# Children Can Use Distributional Cues to Acquire Recursive Structures

**Daoxin Li (daoxinli@sas.upenn.edu)**
Department of Linguistics, University of Pennsylvania
3401-C Walnut Street, Philadelphia, PA 19104 USA

**Kathryn Schuler (kschuler@sas.upenn.edu)**
Department of Linguistics, University of Pennsylvania
3401-C Walnut Street, Philadelphia, PA 19104 USA

## Abstract

This work investigates whether children can use distributional cues from non-embedded examples to learn whether a structure allows recursive embedding. While the ability of recursion is considered universally available, there are considerable cross- and within-linguistic differences regarding the rules for recursive embedding, which must be learned from language-specific experience. One proposal argues that the recursivity of a structure is learnable as a productive generalization from distributional information in non-embedded input, and recent studies have shown that adult learners can indeed use such distributional cues to acquire recursive structures in an artificial language. However, it is not yet known whether children can make use of distributional information in this way. Though children and adults perform similarly in many distributional learning tasks, they are known to behave quite differently in others. In this work, we examine children's distributional learning of recursive structures. We exposed children to non-embedded sentences in an artificial grammar, where we manipulated the productivity of the structure across conditions. At test, we found that children exposed to productive input were more likely to accept recursively embedded test sentences that were unattested during the exposure phase. The results suggest that children can make use of distributional information to acquire recursive structures.

**Keywords:** distributional learning; child language acquisition; recursive structure; syntax; artificial language

## Introduction

Recursion in linguistics refers to the infinite self-embedding of a particular type of linguistic element or structure. While the ability of recursion is considered universally available (Yang, 2013; Berwick & Chomsky, 2017),[1] there are considerable cross- and within- linguistic differences regarding the specific rules for recursive embedding. For instance, in English, the *of*-possessive is more restricted than the *s*-possessive, (1) (see Levi, 1978; Biber et al., 1999; Rosenbach, 2014 for extensive discussion); whereas in German, the *von*(of)-possessive allows free embedding while the *s*-possessive generally cannot embed, (2) (Weiß, 2008; Pérez-Leroux et al., 2022). Therefore, which structures allow recursive embedding must be learned from language specific experience.

(1) a. the man's neighbor's book
  b. *the book of the neighbor of the man
(2) a. *das Manns Nachbarns Buch
    the man's neighbor's book
    'the man's neighbor's book'
  b. das Buch von dem Nachbarn von dem Mann
    the book of the neighbor of the man
    'the book of the neighbor of the man'

However, this poses a difficult learning problem for children: How do children learn rules for infinite embedding from a finite corpus? Children cannot rely on direct observation of recursively embedded examples in the input (e.g., Roeper & Snyder, 2005; Roeper, 2011). Because firstly, corpus studies have found that examples of multi-level embedding are extremely rare in children's input (Giblin et al., 2019); moreover, there is no principled reason why an observed example of a finite level of embedding could entail the possibility of infinite embedding (e.g. witnessing "the man's neighbor's mother" does not entail "the girl's sister's friend's cat's …"). Overall, children need a way to infer the rules for recursive embedding from non-embedded data.

In this study, we investigated whether children can use distributional information from non-embedded input to acquire recursive structures. In the following sections of the Introduction, we will summarize a recent proposal for the distributional learning of recursive structures (Li, Grohe, Schulz & Yang, 2021). We will argue that, though a recent study has demonstrated that adults can indeed learn recursive structures as predicted by the proposal (Li & Schuler, 2023a), more work is needed to determine whether the proposed mechanism is available to children, who are tasked with the challenge of acquiring a first language with more limited cognitive functions than adults. We will test this question through an artificial language learning experiment with 6-8 year-old children.

[1] We are also aware of a long tradition of research on the learning and processing constraints on recursion, such as center embedding (e.g., Roth, 1984; Karlsson, 2007; Christiansen & MacDonald, 2009). Our study, though, does not rely on assumptions of the status of recursion; instead, it explores the learnability problem from a different approach, namely how can recursive structures be learned from non-embedded input data.



In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* 

## The distributional learning proposal

The distributional learning proposal (Li et al., 2021) argues that recursive structures are learnable from distributional cues in non-embedded data from simple child directed speech. First, it argues that recursion derives from structural substitutability: For a structure such as X1's-X2 where X is the head and X1 and X2 stand in a selection relation, it can recursively embed if positions X1 and X2 are productively substitutable, i.e., any word that can appear in one position (X1/X2) can also appear in the other. For example, as demonstrated in Li et al. (2021), for the English *s*-possessive X1's-X2, all nouns used in X1 can be used in X2 as well (denoted as X1↦X2), that is, the possessor can always be possessed, thus allowing infinite embedding to be built in this way (e.g., from "the neighbor's book" to "the man's neighbor's book", etc.).

Li et al., (2021) argue that this way of conceptualizing recursive structures suggests that they can be acquired through distributional learning: If there is sufficient evidence that structural substitutability is generalizable — that is, if a sufficiently large proportion of words attested in one position are also attested in the other position in the input — then children will acquire the generalization that all words that can be used in one position (e.g., X1) can also be used in the other (e.g., X2) and therefore the structure can recursively embed for all words eligible for X1; otherwise, the structure is restricted to certain (types of) words attested in both positions in the input.

Initial support for the proposal comes from corpus studies on a range of structures such as adjectives, possessives and nominal compounds in different languages (Grohe, Schulz & Yang, 2021; Li et al, 2021; Yang, 2022), showing that the proposal can accurately predict the language specific rules given children's realistic input data. These studies demonstrated the availability of the necessary distributional cues in child-directed speech, as a proof a concept for the distributional learning proposal. Critics of the proposal argue that it is necessary to evaluate the proposal on a broader range of natural linguistic phenomena to determine whether the distributional learning mechanism would indeed enable speakers to discover recursive structures; further, it is also necessary to examine whether human learners can utilize the distributional cues as predicted, which is the focus of the present work.

## Adults' distributional learning of recursion

To test the distributional learning proposal against human learning behavior, Li & Schuler (2023a) conducted an artificial language learning experiment with adult participants. They exposed participants to an artificial grammar in the form X1-ka-X2, with 12 different pseudo-words attested in the X1 position. In the Productive condition, nearly all of the 12 words were also attested in the X2 position (10 out of 12); in the Unproductive condition, only some were (6 out of 12). 10/12 and 6/12 were selected because these values were consistent with productivity (or lack of productivity in Unproductive condition) based on several different metrics of productivity (e.g., Aronoff, 1976; Baayen & Lieber, 1991; Bybee, 1995; Yang, 2016). At test, participants were asked to rate on a scale of 1 to 5 the acceptability of one-level (X1-ka-X2) and two-level (X1-ka-X2-ka-X3) attested sentences (i.e., sentences or combinations of two sentences attested during exposure), unattested sentences (i.e., sentences where the post-ka position (X2 or X3) was occupied by a word never attested after ka in the input), and ungrammatical sentences with wrong word order (e.g., ka-X1-X2, ka-X1-X2-X3-ka). The distributional learning proposal predicted that only participants from the Productive condition would learn that any X1 words can also appear in the X2 position (X1↦X2), so they should rate unattested strings higher than participants from Unproductive condition at both one- and two-levels of embedding, although they never heard two-level sentences in the input. As predicted, it was found that participants from the Productive condition generalized significantly more than participants from the Unproductive condition at both embedding levels.

## The current study

Li & Schuler (2023a) demonstrated that adult participants are able to learn recursive structures from distributional information, and argued that this provides further evidence for the distributional learning proposal as a mechanism for language acquisition. However, an important open question is whether younger learners can also fully utilize such distributional information. On one hand, many studies have shown that young children and even infants can acquire linguistic knowledge using statistical information in similar ways (e.g., Saffran, Aslin & Newport, 1996; Marcus, Vijayan, Rao & Vishton, 1999; Shi & Emond, 2023); some have even argued that distributional learning is an ability available from birth (e.g., Gervain et al., 2008; Teinonen et al., 2009; Aslin, 2017). On the other hand, still in the process of development, children are much more limited than adults in a range of cognitive functions such as memory and processing abilities (e.g., Thiessen, 2011; Santolin & Saffran, 2018), and their learning outcomes often differ from adults in first and second language acquisition (Johnson & Newport, 1989; Newport, 1990; Mayberry & Kluender, 2018) as well as in artificial language learning experiments (Weir, 1964; Hudson Kam & Newport, 2005; Austin, Schuler, Furlong & Newport, 2022). Therefore, it is crucial to examine whether young learners exploit the subtle distributional cues in the same way as adults. In the current study, we set to test children's distributional learning of recursive structures in an artificial language.

# Experiment

In this experiment, we examined whether children can use distributional cues to learn which structures allow recursive embedding. The general design was similar to Li & Schuler (2023a) with modifications to make it more children-friendly.

## Methods

**Participants** Participants were 17 children aged 6 to 8 years.[2] We chose this age range because prior work on similar paradigms has shown that children of this age were able to use purely distributional cues to acquire linguistic rules in artificial language learning experiments (e.g., Schuler et al., under review). The children were all native English speakers. 10 of these children were assigned to the Productive condition and 7 to the Unproductive condition. An additional 14 children participated but were excluded from analysis based on our exclusion criteria that we will specify in the Results section.[3] In brief, they failed to learn the basics of the artificial grammar or understand the task. We recruited the children through sharing our advisement online. They participated in the experiment over Zoom and received a $10 Amazon gift card as compensation.

**Stimuli** We generated stimuli sentences from the artificial language used by Li & Schuler (2023a). Originally adapted from Ruskin (2014), the pseudo-words were all bi-syllabic words that conformed to English phonotactics. In the language, sentences are formed X1-ka-X2, and all 9 X-category words were attested in the X1 position. Crucially, we manipulated whether there was sufficient evidence to form a productive generalization of structural substitutability. In the Productive condition, X1 and X2 were productively substitutable: 6 out of the 9 words were also attested in the X2 position; in the Unproductive condition, only 4 of the 9 words were. Based on several different metrics of productivity (Aronoff, 1976; Baayen & Lieber, 1991; Bybee, 1995; Yang, 2016), 6 out of 9 is productive (sufficient evidence to learn the generalization that all of the 9 words used in X1 can also be used in X2 even though they were never attested in X2 position) but 4 out of 9 is not. We decided to use a total of 9 words because our pilot studies showed that this amount of input is feasible for children to learn in a single-day artificial language learning experiment. Therefore, there was a total of 54 unique sentences in the Productive condition, and 36 in the Unproductive condition. The whole corpus was repeated twice during the exposure phase for the Productive condition and three times for the Unproductive condition so that all children heard 108 sentences. The word distribution in each repetition is shown in Table 1.

Similar to Li & Schuler (2023a), the test sentences had either 1-level of embedding ("nogi-ka-sito") or 2-levels of embedding ("mito-ka-sito-ka-tana"), with the prediction being that children could decide the recursivity for 2-level sentences based on substitutability in 1-level sentences. The 1-level sentences tested the children's knowledge of substitutability: Whether all words used in the X1 position could also be used in the X2 position; the 2-level sentences tested their knowledge of recursion. Next, at each embedding level, there were attested, unattested, and ungrammatical sentences (Table 2). Attested sentences meant that the whole sentence (for 1-level sentences) or both parts of the sentence (for 2-level sentences, e.g., both "mito-ka-sito" and "sito-ka-tana" in the 2-level unattested sentence in Table 2) were attested during the exposure phase. In unattested sentences, at least one word following ka was never attested after ka (the X2 position) during the exposure phase. Ungrammatical sentences had completely wrong word order. Children in both conditions were predicted to rate attested sentences high and ungrammatical sentences low at both levels of embedding if they have learned the basic structure of the grammar (e.g., they know X1-ka-X2 sentences are grammatical and ka-X1-X2 — sentences with a completely wrong word order — were not). Crucially, the unattested test sentences allow us to determine whether children have indeed formed the productive generalization that the X1 and X2 positions are substitutable (all the words in the X1 position can also appear in the X2 position). We predicted that only children in the Productive condition would have sufficient evidence to learn that all words used in the X1 position can also be used in the X2 position (even though some were never attested in X2 position in the input). Furthermore, given the productive substitutability in 1-level data, only children in the Productive condition would be predicted to acquire the generalization that X1↦X2 holds for any embedding level to create recursive embedding. Therefore, unattested sentences were predicted to be treated more similarly to attested sentences at both embedding levels in the Productive condition, and more similarly to ungrammatical sentences at both embedding levels in the Unproductive condition.

Table 1: Word distribution in the exposure corpus.

| Word | Productive | | Unproductive | |
|---|---|---|---|---|
| | X1 | X2 | X1 | X2 |
| kewa | 6 | 9 | 4 | 9 |
| tana | 6 | 9 | 4 | 9 |
| sito | 6 | 9 | 4 | 9 |
| bila | 6 | 9 | 4 | 9 |
| tesa | 6 | 9 | 4 | 0 |
| mito | 6 | 9 | 4 | 0 |
| nogi | 6 | 0 | 4 | 0 |
| seta | 6 | 0 | 4 | 0 |
| waso | 6 | 0 | 4 | 0 |
| Total | 54 | 54 | 36 | 36 |

[2] The work is ongoing, and we plan for 20 children in each condition.

[3] The exclusion rate might seem high. However, a special property of the current experiment is that it depends on purely distributional learning with no semantic world. Therefore, it is more similar to infant studies, where exclusion rates have been similarly high (e.g., Aslin, Saffran & Newport, 1998; Shi & Emond, 2023).

The audio of the stimuli sentences was generated by a female voice using an online speech synthesizer, Natural Reader, with a speed of 120 WPM.

Table 2: Sample test sentences.

| Type | One-level | Two-level |
|---|---|---|
| attested | *nogi-ka-sito* | *mito-ka-sito-ka-tana* |
| unattested | *waso-ka-seta* | *waso-ka-nogi-ka-seta* |
| ungrammatical | *ka-seta-mito* | *ka-ka-kewa-bila-waso* |

**Procedure** The experiment was presented as a video game so that it would be more attractive to children. The video game paradigm was adapted from Schwab et al. (2016) and Schuler et al. (under review), which successfully worked for children in distributional learning experiments. At the beginning of the game, a cartoon robot explained that an alien, Zooma, was traveling to another planet, and that the goal of the game was to help Zooma learn the new language "Zilly" which is spoken on that planet. Next, there were some practice trials where children were asked to decide how well Zooma was speaking English using a slider scale from "no" to "yes" corresponding to values of 0 to 100, and an experimenter would provide feedback on how to use the slider scale. This phase was to familiarize the children with the task and the rating scale.

Then, during the exposure phase, children were instructed to listen carefully as Zooma practiced saying sentences in the new language. To keep children attentive, Zooma would get tired and stop practicing at several points in the exposure phase, and children were asked to click on Zooma to wake her up. Every time children clicked on Zooma, they received a star; when they collected enough stars, Zooma would progress in her journey to the new planet. A screenshot of the exposure phase is shown in Figure 1.


Figure 1: Screenshot of the exposure phase.

When children heard all the exposure sentences, Zooma arrived at the distant planet, and the test phase began. On each test trial, Zooma said a test sentence, and children were asked to decide whether they heard this sentence during the exposure phase, using the same slider scale that they used in the practice trials. The 1-level sentences and 2-level sentences were presented in different blocks, with the 1-level sentences presented first. The level 2 instructions were modified slightly, to acknowledge that they would be longer: "Now Zooma will say sentences in Zilly that are longer. Some will be good. Some will be bad. Again, your job is to decide how well Zooma is speaking Zilly." A screenshot of the test phase is shown in Figure 2.

Figure 2: Screenshot of the test phase.

## Results

We excluded children whose meaning rating for ungrammatical sentences was higher than attested sentences at any embedding level, which indicated not learning the fundamentals of the artificial grammar and/or not understanding the task. Results from the remaining children are shown in Figure 3. First, as expected, in both conditions and at both embedding levels, attested sentences were rated high and ungrammatical sentences were rated low. However, crucial for our prediction, the rating for unattested sentences differed significantly across conditions. At both levels of embedding, unattested sentences were rated higher in the Productive condition than the Unproductive condition.

We analyzed the results with mixed effects regression using the *lmerTest* package (Kuznetsova et al., 2017) in *R*. The dependent variable was the rating score. Fixed effects included Condition (Productive vs. Unproductive), Type (attested, unattested, ungrammatical) and Level (1-level vs. 2-level) (in a three-way interaction). All the categorical predictors were simple coded. Participant was included as random intercept to account for by-participant variance. Hierarchical modeling showed that Type (attested, unattested, ungrammatical) ($X^2(2) = 186.54$, $p < 0.001$), Level ($X^2(1) = 7.00$, $p < 0.01$), the interaction between Condition and Type ($X^2(2) = 22.68$, $p < 0.001$), and the interaction between Type and Level ($X^2(2) = 6.97$, $p = 0.03$) were significant predictors of the rating score, but not Condition or any other possible interactions between Condition, Type and Level. The statistics from the regression model are shown in Table 3. Crucially, as predicted, the unattested sentences in the Unproductive condition were rated significantly lower, suggesting that children were less willing to allow unattested sentences when there was not enough evidence for productive substitutability in non-embedded input.

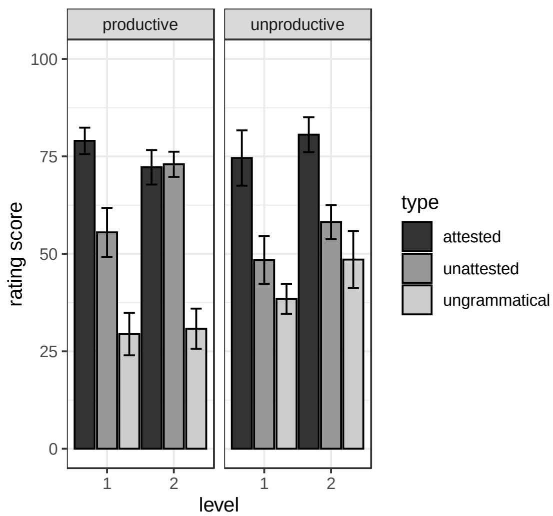


Figure 3: Mean rating scores by embedding level and test sentence type in each condition.

Table 3: Statistics from the regression model.

| Fixed effects | $\beta$ | *SE* | *t* | *p* |
|---|---|---|---|---|
| (Intercept) | 57.28 | 2.57 | 22.31 | <0.001*** |
| Condition – unproductive | 1.05 | 5.14 | 0.20 | 0.84 |
| Type – unattested | -18.73 | 2.78 | -6.73 | <0.001*** |
| Type – ungrammatical | -39.23 | 2.86 | -13.70 | <0.001*** |
| Level – level-2 | 5.47 | 2.25 | 2.43 | 0.02* |
| Condition × Type – unproductive × unattested | -13.56 | 5.57 | -2.44 | 0.02* |
| Condition × Type – unproductive × ungrammatical | 10.73 | 5.73 | 1.87 | 0.06 |
| Condition × Level – unproductive × level-2 | 4.44 | 4.50 | 0.99 | 0.32 |
| Type × Level – unattested × level-2 | 12.19 | 5.57 | 2.19 | 0.03* |
| Type × Level – ungrammatical × level-2 | 5.36 | 5.73 | 0.94 | 0.35 |
| Condition × Type × Level – unproductive × unattested × level-2 | -21.72 | 11.13 | -1.95 | 0.05 |
| Condition × Type × Level – unproductive × ungrammatical × level-2 | -3.310 | 11.46 | -0.29 | 0.77 |

Post-hoc comparisons using the *emmeans* package (Lenth, 2020) further revealed that at 1-level, the rating scores for attested, unattested and ungrammatical sentences all differed from each other in the Productive condition ($p$ < 0.01 for all comparisons), while in the Unproductive condition, the unattested and ungrammatical sentences did not differ from each other ($\beta$ =3.61, $SE$ = 4.86, $t$ = 0.74, $p$ = 1.00), and they were both rated lower than the attested sentences ($p$ < 0.001 for both comparisons); at 2-level, in the Productive condition the unattested sentences did not differ from the attested sentences ($\beta$ = 4.93, $SE$ = 4.60, $t$ = 1.07, $p$ = 1.00), both rated higher than the ungrammatical sentences ($p$ < 0.001 for both comparisons), whereas in the Unproductive condition, the unattested sentences patterned with the ungrammatical sentences ($\beta$ = 12.07, $SE$ = 4.67, $t$ = 2.58, $p$ = 0.29). These results align with what we see in Figure 3. Overall, the results showed that the unattested sentences — sentences that did not occur in the language input — were rated lower in the Unproductive condition than the Productive condition. In particular, although children never heard recursively embedded sentences in the learning phase, children in the Productive condition judged 2-level unattested sentences to be similarly well-formed to attested sentences; by contrast, children from the Unproductive condition regarded 2-level unattested sentences essentially as ungrammatical sentences. Therefore, the results suggested that children can indeed use distributional cues to acquire recursive structures.

## General Discussion

In this study, we asked whether children can use distributional cues from non-embedded examples to learn whether a structure allows recursive embedding. Previous work has shown that adults can acquire recursive structures through distributional learning, but we argued that it is necessary to examine whether children can also do it in order to determine whether such a learning mechanism could be helpful during child language acquisition.

Through an artificial language learning experiment, we demonstrated that children can indeed acquire recursive structures using distributional information: Children who received sufficient input for structural substitutability rated unattested embedded sentences higher than children from the other condition, although no recursively embedded sentences were attested in their input. Importantly, children who received productive input rated unattested embedded sentences as high as attested ones, while for children who received unproductive input, the unattested embedded sentences patterned with ungrammatical ones. Overall, the results suggest that children are sensitive to distributional cues on structural substitutability in non-embedded data, and can use these cues to determine whether the structure permits recursive embedding.

Taken together, this work demonstrated that the ability to track and utilize sophisticated and subtle distributional information to discover the underlying rules is available from an early age. Therefore, it could be a useful mechanism that helps children with the extremely challenging task of language acquisition. With this claim, we do not intend to suggest that distributional cues are the only kind of information that is helpful or that distributional learning can give the children everything. In natural language acquisition, we agree that other factors such as semantic, pragmatic and phonetic cues will also play a role in the acquisition of recursive structures (e.g., Rosenbach, 2014), and that future research should investigate how these different cues are exploited and coordinated. On the other hand, our research suggests that children can learn important rules for recursive embedding from purely distributional information, even when other cues may not be accessible or completely reliable (Marastos & Chalkley, 1980; Braine, 1987).

While it has been found that both children and adults can acquire recursive structures through distributional learning, children's learning behavior in our experiment also showed some differences from adults' behavior in Li & Schuler (2023a). Compared to adults, children exhibited categorical generalization, i.e., they rated unattested 2-level sentences as high as attested ones in the Productive condition, and as low as ungrammatical ones in the Unproductive condition; whereas for adults, although they did generalize more in the Productive condition, their rating scores for unattested 2-level sentences were still significantly lower than attested ones.

Though more work is needed to compare children and adults directly, this observation is perhaps unsurprising given other studies noting differences between children and adults generalization behavior. First, studies on the acquisition of regular rules have found that children are more likely to form categorical generalizations. For example, in Berko's (1958) study where participants were asked to produce the past tense form of pseudo-verbs ending with "-ing" such as "gling", most adults produced an irregular form "glang", following the irregular pattern in English (e.g., "sing, sang"; "ring, rang"), but children predominantly produced the regular form "glinged". And in Schuler, Yang & Newport (2016), when participants learned a noun plural rule in an artificial language, almost all the children applied it to either *all* or *none* of the novel words during test depending on the productivity of their input; by contrast, adults from both conditions matched the token frequency of the plural markers from the input. Furthermore, in another line of work on learning from input that contains inconsistent use of grammatical forms, in both natural language acquisition and artificial language learning experiments, children have been observed to regularize, i.e., to produce these forms more consistently, whereas adults tend to reproduce the inconsistencies: For instance, when there is more than one grammatical form in variational use, adults closely reproduce the probabilistic patterns, while young children only use the most consistent form almost all the time (e.g., Singleton & Newport, 2004; Hudson Kam & Newport, 2005, 2009). In general, children seem to be particularly inclined to form categorical rules.

One important and open question from our work is children's structural representation for the artificial grammar in our experiment. According to the distributional learning proposal, it is important that the X word is the head of the structure. For example, in the English *s*-possessive N1's-N2, N2 is the head, e.g., "the kid's book" is essentially an instance of "book". This notion of the head establishes an equivalence relation between a head noun and all syntactic objects headed by that noun of the phrase, and therefore guarantees recursion. Otherwise, structural substitutability itself would not necessarily lead to recursion. For instance, in English NP1-V-NP2 structures (e.g., "dogs chase cats"), neither of the two NPs is the head of the structure. Therefore, even though NP1 and NP2 can be substitutable in this case, learners would not learn this as a recursive structure (e.g., "*dogs chase cats chase rats"). Given our current design, we recognize it is possible for children to represent the grammar as a linear language, and we agree that more work is needed to determine the role of structural representation in children's distributional learning of recursive structures. In another study on adults' distributional learning recursive structures, Li & Schuler (2023b) presented participants with distributional information not just on structural substitutability but also on headedness: In their A1-B-A2 grammar, A1 and A2 were productively substitutable, but distributional cues suggested that A2 was the head in one condition and B was the head in the other condition. Participants were found to integrate the information, only allowing recursive embedding in the A-head language, where the results were similar to those in the Productive condition from Li & Schuler (2023a). Given this, we suggest that it is likely that children treated the artificial grammar as a headed language where X was the head; and even if some children learned a linear structure, it is likely that they would also be able to learn a headed recursive structure with the same mechanism provided information for headedness. And in ongoing work, we are conducting similar experiments with children to explicitly test their learning behavior given a headed language.

In conclusion, our results suggest that as adults, children can also use distributional cues in principled ways to acquire recursive structures, even when there was no accompanying semantic world. When there was sufficient information supporting structural substitutability in non-embedded data, children would be willing to allow this structure to be recursively embedded even though embedded examples were never attested in their input. On the other hand, children also exhibited interesting and subtle differences from adults which invite further investigation.

## Acknowledgments

Funding for this work was provided by the University of Pennsylvania to K. Schuler. Thank you to our participants and their family for participating in the experiment; to the lab managers and research assistants at the Child Language Lab of the University of Pennsylvania for help with data collection; to Charles Yang and members of the Child Language Lab and the Language and Cognition Lab at the University of Pennsylvania for helpful discussion; and to CogSci reviewers whose comments improved the paper.

## References


Aronoff, M. (1976). *Word formation in generative grammar*. MIT Press.

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *WIREs Cognitive Science, 8*, e1373.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321-324.

Austin, A. C., Schuler, K. D., Furlong, S., & Newport, E. L. (2022). Learning a language from inconsistent input: Regularization in child and adult learners. *Language Learning and Development*, *18*(3), 249-277.

Baayen, H., & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. *Linguistics, 29*(5), 801-843.

Berko, J. (1958). The child's learning of English morphology. *WORD, 14*(2-3), 150-177.

Berwick, R., & Chomsky, N. (2017). *Why only us*. MIT Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Braine, M. D. S. (1987). What is learned in acquiring word classes – A step toward an acquisition theory. In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* Lawrence Erlbaum Associates.

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425-455.

Christianson, M. H. & MacDonald, M. C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning, 59*, 126-161.

Gervain, J., Macagno, F., Cogoi, S., Pena, M., & Mehler, J. (2008). The neonate brain detects speech structure. *The Proceedings of the National Academy of Sciences*, *105*, 14222-14227.

Giblin, I., Zhou, P., Bill, C., Shi, J., & Crain, S. (2019). The spontaneous eMERGEnce of recursion in child language. In *Proceedings of the 43rd annual Boston University Conference on Language Development* (pp. 270-286). Cascadilla Press.

Grohe, L., Schulz, P., & Yang, C. (2021). *How to learn recursive rules: Productivity of prenominal adjective stacking in English and German.* Paper presented at the 9th bi-annual conference on Generative Approaches to Language Acquisition – North America.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*(2), 151-195.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*(1), 30-66.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

Karlsson, F. (2007). Constraints on multiple center embedding of clauses. *Journal of Linguist, 43*, 365–392.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26.

Lenth, R. V. (2020). emmeans: Estimated marginal means, aka least-squares means. R package version 1.5.3. https://CRAN.R-project.org/package=emmeans

Levi, J. N. (1978). *The syntax and semantics of complex nominals.* Academic Press.

Li, D., Grohe, G., Schulz, P., & Yang, C. (2021). The distributional learning of recursive structures. In *Proceedings of the 45th annual Boston University Conference on Language Development* (pp. 471-485). Somerville, MA: Cascadilla Press.

Li, D., & Schuler, K. D. (2023a). Acquiring recursive structures through distributional learning. *Language Acquisition, 30*(3-4), 323-336.

Li, D., & Schuler, K. (2023b). Distributional learning of recursive structures: The role of the structural representation. *Proceedings of the 47th annual Boston University Conference on Language Development* (pp. 487-498). Somerville, MA: Cascadilla Press.

Maratsos, M. P., & Chalkley, M. A. (1980). The internal language of children's syntax: The nature and ontogenesis of syntactic categories. In K. Nelson (Ed.), *Children's language* (Vol. 2). Gardner Press.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven month-old infants. *Science, 283*(5398), 77-80.

Mayberry, R. I., & Kluender, R. (2018). Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism: Language and Cognition, 21*(5), 886-905.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive science, 14*(1), 11-28.

Pérez-Leroux, A., Roberge, Y., Schulz, P., & Lowles, A. (2022). Structural diversity does not affect the acquisition of recursion: The case of possession in German. *Language Acquisition*, *29*, 54-78.

Roeper, T. (2011). The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics, 5*(1-2), 57-86.

Roeper, T., & Snyder, W. (2005). Language learnability and the forms of recursion. In A. M. DiScullo (ed.), *UG and external systems: Language, brain and computation* (pp. 155-169). John Benjamins.

Rosenbach, A. (2014). English genitive variation – The state of the art. *English Language and Linguistics*, *18,* 215-262.

Roth, F. P. (1984). Accelerating language learning in young children. *Child Language, 11,* 89–107.
Ruskin, D. (2014). *Cognitive influences on the evolution of new languages.* Doctoral dissertation, Department of Brain and Cognitive Sciences, University of Rochester.
Santolin, C., & Saffran, J. R. (2018). Constraints on statistical learning across species. *Trends in Cognitive Sciences, 22*(1), 52-63.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.
Schuler, K. D., Lukens, K., Reeder, P. A., Newport, E. L., & Aslin, R. N. (under review). Children can use distributional cues to acquire grammatical categories. *Open Mind.*
Schuler, K. D., Yang, C., & Newport, E. L. (2016). Testing the Tolerance Principle: Children form productive rules when it is computationally more efficient to do so. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2321–2326). Austin, TX: Cognitive Science Society.
Schwab, J. F., Schuler, K. D., Stillman, C. M., Newport, E. L., Howard, J. H., & Howard, D. V. (2016). Aging and the statistical learning of grammatical form classes. *Psychology and Aging, 31*(5), 481-487.
Shi, R., & Emond, E. (2023). The threshold of rule productivity in infants. *Frontiers in Psychology, 14*, 1251124.
Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American sign language from inconsistent input. *Cognitive Psychology, 49*(4), 370–407.
Teinonen, T., Fellman, V., Naatanen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience, 10,* 21.
Thiessen, E. D. (2011). Domain general constraints on statistical learning. *Child Development, 82*(2), 462-470.
Weir, M. W. (1964). Developmental changes in problem-solving strategies. *Psychological Review, 71*(6), 473.
Weiß, H. (2008). The possessor that appears twice? Variation, structure and function of possessive doubling in German. In *Microvariation in syntactic doubling* (pp. 381-401). Emerald.
Yang, C. (2013). Ontogeny and phylogeny of language. *The Proceedings of the National Academy of Sciences, 110,* 6324-6327.
Yang, C. (2016). *The price of linguistic productivity: How children learn to break rules of language*. MIT Press.
Yang, C. (2022). *Productivity and recursion in English compounding.* Poster presented at the 37th annual Boston University Conference on Language Development.