedings of the 36th Annual Conference of the Cognitive Science Society (CogSci 2014)

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Studying Frequency Effects in Learning Center-embedded Recursion

Jun Lai (J.Lai@ tilburguniversity.edu)
Tilburg Center for Logic and Philosophy of Science; Tilburg Center for Cognition and Communication,
Tilburg University, the Netherlands

Emiel Krahmer (E.J.Krahmer@ tilburguniversity.edu)
Tilburg Center for Cognition and Communication, Tilburg University, the Netherlands

Jan Sprenger (j.sprenger@tilburguniversity.edu)
Tilburg Center for Logic and Philosophy of Science, Tilburg University, the Netherlands

Abstract

Long-distance dependencies in center-embedded recursion are among the most typical but also most difficult structures in human language (Corballis, 2007; Hauser, Chomsky, & Fitch, 2002). Concerning the impact of the learning sample on grasping object-action relations, there are two opposing arguments: *more is better* vs. *fewer is better* (Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008). The former theory assumes that a large number of different exemplars facilitates learning (Gentner, 2003), while the latter theory suggests that a more restricted set of unique exemplars with repetitions advances the learning of these patterns (Casasola, 2005; Kersten & Smith, 2002). In the current study, we designed a grammaticality-judgment task and test both theories using an artificial grammar learning paradigm. We found that when participants were trained on fewer unique exemplars, but with repetitions, they could still perform significantly better than at chance level. Moreover, when the few unique exemplars were repeated for an unequal number of times, their performance was boosted to a higher level. In line with the fewer is better theory, our findings point to a repetition effect and frequency distribution effect in processing hierarchical center-embedded recursion.

Keywords: center-embedded recursion; statistical learning; starting small; repetition; frequency distribution

Introduction

In early language acquisition, children display remarkable capabilities in learning and producing new sentences. One of the most discouraging obstacles in language learning lies in making associations between objects and actions (Kersten & Smith, 2002). Children tend to initially concentrate on objects and agents, rather than on the relations between them (Behrend, 1990). One of the most difficult sentential relations to learn is center-embedded (CE) recursion, which is proposed to be a crucial factor distinguishing humans and nonhumans (Corballis, 2007; Hauser, Chomsky, & Fitch, 2002). For instance, "*The rabbit that the fox chased ran away.*" is a typical CE recursive sentence, with one object-action pair (*fox-chased*), inserted as a relative clause in the middle of another pair (*the rabbit-ran away*) that forms the main sentence. The complexity of the sentence poses great challenges for sentence processing and understanding, since human parsers not only need to store all the relevant information in memory, but also need to integrate the associated elements, in order to detect who did what to whom.

There is a large amount of research attempting to explain why the relations in CE recursion are so notoriously difficult to comprehend (Christiansen & MacDonald, 2009). For instance, Chomsky (1965) argued that the way that CE recursion is constructed requires human parsers to associate related elements, which is complicated for remote dependencies. Moreover, Johnson (1998) and Morrill (2000) indicated that the comprehension collapses at the moment when too many dependencies are waiting to be paired in memory.

Though processing CE recursion seems extraordinarily challenging, it has been shown that acquisition of such complicated structures could be enhanced to a certain extent by varying the learning sample, for instance, when the complexity of the learning material is arranged incrementally (Conway, Ellefson, & Christiansen, 2003; Elman, 1993; Lai & Poletiek, 2011). It has also been shown that the high variability of the intervening middle element in a non-adjacent structure could result in better understanding and discrimination of the remote dependencies (Gomez, 2002). Furthermore, the frequency distribution of the co-occurrence of related dependencies has been observed to affect the categorization of the intervening elements (Mintz, 2003). In addition, it has been shown that people could master the sentential structures easier when verbs were in high frequencies (Kidd, Lieven, & Tomasello, 2010). Thus, a skewed frequency distribution, which could result in different constructions of the sample, e.g. size and probability, is also an aid in learning complex structures.

With regard to the effects on sample diversity on facilitating learning, there are primarily two opposite points of view: *more is better* vs. *fewer is better* (see Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008, for a review). The first perspective assumes that the variety of learning samples helps children in unpacking the sentential structures. Therefore, exposure to a larger amount of different exemplars advances children's development in extracting relations (Gentner, 2003). Evidence for this position comes from studies on natural language production (Gallivan, 1987; Rinaldi, Barca, & Burani, 2004), but also experiments with nonlinguistic materials (Gomez, 2002). In a statistical learning task, Gomez (2002) trained both adults and children to learn the a-b relation in the *aXb* pattern, with sample size ranging from 2, 6, 12, to 24 items. The highest learning scores were observed with the largest sample size, of 24 items. Similarly, in a series of artificial grammar learning (AGL) studies, participants who were presented with diverse training exemplars showed increasing sensitivity towards the sequential structure, while those who were trained with a few exemplars were only able to memorize certain individual fragments (McAndrews & Moscovitch, 1985; Meulemans & VanderLinden, 1997). Furthermore, with computational simulations, Harris (1991) found that a network succeeded in learning and making generalizations when it was exposed to a complete "big" set of data, instead of a "small" subset of data. Arguably, the larger sample variety might provide a better estimate of the general population.

By contrast, the alternative perspective regards the repetition of a limited set of exemplars as an accelerator of abstracting relational information (Kersten & Smith, 2002). Experiments with 14-month-olds showed a learning effect in a spatial categorization task only when the infants were exposed to two-pair-relations, instead of six-pair-relations (Casasola, 2005). In addition, Elman (1993) found a "starting small" (SS) effect in simulations with neural networks. His Simple Recurrent Networks could not learn the new language when they were given the entire training input at once, whereas the network showed learning when given restricted data initially, namely staged-input increasing gradually. Hence, the debate remains inconclusive as to whether the exposure to a large or a small amount of various exemplars helps learners to acquire associations between object and actions in CE recursion and to form the correct relational categories.

In the current study, we address this issue by investigating the influence of sample properties on comprehending CE recursion with the AGL paradigm. AGL is a well-known and often used paradigm to study natural language learning. Recent research (Pelucchi, Hay, & Saffran, 2009) found that in a task of learning novel words, 8-month-olds showed a highly consistent pattern with that from statistical learning using an artificial language. Nevertheless, natural language is an extremely complex system, integrating information from many different sources, such as lexicon, semantics, syntax, etc., whereas our artificial language is quite straightforward. However, it is precisely this difference between artificial and natural languages that allows us to manipulate various aspects of the learning conditions. In this manner, we were able to analyze the specific influence of linguistic variables on language learning, and avoid being confounded by the co-occurrence of multiple linguistic components.

In the grammar-learning experiment of the current study, we manipulated two aspects of the sample set, i.e. the diversity of unique learning exemplars, and the occurrence of the learning exemplars according to their frequencies. In a previous study with CE recursion (Lai & Poletiek, 2013), we observed a skewed frequency distribution effect, i.e. participants showed better performance when exemplars with different levels of embedding (LoE) were presented unequally than equally. In that study, with sufficient exposure to the basic associations, participants were still able to learn the structure when given fewer items of higher complexity. Accordingly, on the one hand, we are interested in whether we could maintain the effect of "fewer exemplar", when decreasing the number of learning items for all levels of complexity. On the other hand, we are curious about the interaction between the decreased sample variety and the skewed frequency distribution.

Therefore, in the present study, we hypothesize that: 1) People might not necessarily need a large number of different exemplars when learning a new complex structure. Instead, fewer unique exemplars, which are repeated, could facilitate the processing of CE recursion and detection of the inherent relations. 2) The occurrence of exemplars with different frequencies might affect the learning performance: the more high-frequency exemplars occur, the better participants perform.

Our current study is a crucial complement to the previous research on statistical learning and language development in a number of aspects. Firstly, referring to sample size, previous studies have focused mainly on the absolute number of sample items in total (Gomez, 2002;McAndrews & Moscovitch, 1985; Poletiek & van Schijndel, 2009), but not on *the diversity and the relative frequency of the exemplars in the sample*. Although a few experiments addressed this issue in verb learning (Maguire et al., 2008), in spatial categorization (Casasola, 2005), or in objects relations (Kersten & Smith, 2002; Quinn, Polly, Furer, Dobson, & Narter, 2002), no research has pinpointed the exemplars' uniqueness in processing the non-adjacent relations of *hierarchical CE structures*.

Secondly, previous research mostly focused on human parsers' cognitive limitation and the facilitative effect that it might bring about. For example, it has been shown that children's limited working memory and processing abilities only direct them to individual linguistic segments. As their cognitive abilities develop, they become able to analyze and produce more complex structure of a language (Newport, 1990). Similarly, Goldowsky and Newport (1993) presented a statistical model and demonstrated the advantage of memory limitations in learning a morphological system; Kareev, Lieberman, and Lev (1997) found that participants with lower memory span provided larger response correlations and higher accuracies. Moreover, in a study of American Sign Language, Cochran, McDonald, and Parault (1999) found that participants could show better learning of novel verbs with the aid of cognitive limitations.

Nevertheless, instead of focusing on the internal limited cognitive resources, only a few studies investigated *the properties of the actual external input per se*. For instance, Elman (1993) stressed the importance of starting with a small subset at an initial stage of learning, instead of the entire set as a whole. Lai and Poletiek (2011) replicated this SS effect for learning CE recursion by showing that the gradual increasing ordering of input complexity assisted participants in learning this construction. Lai and Poletiek also showed that the sufficient exposure to the basic adjacent-dependencies, i.e. 0-LoE (for instance, sentences like "*The dog runs*.") was another necessity for successful learning. Furthermore, Lai & Poletiek (2013) verified the SS effect again with a skewed frequency distribution. The frequencies of different LoE items were inversely related to their complexity (i.e. 50% 0-LoE items, 33% 1-LoE and 17% 2-LoE items). Nevertheless, as far as we know, no previous studies have explored the sufficient proportion of unique exemplars *within each LoE*. For instance, in order to learn the CE recursive sentences with one embedding, we do not know yet whether we need to see as many different unique combinations as possible, or merely a certain limited combination, but with a lot of repetition.

Thirdly, with regard to frequency distribution, a number of highly frequent subject-verb relations appear much more often than the others. For example, in a sentence like "*The horse that the astronaut that the veterinarian chiseled possessed ran*," three subject-verb pairs represent three different levels of relative frequencies: high, medium and low, respectively. Previous studies using the AGL paradigm attempted to simulate the skewed distribution of natural language input, merely by presenting the simple and short sentences more frequently than the long and complex ones, either with a phrase structure grammar (Lai & Poletiek, 2013), or with a finite grammar (Poletiek & Chater, 2006). However, previous studies have not specified the skewed frequency distribution *for certain level of complexity*, e.g. more A_1B_1 than A_2B_2, or more $A_1A_2B_2B_1$ than $A_2A_3B_3B_2$, etc. Studying this influence is a major novel aspect of our research.

In response to the issues discussed above, in the current experiment, we compared the performance of learning an artificial CE recursion under three learning conditions, with exactly the same amount of learning input but different content. The first condition is an incrementally staged input with multiple various exemplars (Starting-Small), for which we replicated the learning material used by Lai and Poletiek (2011); the second is a smaller set of fewer unique exemplars, which are repeated for an equal number of times (Starting-Less); and the third contains the same unique exemplars as the second condition, but the exemplars were repeated for an unequal number of times, according to their frequency distribution (Starting-High). Since previous research has shown the advantage of staged-input ordering (Lai & Poletiek, 2011; Lai & Poletiek, 2013), we also presented our learning material incrementally in complexity in all conditions.

Experiment

Method

Participants. Sixty-nine students (33 female, mean age 24 year, SD 6.4)[1] from Tilburg University participated for course credit or payment. All were native Dutch speakers. All had normal or corrected to normal vision.

Materials and design. For ease of comparison, we applied the artificial CE Grammar G with the type of A_nB_n in Lai and Poletiek (2011), but generated three novel sets of artificial sequences, which were composed of non-sense syllables. Syllables from Category A [be, bi, de, di, ge, gi] were paired with specific syllables from Category B [po, pu, to, tu, ko, ku] according to their consonants. The pairs were [be/bi-po/pu]; [de/di-to/tu]; [ge/gi-ko/ku]. The complexity differs from 0-, 1-, to 2-LoE. Accordingly, the length of sequences ranged from two-, four-, to six-syllable (e.g. bipo, bebepopo, gebiditopoku). The occurrence number of all possible syllables was balanced in all conditions.

All groups were exposed to a learning set with the same sample size, but crucially, each set contains different items. The original "Starting-Small" (SS) group was trained with 144 items[2] as in Lai and Poletiek (2011). The "Starting-Less" (SL) group was trained with 36 unique items, each presented equally four times. Among these 36 items, there are 12 unique 0-LoE items, 12 1-LoE items and 12 2-LoE items. The "Starting-High" (SH) group also received the same 36 unique items as the SL group, each presented for an unequal number of times, which was determined by their relative frequencies of occurrence. In order to create various relative frequencies, only for the SH condition, we defined three categories of frequency for the basic 0-LoE relations in the grammar: the pairs [be/bi-po/pu] as high frequency; [de/di-to/tu] as medium frequency; [ge/gi-ko/ku] as low frequency. The number of occurrences of an item depended on its defined frequency. For

[1] The data from 7 additional participants could not be included in the analysis, due to computer failure (2), external disturbance (2), and experimental pilots (3).

[2] In Lai and Poletiek (2011), there were 144 items in total, but actually only 108 unique items, since the grammar decided that there were only 12 unique 0-LoE items. Those 108 items consisted of 12 unique 0-LoE, 48 unique 1-LoE and 48 unique 2-LoE. The remaining 36 items were the repetition of 0-LoE. For ease of comparison, we used the same combination.

example, the unequal frequency would result in more occurrences of bepo than geku, and in turn more bebipopu than gegikoku, more bebebipopupu than gegegikokuku etc.

All groups received the same set of 72 test items, half grammatical and half ungrammatical. There were an equal number of test items for each level of complexity (i.e. 0-, 1-, and 2-LoE). The violations were implemented by mismatching an A-syllable with another B-syllable. The numbers of A- and B-syllables were the same for the test sequence in order to avoid salient cues for detecting errors merely by simple strategies, such as counting.

Procedure. Participants were randomly assigned to one of the three groups (23 each). In the training phase, participants were presented with the visual sequences randomly and were informed that there was an underlying rule. Each trial started with a fixation cross (500 ms) and presented the learning sequence syllable-by-syllable (each syllable 800 ms, no in-between interval). In the test phase, participants received a set of new sequences. Participants were instructed to respond whether the test sequences obeyed the same rule as the one in the training phase or not. No feedback on answers was given during the test.

The whole experiment took approximately 35 minutes.

Results

All three groups performed significantly above chance level: M_{SS}= .60, SE_{SS}=.02, t (20) = 5.92, p < .001, r^2 = .64; M_{SL}= .61, SE_{SL}= .02, t (22) = 5.20, p < .001, r^2 = .55; M_{SH}= .67, SE_{SH}=.03, t (21) = 6.82, p < .001, r^2 = .69, respectively, (Figure 1). This shows that in all conditions participants were able to learn processing CE recursion to some extent.

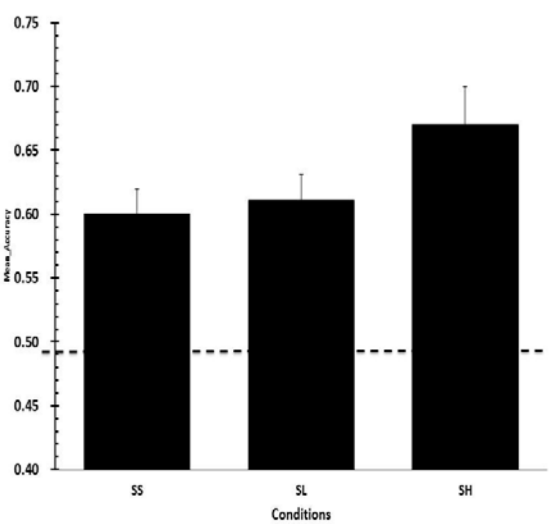

Figure 1. Mean accuracy of all conditions. The dotted line represents chance level (M= .50). Error bars indicate standard error of the mean.

The current results showed a main effect of condition on accuracy, F (2, 63) = 3.44, p < .05, $\eta_p{}^2$ = .10. A post hoc Tukey test showed that the SH group outperformed the SS group significantly (p =.048), while no significant difference between the SH and the SL group (p = .102), nor the SL and the SS group (p = .920).

Next we performed separate analyses for the grammatical and ungrammatical items in the test. When only looking at the grammatical items, there was no main effect of condition on accuracy, F (2, 63) = .21, n.s., (M_{SS}= .66, SE_{SS}= .03; M_{SL}= .67, SE_{SL}= .03; M_{SH}= .69, SE_{SH}= .03). However, regarding the performance on ungrammatical items, there appeared a significant effect of condition on accuracy, F (2, 63) = 5.61, p < .01, $\eta_p{}^2$ = .15. It indicates that these three conditions scored differently when judging ungrammatical items. About the performance on ungrammatical items, a post hoc Tukey test revealed that the SH group (M_{SH}= .66, SE_{SH}= .04) was significantly better than the SS group (M_{SS}= .53, SE_{SS}= .04), p = .009, and also better than the SL group (M_{SL}= .55, SE_{SL}= .03), p = .025. There was no significant difference between the SS and SL group, p = .891.

For the high accuracy of the SH group, we probed into different LoE test items to see whether the effect was due to certain simple and highly frequent ones. An ANOVA indeed showed a main effect of LoE on accuracy, t (63) = 9.22. A post hoc Tukey test indicated that performance on 0-LoE (M= .78, SE= .03) was significantly better than that on 1-LoE (M= .65, SE= .03), p = .010, and also 2-LoE (M= .59, SE= .03), p = .001, though no significant difference between 1- & 2-LoE, p = .498. However, performance on 0-, 1-, and 2-LoE test items were all significantly above chance, respectively t (21) = 8.11, p = .001, t (21) = 4.71, p = .001, and t (21) = 3.18, p = .005. Briefly, participants scored better on simple and more frequent items, but they also showed learning on the more complex and less frequent items.

General Conclusion and Discussion

The current study highlights the influence of sample properties in detecting relational structures of CE recursion. In our study, we first observed that staged-input facilitated learning CE recursion in all three conditions. When participants were exposed to the training input, which was arranged in an increasingly complex manner, they were able to learn hierarchically recursive structures. Moreover, when participants in the SL group received a sample set which contains fewer unique exemplars compared to the original SS group, they were still able to show above-chance performance in the grammaticality-judgment task. This suggests that fewer unique learning exemplars, but with repetition, are at the very least not detrimental to learning center-embedded recursion.

Interestingly, for the SH group, which was provided with a sample set of fewer unique exemplars, which were repeated for an unequal number of times, the performance was significantly boosted.

Our results suggest that we might actually not need a large number of unique learning exemplars when being exposed to a new language construction. Our observations provide empirical supports for the "*fewer is better*" side of the debate (Casasola, 2005; Kersten & Smith, 2002). Our prominent finding is that the diversity of exemplars turned out not to be the most important factor, but the repetition of a smaller set of unique exemplars with differential frequencies stands out as a crucial element. The advantage appears to be twofold: on the one hand, the restricted number of unique exemplars seems more natural in accordance with the cognitive processing window of a human parser. Due to the intrinsic constraints of memory (Christiansen & MacDonald, 2009), people are directed to smaller segments of information more efficiently. On the other hand, the repetition might help to consolidate memory traces. The recurrent exemplars seem to help participants to focus on the pattern and to detect the underlying rule. In this way, the repetition assists people to store and recognize their acquired knowledge.

The distribution of linguistic levels of complexity in natural language is not uniform, but fairly skewed. For instance, child-directed speech contains a large amount of simple and short utterances, but fewer complex or long sentences (Snow, 1972). Also concerning CE recursive sentences, it has been shown that in oral language, people tend to use simpler structures, and that they seldom produce sentences with two or more embeddings (Karlsson, 2010), although proficient language users are able to understand sentences with more embeddings. Compared to other studies, our design is more realistic in the sense that our SH condition simulated these properties of natural language, as discussed above. The SH group was confronted with more basic 0-LoE exemplars (84 items), fewer 1-LoE (45 items) and even fewer 2-LoE (15 items); while the SS and the SL group respectively received 48 0-LoE, 48 1-LoE and 48 2-LoE items. Here, once again we replicated the skewed frequency distribution effect as also observed in Lai and Poletiek (2013).

Note that our results also showed that the SH group primarily benefited for ungrammatical items. To correctly judge on ungrammatical items, participants need to know about the underlying rule that they might form in the learning phase. The SH group might have acquired more confidence in detecting the inherent rule from their set of learning material.

However, while our artificial simulation is ideal in the lab setting, natural language is much more chaotic. Firstly, we created an error-free environment and our training exemplars were all grammatical items. In reality, however, children are exposed to a noisy linguistic environment, full of correct and incorrect examples. When they make grammar mistakes, they might be corrected and given the grammatical examples by parents. Previous studies suggested that this type of negative evidence was a useful part to correct errors in language learning (Schachter, 1991). We would like to further test the influence of (negative) feedback in forthcoming studies. Secondly, our staged-input in all conditions followed a strictly incremental ordering, while in natural language there are more mixtures of different complexities. In future studies, it might be worthwhile testing whether the effect could be replicated without stage-input. In that manner, we would be able to disentangle the influence of incremental ordering from that of the learning sample with fewer exemplars. Thirdly, though we endeavored to mimic the statistical environment at the beginning of language learning, we tested adults with a simplified artificial language. Yet, the ecological validity still needs to be testified. Also, natural language is a sophisticated system, which contains enormous information from a large vocabulary, phonology, syntax, and semantic, etc. Given the complexity of the linguistic properties, it seems extremely difficult for children to discover and extract the underlying grammatical rules precisely. Further testing is needed to verify how well the model could depict children's natural language learning.

Moreover, it would be worthwhile to further test the effect across modalities to see whether the results under various modalities are comparable.

In sum, we found that a limited set of exemplars but with repetition could largely boost the learning of complex structures. Our results shed light on how language learners utilize the statistical properties of the sample set to detect the associated relations of complex hierarchical recursion and the underlying rules in a cognitively efficient and economical way.

References

Behrend, D. A. (1990). The Development of Verb Concepts - Childrens Use of Verbs to Label Familiar and Novel Events. *Child Development, 61*(3), 681-696.

Casasola, M. (2005). When less is more: How infants learn to form an abstract categorical representation of support. *Child Development, 76*(1), 279-290.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts : The MIT Press.

Christiansen, M. H., & MacDonald, M. C. (2009). A Usage-Based Approach to Recursion in Sentence Processing. *Language Learning, 59*, 126-161.

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.

Conway, C. M., Ellefson, M. R., & Christiansen, M. H. (2003). When Less is Less and When Less is More: Starting Small with Staged Input. *Proceedings of the 25th Annual Conference of the Cognitive Science Society*, 270-275, Mahwah, NJ: Lawrence Erlbaum.

Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science, 31*(4), 697-704.

Elman, J. L. (1993). Learning and Development in Neural Networks - the Importance of Starting Small. *Cognition*, *48*(1), 71-99.

Gallivan, J. (1987). Correlates of order of acquisition of motion verbs. *Perceptual and Motor Skills, 64*, 311-318.

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195-236). Cambridge, MA: MIT Press.

Goldowsky, B.N., & Newport, E.L. (1993). Modeling the effects of processing limitations on the acquisition of morphology: the less is more hypothesis. In E. Clark (Ed.), *The proceedings of the 24th annual Child Language Research Forum* (pp. 124-138). Stanford, CA: Center for the Study of Language and Information.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431-436.

Harris, C. (1991). *Parallel distributed processing models and metaphors for language and development*. Ph.D. dissertation, University of California, San Diego.

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*(5598), 1569-1579.

Johnson, M. (1998). Proof Nets and the Complexity of ProcessingCenter-Embedded Constructions. *Journal of Logic, Language and Information, 7*(4), 433-447.

Kareev, Y., Lieberman, L., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology, 126*(3), 278-287.

Karlsson, F. (2010). Syntactic recursion and iteration. In H. v. d. Hulst (Ed.), *Recursion and Human Language* (pp. 43–68). Berlin/New York: Mouton de Gruyter.

Kersten, A. W., & Smith, L. B. (2002). Attention to novel objects during verb learning. *Child Development, 73*(1), 93-109.

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.

Lai, J., & Poletiek, F. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition, 118*(2), 265-273.

Lai, J., & Poletiek, F. H. (2013). How "small" is "starting small" for learning hierarchical centre-embedded structures? *Journal of Cognitive Psychology, 25*(4), 423-435.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science, 11*(4), 628-634.

McAndrews, M. P., & Moscovitch, M. (1985). Rule-based and exemplar-based classification in artificial grammar learning. *Memory & Cognition*, *13*(5), 469-475.

Meulemans, T., & VanderLinden, M. (1997). Associative chunk strength in artificial grammar learning. Journal of Experimental *Psychology-Learning Memory and Cognition, 23*(4), 1007-1028.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91-117.

Morrill, G. (2000). Incremental processing and acceptability. *Computational Linguistics, 26*(3), 319-338.

Newport, E. L. (1990). Maturational Constraints on Language-Learning. *Cognitive Science, 14*(1), 11-28.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical Learning in a Natural Language by 8-Month-Old Infants. *Child Development*, *80*(3), 674-685.

Poletiek, F., & Chater, N. (2006). Grammar induction proWts from representative stimulus sampling *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1968–1973). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Quinn, P. C., Polly, J. L., Furer, M. J., Dobson, V., & Narter, D. B. (2002). Young Infants' Performance in the Object-Variation Version of the Above-Below Categorization Task: A Result of Perceptual Distraction or Conceptual Limitation? *Infancy, 3*(3), 323-347.

Rinaldi, P., Barca, L., & Burani, C. (2004). A database for semantic, grammatical, and frequency properties of the first words acquired by Italian children. *Behavior Research Methods Instruments & Computers, 36*(3), 525-530.

Schachter, J. (1991). Corrective feedback in historical perspective. *Second Language Research, 7*, 89-102.

Snow, C. E. (1972). Mothers Speech to Children Learning Language. *Child Development, 43*(2), 549-565.