Distributional learning of recursive structures is constrained by structural representation

Daoxin Li[a,b,*] & Kathryn D. Schuler[a]

[a]Department of Linguistics, University of Pennsylvania
[b]Current address: Institute for Policy Research & Department of Psychology,
Northwestern University
[*]Corresponding author at: daoxinli@sas.upenn.edu

Abstract:
This work investigates the mechanism of learning recursive structures. The ability of recursion is considered crucial for language and universally available, but there are considerable with- and cross-linguistic differences regarding the rules for recursive embedding, which poses a learnability problem. Previous research has shown that adults can learn the recursivity of linear structures in an artificial language from distributional cues in non-recursively embedded data. However, an important open question is participants' structural representation of the grammar, which is considered crucial for linguistic recursion. In this study, we examined the hypothesis that representation of the structural head is necessary for the distributional learning of recursive structures. Adult participants were exposed to one of the two artificial languages which had identical linear orders but different heads. At test, participants were asked to rate test strings which examined their knowledge of the head and recursion. As predicted, we found that the learning mechanism is constrained: participants who learned the language with the correct head were more likely to allow recursive embedding. The findings suggest that human learners need structural representations beyond surface-level distributional cues to acquire recursive structures.

1. Introduction

A fundamental property of human language is its capacity to create an infinite number of sentences using a finite set of words and rules. This capacity is in part due to *recursion*, the infinite self-embedding of a particular type of linguistic element or structure. For instance, in the 's'-possessive example in (1a), noun phrases are recursively embedded inside noun phrases: 'book' is embedded in 'neighbor's book', which is further embedded in 'man's neighbor's book', and this process can continue infinitely.

(1) a. the man's neighbor's book
b. *the book of the neighbor of the man

The ability of recursion is considered crucial for language and universally available to human learners (e.g., Berwick & Chomsky, 2017).[1] However, this ability cannot be applied freely, since

---

[1] There are also long lines of work on whether recursion is a property of the human mind or an historical achievement (e.g., Goldin-Meadow, 1982; Deutscher, 2000; Kocab et al., 2023), and on the learning and

there are often constraints on the rules for recursive embedding given a specific structure. For example, while both the ‘s’-possessive and ‘of’-possessive in English can express the general ‘possession’ meaning, it is agreed that the ‘of’-possessive is much more limited than the ‘s’-possessive when expressing ownership (alienable possession), (1) (e.g., Levi, 1978; Biber et al., 1999; Rosenbach, 2014). Furthermore, such knowledge cannot be universal or innate given the cross-linguistic differences. For instance, different from the English examples above, the ‘s’-possessive in German can only be used with a narrow set of words such as proper names and some kinship terms and usually cannot embed at all (Weiß, 2008; Pérez-Leroux et al., 2022), while the possessive with the preposition ‘von’ (‘of’) is almost freely recursive, (2).

(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’

Beyond the case of possessives, there are many other well-documented cross-linguistic differences regarding the rules for recursive embedding (e.g., Bauer, 1978; Everett, 2005; Roeper, 2011; Haspelmath, 2016). Therefore, despite the probable universality of recursion, learners need to learn from language-specific experience whether a structure allows recursive embedding.

One prominent proposal suggests that children learn which structures allow recursion by witnessing examples of recursive embedding in their input (Roeper & Snyder, 2004; Roeper, 2007, 2011). However, corpus studies show that multiple embeddings are extremely rare in children’s input (e.g., Pérez-Leroux et al., 2018; Giblin et al., 2019; Li, 2024), as are many complex linguistic structures (Jelinek, 1998; Yang, 2013). Moreover, even if children do encounter recursive embedding in their input, there is no clear explanation for how this exposure would generalize to allow deeper embeddings. In other words: how do children learn that a structure allows infinite embedding, from examples in their input which are always of a finite length?

Recently, an alternative account proposed that rules for recursive embedding are learnable from distributional cues in non-recursively-embedded input (Li, Grohe, Schulz & Yang, 2021). A long line of research has demonstrated that human learners can track and use distributional cues to acquire linguistic knowledge (e.g., Maratsos & Chalkley, 1980; Braine, 1987; Mintz, 2003; Thompson & Newport, 2007; Reeder, Newport & Aslin, 2013; Schuler, Reeder, Newport & Aslin, 2017). Given this, Li et al (2021) proposed that recursion can be viewed as *structural substitutability*: a structure like ‘X_1’s-X_2’ is recursive if the elements in positions X_1 and X_2 are

---

processing constraints on recursion (e.g., Roth, 1984; Karlsson, 2007; Christiansen & MacDonald, 2009), but these are not the focus in the present study.

productively substitutable. In other words, recursion is licensed when any word that appears in X_1 can also appear in X_2.[2] For example, if a learner encounters 20 nouns in the X_1 position, and all can also appear in X_2, then the structure allows infinite recursion (e.g., from "the neighbor's book" to "the man's neighbor's book", etc.). This means that a possessor can always be possessed, enabling infinite recursive embedding.

Li et al. (2021) further argued that this way of conceptualizing recursive structures allows them to be acquired through distributional learning: While the fact that a word *can* appear in a position does not guarantee that it will actually be attested in that position in a child's input, children can learn substitutability as a productive generalization if a *sufficiently high* proportion of words in their input are attested to follow the pattern. For example, returning to the toy grammar where 20 different nouns are used in the X_1 position: if 18 out of these nouns (though not all 20 of them) are also attested in the X_2 position, then children will still learn the generalization of substitutability (and thereby recursion) and will allow the 2 unattested words to also be used in the X_2 position and participate in recursion. Otherwise, if only a small proportion of words are attested in both positions in the input, then children will limit substitutability and recursion to the words that they have witnessed in both positions, but will not generalize to novel instances.

Corpus studies on a variety of structures such as possessives, nominal compounds, and serial verb constructions across languages have confirmed that such distributional cues on structural substitutability are available in child-directed speech (Grohe, Schulz & Yang, 2021; Li et al, 2021; Yang, 2022; Li, 2024). Moreover, a previous artificial language learning experiment demonstrated that learners can indeed use this distributional information to determine whether a structure allows recursive embedding, even from non-recursively-embedded input. Li and Schuler (2023) exposed adult participants to X_1-ka-X_2 strings in an artificial grammar, where they manipulated whether there was sufficient evidence for structural substitutability in the exposure. In the Productive condition, nearly all the words attested in the X_1 position were also attested in the X_2 position (10 out of 12); in the Unproductive condition, only some were (6 out of 12). At test, only participants from the Productive condition allowed recursive embedding (X_1-ka-X_2-ka-X_3), even though they were never exposed to recursively embedded examples; By contrast, participants from the Unproductive condition treated these recursively embedded strings as ungrammatical.

Crucially, the distributional learning proposal (Li et al., 2021) argues that recursion is not freely applied but is constrained to operate over *heads* of phrases. The concept of heads is fundamental in linguistics. The head determines the syntactic category and basic meaning of a phrase or word. For example, in 'big red flower', 'flower' is the head, which determines that this phrase is a noun

---

[2] This is not a theoretical proposal *why* some structures allow recursive embedding and some do not. We are aware of theoretical studies on this *why* problem (e.g., Hartmann & Zimmermann, 2002; Adger, 2003; Arsenijević & Hinzen, 2012) and our approach is not inconsistent with these, but we are focusing on a different perspective, i.e., how to learn the rules for recursive embedding from realistic input data.

phrase, not an adjective phrase. Similarly, in ‘the neighbor’s book’, ‘book’ is the head, since ‘the neighbor’s book’ is essentially a kind of book, not a kind of neighbor. The notion of headedness also has parallels in cognitive science, though often described differently (e.g., Miller, 1956; Rosch, 1973; Murphy, 2002). This restriction is important for the distributional learning proposal (Li et al., 2021) because it prevents overgeneralization, ensuring that learners only extend recursion where it is structurally valid. For example, the two NPs in ‘NP_1-V-NP_2’ in English can be substitutable (e.g., many words like ‘cats’ can appear in both NP_1 and NP_2 positions), but the structure cannot recursively embed (e.g., ‘*dogs chase cats chase rats…’) because neither NP is the head (‘dogs chase cats’ is not a kind of dog or cat) and therefore the self-embedding definition of recursion is not satisfied. Consequently, this *linear* structural substitutability will not guarantee recursion.

Therefore, an important open question is whether the learning mechanism itself is constrained by head information. In previous artificial language learning experiments, what is participants’ *structural representation* for the artificial grammar: Do they represent the head correctly? Can they indeed utilize the head information during the distributional learning of recursive structures? In the present study, we examine this question by exposing adult participants to two artificial languages that both allowed structural substitutability in linear positions but differed in their head. To preview the results, we found that, as predicted, participants in the head-substitutability condition were more willing to allow recursive embedding, suggesting that both substitutability of linear positions and higher-order representation of the head are important for the acquisition of recursive structures.

2. Methods

2.1. Participants

Participants were 50 adult native English speakers with typical hearing and vision (or corrected vision) and no documented language disorders. All participants were recruited and run online via Prolific Academic (https://www.prolific.com) and paid $9/hour as compensation. The 50 participants were evenly assigned to two language conditions, A-head language (mean age = 31.2, range = 20-46) and B-head language (mean age = 29.5, range = 20-45).

2.2. Stimuli

The artificial grammar was modeled after Li & Schuler (2023) and consists of two categories: A and B. The A category contains 12 pseudowords, while the B category contains one. From this basic structure, we generated sentences to create two languages, each providing two types of distributional cues: (1) cues to structural substitutability and (2) cues to the head.

To provide distributional cues to structural substitutability, we generated three-word strings of the form A_1-B-A_2, where A_1 and A_2 are positions that can be filled by any A category word. In both language conditions, all 12 A-words were attested in the A_1 position, and 9 of those were also attested in the A$_2$ position. According to several metrics of productivity, this distribution provides sufficient evidence for the productive generalization that all words appearing in the A$_1$ position could also appear in the A$_2$ position (e.g., Bybee, 1995; Yang, 2016). Crucially, these cues to structural substitutability were the same in both language conditions, and neither language contained examples of multiple embeddings.

Turning next to cues to the head, recall that the proposal argues that learners will apply the learning mechanism selectively to the head of the phrase. To test this constraint, we included distributional cues to the structural head, which were conveyed via shorter one- and two- word sentences. To approximate the distribution of heads in natural language, we borrowed an approach used by Thompson & Newport (2007) and Fetch (2020) in which (1) heads are obligatory while other elements are optional, and (2) heads appear consistently in the same linear position.[3] In addition, we used syllable length as a further cue for headedness: Heads contain two syllables while non-heads contain one.[4] In this way, we created two distinct language conditions with different possible one- and two-word strings (Table 1): one in which A$_2$ was the head (A-head language) and one in which B was (B-head language). Thus, both languages shared the same A$_1$-B-A$_2$ strings (and therefore the same cues to structural substitutability over the A$_1$ and A$_2$ positions), but only in the A-head language were those cues available over the phrasal head. If learners can indeed learn recursive structures based on distributional cues to both substitutability *and* headedness, they should be more likely to license recursion (A$_1$-B-A$_2$-B-A$_3$) in the A-head language. If the learning mechanism has no such head constraint, then participants should license recursion in both conditions.

	A-head language	B-head language
one-word	**A**, *B	*A, **B**
two-word	A$_1$-$\mathbf{A_2}$, B-$\mathbf{A_2}$, *A$_1$-B	*A$_1$-A$_2$, **B**-A$_2$, A$_1$-**B**
one-level	A$_1$-B-$\mathbf{A_2}$	A$_1$-**B**-A$_2$

Table 1. Headedness and grammaticality of strings of different lengths in each language. Head is bolded. Ungrammatical strings are indicated with an asterisk.

---

[3] These characterizations of heads do not include all of the features that define heads; and they may not always apply in all languages. We chose to use these rules because they include key features that define heads in the theoretical work on natural languages and have proven useful in previous distributional learning studies.

[4] A-head language - A-category words: 'nogi', 'tesa', 'waso', 'mito', 'bila', 'sane', 'sito', 'kosi', 'kewa', 'seta', 'sasa', 'tana', B-category word: 'ka'; B-head language - A-category words: 'ka', 'bo', 'ru', 'ni', 'fei', 'pao', 'sa', 'mo', 'gu', 'di', 'tei', 'lao', B-category word: 'nogi'.

All strings used in the experiment were generated by a female voice using an online speech synthesizer, Natural Reader (https://www.naturalreaders.com). We generated each unique string separately such that all strings were generated with the same speed, volume, and pitch.

2.3. Procedure

The experiment consisted of an exposure phase and a test phase. During exposure, we constructed a 144-string exposure corpus for each language condition, which consisted of 36 one-word strings, 74 two-word strings, and 36 A_1-B-A_2 strings that were selected from all the 108 possible A_1-B-A_2 strings such that the frequency of each category-A in each position was balanced. For one-word and two-word strings, all words were attested in the position where they can appear. During the exposure phase, participants heard two repetitions of the exposure corpus presented in random order as they viewed a still, unrelated nature scene (i.e., there was no accompanying referential world). In order to ensure that the participants were paying attention, other sounds such as birds chirping were randomly dispersed among the linguistic strings, and participants were later asked how many such sounds they heard. All participants answered those questions correctly.

The test phase began once the exposure phase was completed. On each test trial, participants heard a test string, and were asked to rate the acceptability of the string on a scale of 1 to 5. Participants were told to decide if the string came from the language they had just heard, where 1 meant definitely not and 5 meant definitely yes. There were 18 two-word test strings (hereby *zero-level* strings) to examine participants' knowledge of the head, 18 one-level test strings (A_1-B-A_2) to examine participants' knowledge of substitutability, and 18 two-level test strings (A_1-B-A_2-B-A_3) to examine participants' knowledge of recursion; at each level beyond zero, the test strings were further divided into types of attested, unattested, and ungrammatical strings (Table 2). Attested strings were strings or, for two-level ones, combinations of two strings that were heard during the exposure phase. Unattested strings were strings where the A_2 or A_3 position was occupied by a word that never appeared after the category-B word during exposure. Ungrammatical strings had completely wrong word order. The prediction is that participants in both conditions should learn headedness at level zero and substitutability at level one, but at level two, participants from the A-head language condition should be more willing to license recursion with both attested and unattested words.

Type	Zero-level	One-level	Two-level
Attested	*sito-mito* (AA), *ka-kewa* (BA)	*nogi-ka-mito*	*nogi-ka-mito-ka-tesa*
Unattested		*bila-ka-tana*	*waso-ka-tesa-ka-sasa*
Ungrammatical	*tesa-ka* (AB)	*nogi-tesa-ka*	*nogi-waso-bila-ka-ka*

Table 2. Sample test strings in A-head language condition. *ka* is the category-B word. The three category-A words *tana*, *tesa* and *sasa* were never attested in the A_2 position of A_1-B-A_2 during exposure; all the other category-A words were attested in both A_1 and A_2 positions.

3. Results

In all our analyses, mixed effects regression was conducted using the lmerTest package (Kuznetsova et al., 2017) in R, where all categorical predictors were simple coded, and participant was included as a random intercept to account for by-participant variance. Post-hoc comparisons were conducted using the emmeans package (Lenth, 2020).

3.1. Headedness (Zero-level)

First, before we investigate whether learners license recursion, it is crucial to determine whether they have learned the intended representation of the head in their assigned language. To test this, we ran a mixed effects regression predicting ratings on zero-level strings by input Condition (A-head vs. B-head), test string Type (attested vs. ungrammatical), and their interaction. There was a significant main effect of test string Type (attested vs. ungrammatical) ($\chi^2(1) = 587.42$, $p < 0.001$), with ungrammatical strings rated significantly lower than attested strings ($\beta = -2.20$, $SE = 0.06$, $t = -34.19$, $p < 0.001$). These results suggest that participants from both conditions have learned the head of the structure.

The model also revealed a significant main effect of Condition ($\chi^2(1) = 4.02$, $p = 0.045$) and a significant interaction between Condition and Type ($\chi^2(1) = 276.43$, $p < 0.001$), indicating that ungrammatical strings were rated higher in the A-head language condition ($\beta = 2.33$, $SE = 0.13$, $t = 18.05$, $p < 0.001$). This result was expected, because the ungrammatical strings in the B-head language were indeed worse than those in the A-head language: The ungrammatical strings in the B-head language were AA, which did not contain the head at all, while the ungrammatical strings in the A-head language were AB, which did contain an A-category word but violated the heads-in-same-linear-order constraint. As shown in Figure 1, all participants rated grammatical zero-level strings significantly higher than ungrammatical, the separation between them is simply greater in the B-head language.

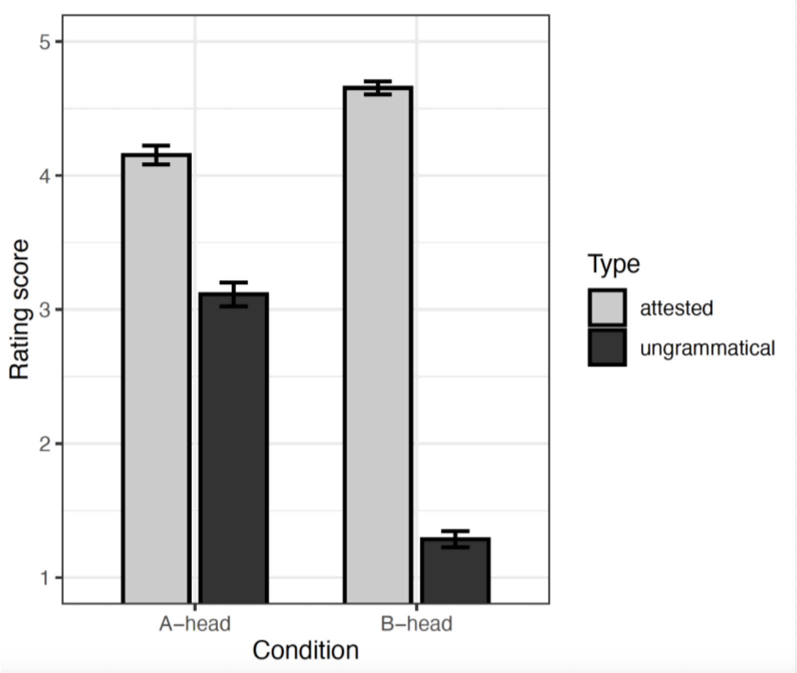

Figure 1. Mean rating scores by test string type in each condition at level zero.

3.2. Substitutability (One-level)

Turning next to the one-level strings, we predicted that participants in both conditions should learn and generalize the rule of linear substitutability for A_1 and A_2 positions. To test this, we computed a learning index and a generalization index to measure how much the participants learned and generalized (Tandoc et al., 2021). The learning index is the difference score of a participant's mean response on Attested test strings minus their mean response on Ungrammatical test strings, (3), and the generalization index is the difference score of a participant's mean response on Unattested test strings minus their mean response on Ungrammatical test strings, (4).

(3) *Learning index* = $M_{attested} - M_{ungrammatical}$

(4) *Generalization index* = $M_{unattested} - M_{ungrammatical}$

Next we ran a mixed effects regression predicting index by Condition (A-head vs. B-head), Type (learning vs. generalization), and their interaction. Model comparison revealed that Type (learning vs. generalization) ($\chi^2(1) = 9.38$, $p = 0.002$) but neither Condition ($\chi^2(1) = 0.72$, $p = 0.40$) nor the interaction of Type and Condition ($\chi^2(1) = 1.10$, $p = 0.29$) was a significant predictor of the index. As shown in Figure 2, this suggests that participants in both conditions have learned and generalized the rule of substitutability for the A_1-B-A_2 structure. The significant main effect of Type showed that the generalization index was generally lower than the learning index, suggesting participants were more willing to accept strings they heard during exposure than unattested strings. Overall, at level one, the two language conditions patterned similarly in both learning and generalization.

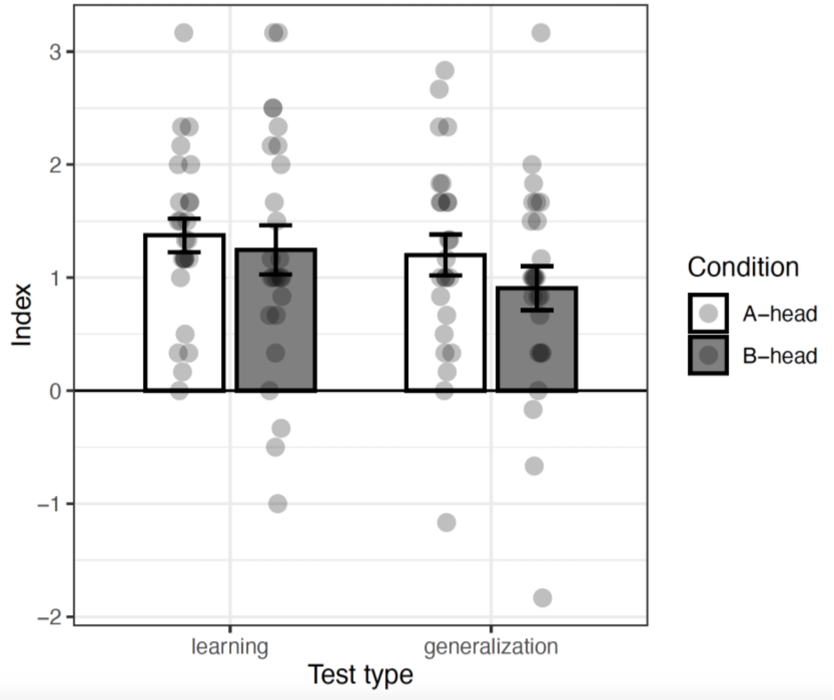

Figure 2. Effects of input condition on learning and generalization at level one. Dots are individual participants and error bars are standard error.

3.3. Recursion (Two-level)

Finally, to test whether participants licensed recursion, we focus on two-level strings (A_1-B-A_2-B-A_3). Participants from the A-head language condition should rate both attested and unattested strings higher than participants from the B-head language condition: although participants from the B-head language condition have heard examples analogous to 'dogs chase cats' and 'cats chase rats', they would not be willing to accept 'dogs chase cats chase rats' because of the head constraint; neither would they be willing to allow recursion for unattested words although they have learned linear substitutability at one-level.

Results at level two are shown in Figure 3. We analyzed the results using the same methods as with one-level strings. Mixed-effects regression showed that both Condition (A-head v. B-head) ($\chi^2(1) = 5.04$, $p = 0.025$) and Type (learning v. generalization) ($\chi^2(1) = 12.46$, $p < 0.001$) but not their interaction ($\chi^2(1) = 1.66$, $p = 0.20$) were significant predictors of the index. The significant main effect of Condition suggested that the learning and generalization indices were higher in the A-head language condition than in the B-head language condition ($\beta = 0.74$, $SE = 0.33$, $t = 2.26$, $p = 0.03$). Therefore, although participants from both conditions have learned linear substitutability in one-level strings (see Section 3.2), participants from the A-head condition were more willing to accept recursively embedded strings using both attested and unattested words. Finally, similar to one-level data, there is also a significant main effect of Type, suggesting the generalization index was lower than the learning index, though post-hoc analyses reported that the learning and generalization index only differed significantly in the B-head language condition ($\beta = 0.49$, $SE = 0.14$, $t = 3.55$, $p < 0.001$).

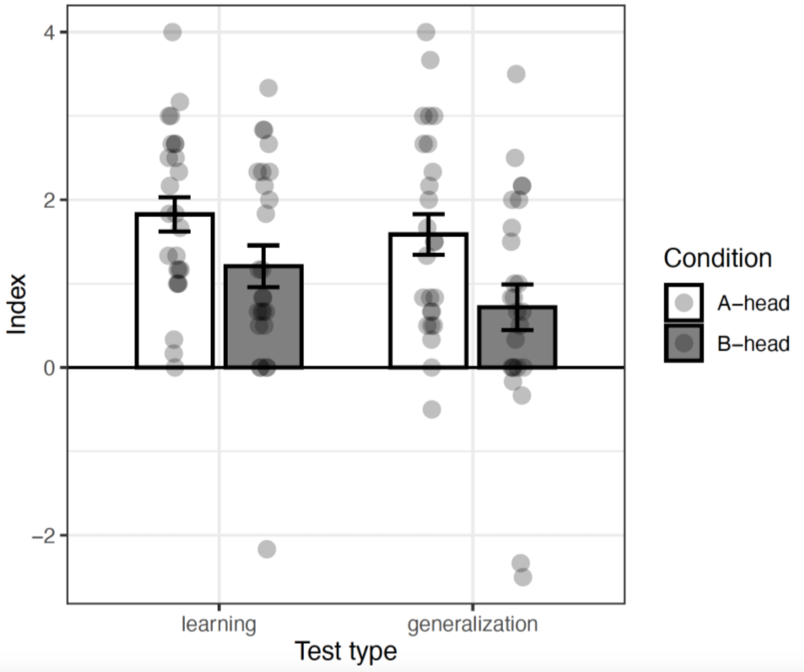

Figure 3. Effects of input condition on learning and generalization at level two. Dots are individual participants and error bars are standard error.

4. General discussion

In summary, this study investigated the role of structural representation in the distributional learning of recursive structures. Previous research has shown that adult learners can use distributional cues for the substitutability of two linear positions to learn recursive structures from non-recursively-embedded data (Li & Schuler, 2023). However, we argued that it is necessary to examine participants' structural representation of the grammar, because the notion of the head is crucial for linguistic recursion: By definition, recursion requires *self*-embedding, so linear substitutability will not guarantee recursion if the substitutable elements are not the head (Li et al., 2021).

Through an artificial language learning experiment, we demonstrated that adult participants can indeed integrate knowledge of linear patterns and syntactic representation to determine whether recursive embedding would be possible: Given sufficient evidence for linear substitutability but different heads, while participants in both languages learned and generalized the rule of substitutability to similar extent, participants in the head-substitutability language condition were significantly more likely to accept recursively embedded strings for either attested words or unattested words.

The current results show that learners can use distributional information to learn the head of a linguistic structure, and integrate this knowledge with other distributional information to acquire complex rules such as recursion. This finding adds to a body of work that investigates how distributional information can be utilized to acquire higher-order linguistic representations (e.g., Takahashi & Lidz, 2007; Thompson & Newport, 2007; Reeder, Newport, & Aslin, 2013; Schuler, Reeder, Newport, & Aslin, 2017; Fetch, 2020). By emphasizing the role of formal learning, though, we do not intend to deny the role of other factors in learning the head and learning recursion. The present study only focused on the role of specific distributional information, and it is worthwhile for future studies to investigate how different types of cues are coordinated and exploited by the learner. Meanwhile, our findings do support an important role of formal, distributional cues during language acquisition, since the participants were able to learn complex linguistic rules when other kinds of cues were not available.

Importantly, the results also show that learning mechanisms like distributional learning are constrained: Even with identical cues to linear substitutability, the participants only licensed recursion when substitutability applied to the head. The results also demonstrate that these constraints can be formally investigated with artificial language experiments. These constraints are crucial to cognition/language acquisition, as they help to restrict the domain over which such learning mechanisms are applied. This allows human learners to efficiently acquire complex systems with minimal resources by focusing them on meaningful patterns.

Finally, the present experiment was conducted with adult participants. However, an important question is whether younger learners can also fully utilize this distributional learning mechanism, since children are known to behave differently from adults in some language learning tasks (e.g., Weir, 1964; Johnson & Newport, 1989; Newport, 1990; Hudson Kam & Newport, 2005; Mayberry & Kluender, 2018; Austin, Schuler, Furlong & Newport, 2022). In ongoing work, we are adapting the current paradigm for child participants, and preliminary results suggest that children can use distributional cues to acquire recursive structures in very similar ways to adults, indicating that this distributional learning mechanism could indeed be helpful for child language acquisition.

Data Availability Statement: The data and code supporting the findings of this study are available to reviewers via the following link: https://osf.io/n54v6/?view_only=2aab3fe44cec4fd29ae036a5c027b3f6 We will make them publicly available when this work is published.

Declaration of Interest Statement: The authors declare no conflict of interest.

Funding: Funding for this work was provided by the University of Pennsylvania to K. Schuler.

Acknowledgements: Omitted for peer review.

References

Adger, D. (2003). *Core syntax*. Oxford University Press

Arsenijević, B., & Hinzen, W. (2012). On the absence of X-within-X recursion in human grammar. *Linguistic Inquiry, 43*(3), 423-440.

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.

Bauer, L. (1978). *The grammar of nominal compounds, with special reference to Danish, English, and French.* Odense University Press.

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 *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.

Deutscher, G. (2000). *Syntactic change in Akkadian: The evolution of sentential complementation*. Oxford University Press.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology, 46*(4), 621-634.

Fetch, A. (2020). *Does learnability predict syntactic universals? An investigation using artificial languages.* Doctoral dissertation, Georgetown University.

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

Goldin-Meadow, S. (1982). The resilience of recursion: A study of a communication system developed without a conventional language model. In *Language acquisition: The state of the art* (pp. 51-77). Cambridge University 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.

Hartmann, K., & Zimmermann, M. (2002). Syntactic and semantic adnominal genitive. In *A-symmetrien – A-symmetries* (pp. 171–202). Stauffenburg.

Haspelmath, M. (2016). The serial verb construction: Comparative concept and cross-linguistic generalizations. *Language and Linguistics, 17*(3), 291-319.

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.

Jelinek, F. (1998). *Statistical methods for speech recognition*. MIT Press.

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.

Kocab, A., Senghas, A., Coppola, M., & Snedeker, J. (2023). Potentially recursive structures emerge quickly when a new language community forms. *Cognition*, *232*, 105261.

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.

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

Li, D. (2024). *Distributional learning of syntactic generalizations.* Doctoral dissertation, University of Pennsylvania.

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). Cascadilla Press.

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

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.

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.

Miller, G. A. (1956). Information and memory. *Scientific American, 195*(2), 42-47.

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

Murphy, G. L. (2002). *The big book of concepts.* Boston Review.

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

Pérez-Leroux, A. T., Peterson, T., Castilla-Earls, A., Béjar, S., Massam, D., & Roberge, Y. (2018). The acquisition of recursive modification in NPs. *Language, 94*(2), 332-359.

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.

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology, 66*(1), 30-54.

Roeper, T. (2007). *The prism of grammar: How child language illuminates humanism.* MIT Press.

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

Roeper, T., & Snyder, W. (2004). Recursion as an analytic device in acquisition. *LOT Occasional Series, 3*, 401-408.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328-350.

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.

Schuler, K. D., Reeder, P.A., Newport, E.L., & Aslin, R.N. (2017). The effect of Zipfian frequency variations on category formation in adult artificial language learning. *Language Learning and Development*, *13*, 357-374.

Takahashi, E., & Lidz, J. (2007). Beyond statistical learning in syntax. In *Proceedings of GALA 2007* (pp. 446–456).

Tandoc, M. C., Bayda, M., Poskanzer, C., Cho, E., Cox, R., Stickgold, R., & Schapiro, A. C. (2021). Examining the effects of time of day and sleep on generalization. *PloS one*, *16*(8), e0255423.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3,* 1-42.

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.