Distributional Learning of Recursive Structures: The Role of the Structural Representation

Daoxin Li and Kathryn Schuler

1. Introduction

Recursion in linguistics refers to infinite self-embedding of a particular type of linguistic element or grammatical structure. The ability for recursion is a crucial part of the language faculty and is considered universally available (e.g., Berwick & Chomsky, 2017). However, languages differ regarding the syntactic domains of recursive structures. For instance, the genitive *-s* is much more restricted in German (1) than in English (2), (Pérez-Leroux, Roberge, Lowles & Schulz, 2022). Therefore, it must be learned from language-specific experience whether a structure allows recursion. So how can learners learn whether a structure can be recursively embedded or not?

(1) *das Manns Nachbars Buch
 the man's neighbor's book
 'the man's neighbor's book'
(2) the man's neighbor's book

In the present study, we examine the role of structural representation in the distributional learning of recursive structures. In Section 2, we introduce the distributional learning proposal of recursive structures (Grohe, Schulz & Yang, 2021; Li, Grohe, Schulz & Yang, 2021), which argues that recursion can be learned from the distributional information of structural substitutability. We present experimental findings that support the proposal, showing that learners can integrate syntactic knowledge of the head in order to distributionally acquire recursion. Specifically, we conducted an artificial language learning experiment introduced in Section 3. Participants learned one of the two artificial languages which both satisfied the substitutability requirement but had different heads. Section 4 reported the results: As predicted, cues of both structural substitutability and the head are needed for the participants to license recursion. Section 5 concludes the paper and discusses the implications.

---

[*] Daoxin Li, University of Pennsylvania, daoxinli@sas.upenn.edu. Kathryn Schuler, University of Pennsylvania. Thank you to Charles Yang and members of the Child Language Lab and the Language and Cognition Lab at University of Pennsylvania for helpful comments and discussion.

 Proceedings of the 47th annual Boston University Conference on Language Development, ed. Paris Gappmayr and Jackson Kellogg, 487-498. Somerville, MA: Cascadilla Press.

2. The distributional learning proposal

In this section we describe a recent proposal of learning recursive structures (Grohe, Schulz & Yang, 2021; Li, et al., 2021), which relies on distributional learning (e.g., Maratsos & Chalkley, 1980; Braine, 1987). The proposal suggests that recursion can be viewed as structural substitutability: a structure such as X_1*'s*-X_2 is recursive if position X_1 and X_2 are productively substitutable, i.e., any word that appears in one of those positions can also be used in the other position. The proposal further claims that children can learn this structural property of substitutability if there are sufficient words attested in both positions in their early language input. Corpus studies on a variety of structures across languages have confirmed that such distributional information is available in simple child-directed speech (English and German adjectives in Grohe, Schulz & Yang (2021); possessives in English, Mandarin and German in Li et al. (2021); English nominal compounds in Yang (2022)). Moreover, a previous artificial language learning experiment demonstrated that learners can indeed use this distributional information to determine which structures allow recursion: Li and Schuler (2021) found that adult participants rated recursive strings (X_1-*ka*-X_2-*ka*-X_3) in an artificial language significantly higher when X_1 and X_2 positions were productively substitutable in their exposure to the artificial language.

However, structural substitutability itself does not necessarily lead to recursion. There are structures that exhibit substitutability in linear position but do not allow recursion. For example, the two NPs in NP_1-*V*-NP_2 in English can be substitutable, but the structure cannot be recursively embedded (e.g., '*dogs chase cats chase rats…'). How does a distributional learner avoid such wrong generalizations? In the distributional learning proposal, an important prerequisite is that the substitutable element must be the head of the structure, since the 'self-embedding' definition of recursion is only satisfied in that case, e.g., N_2 is the head in N_1*'s*-N_2, but neither NP is the head in NP_1-*V*-NP_2. However, it is unknown whether learners will indeed utilize this head information during the distributional learning of recursive structures. In order to test this, we exposed participants to two languages that both had structural substitutability in linear position but differed in whether the substitutable element was the head. The experiment is described in the next section.

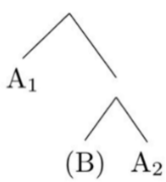
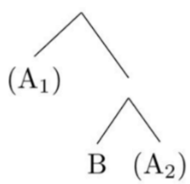
3. Methods
3.1. Participants

Participants were 50 adult native English speakers with typical hearing and vision (or corrected vision). All participants were recruited and run online via Prolific Academic (www.prolific.ac) and paid $9/hour as compensation. The 50 participants were evenly assigned to two language conditions, A-head language (age = 31.2, range = 20-46) and B-head language (age = 29.5, range = 20-45).

3.2. Stimuli

We designed two artificial languages, namely the A-head language and the B-head language. The two languages were constructed such that they can both form one-level A_1-B-A_2 strings, but the head of the A_1-B-A_2 string is different in the two languages: it is A, in particular A_2, in the A-head language, and is B in the other language (Table 1). We approximated the distributional character of heads by implementing the rules that the head of the phrase obligatorily appears whenever the phrase is present, and that non-head elements are optional. By implementing those rules, we do not mean that any non-head element in any language must be omittable. There could be language specific rules that complicates those fundamental rules. Neither do we mean those are the only cues in natural languages for learners to identify the head. We choose those rules because they are key features that define heads in theoretical work on natural languages, and they have been proven useful for learners to identify the head in distributional learning studies (Fetch, 2020). Therefore, the two languages allow different linear strings as shown in Table 1: For one-word strings, the single word must be the head; for two-word strings, the hierarchy and the head determined that *AA* and *BA* but not *AB* are possible in the A-head language, whereas the B-head language allows *BA* and *AB* but not *AA*.

Table 1. Design of two languages.

	A-head language	B-head language
structure		
one-word	A, *B	*A, B
two-word	AA, BA, *AB	*AA, BA, AB

There are 12 category-A words and 1 category-B word in each language. To help participants learn the distinction between head and non-head elements, we used bi-syllabic words for the head and mono-syllabic words for the non-head. All the nonsense words conformed to English phonotactics and are provided in Table 2-3 (adapted from Ruskin (2014)). For the A_1-B-A_2 structure in both languages, A_1 and A_2 are productively substitutable: All of the 12 different words were attested in A_1 position, and 9 of them were attested in A_2 position, clearing the productivity threshold predicted by common measures of productivity (e.g., Bybee, 1995; Yang, 2016). If learners indeed learn recursive structures as predicted by the proposal, i.e., they learn a structure can be recursively embedded if the head positions are substitutable, then they should license recursion (A_1-B-A_2-B-A_3) in the A-head language but not in the B-head language.

For both languages, we constructed a 144-string exposure corpus (Table 2-3). In specific, there were three different types of exposure strings: one-word strings, two-word strings, and A_1-B-A_2 strings. The one-word string for the B-head language was the single category-B word, which was repeated 36 times; for the A-head language, that included each of the 12 category-A words repeated 3 times, thus also making 36 strings in total. The two-word strings for the A-head language included *AA* strings and *BA* strings, and for the B-head language included *AB* strings and *BA* strings. For both languages, the BA strings consisted of 3 repetitions of each of the 12 category-A words following the category-B word. The *AB* strings in the B-head languages consisted of 3 repetitions of each of the 12 category-A words preceding the category-B word. The 36 *AA* strings in the A-head language were selected from all the possible *AA* combinations such that each category-A word appeared in A_1 position three times and appeared in A_2 position three times and A_1 and A_2 were not occupied by the same word. There were also 36 A_1-B-A_2 strings for each language. They were selected from all possible A_1-B-A_2 combinations such that all 12 category-A words were attested in A_1 position 3 times; 9 category-A words were attested in A_2 position 4 times; A_1 and A_2 were occupied by two different words. The 144 exposure strings were divided into three blocks: each block contained 12 one-word strings, 12 *AA* or *AB* strings, 12 *BA* strings, and 12 A_1-B-A_2 strings. The frequency of each word was balanced across three blocks.

Table 2. The distribution of category-A words in the exposure corpus and word frequency in each position in the A-head language.

Word	Freq	One-word	Two-word			A_1-B-A_2	
		A	A_1 in AA	A_2 in AA	A in BA	A_1	A_2
nogi	19	3	3	3	3	3	4
tesa	19	3	3	3	3	3	4
waso	19	3	3	3	3	3	4
mito	19	3	3	3	3	3	4
bila	19	3	3	3	3	3	4
sane	19	3	3	3	3	3	4
sito	19	3	3	3	3	3	4
kosi	19	3	3	3	3	3	4
kewa	19	3	3	3	3	3	4
seta	15	3	3	3	3	3	0
sasa	15	3	3	3	3	3	0
tana	15	3	3	3	3	3	0

The category-B word in the A-head language is *ka*.

Table 3. The distribution of category-A words in the exposure corpus and word frequency in each position in the B-head language from Exp 2.

Word	Freq	One-word	Two-word		A_1-B-A_2	
		B	A in AB	A in BA	A_1	A_2
ka	13	0	3	3	3	4
bo	13	0	3	3	3	4
ru	13	0	3	3	3	4
ni	13	0	3	3	3	4
fei	13	0	3	3	3	4
pao	13	0	3	3	3	4
sa	13	0	3	3	3	4
mo	13	0	3	3	3	4
gu	13	0	3	3	3	4
di	9	0	3	3	3	0
tei	9	0	3	3	3	0
lao	9	0	3	3	3	0

The category-B word in the B-head language is *nogi*.

The test strings included two-word strings to test participants' knowledge of the hierarchical structure (hereby zero-level strings), one-level embedded strings (A_1-B-A_2) to test their knowledge of substitutability, and two-level strings (A_1-B-A_2-B-A_3) to test their knowledge of recursion (Table 4). For zero-level strings, participants in both conditions would be tested on 6 *AB* strings, 6 *BA* strings, and 6 *AA* strings. All the category-A words used for those two-word test strings were among the 9 category-A words that were attested in both A_1 and A_2 positions in A_1-B-A_2. For participants in each condition, two types of zero-level strings would be attested, one type would be ungrammatical. For example, for participants in the A-head language condition, as shown in Table 4, *AA* strings (e.g., *sito-mito*) and *BA* strings (e.g., *ka-kewa*) were attested, while *AB* strings (e.g., *tesa-ka*) would be ungrammatical. One-level strings tested participant's knowledge of substitutability, and two-level strings tested their knowledge of recursion. For both one- and two-level strings, there were attested strings, unattested strings, and ungrammatical strings. For one-level strings, the attested strings were strings that had been heard during exposure phase. For example, in Table 4, *nogi-ka-mito* is a string selected from the exposure corpus. Unattested strings were strings where the A_2 position was occupied a word that was never attested in A_2 position during exposure. For instance, *tana* has never appeared after *ka* in the A_1-B-A_2 structure in the exposure corpus. Ungrammatical strings had the word order A_1-A_2-B, which was not allowed in either of the two languages. For two-level strings, the attested strings were combinations of two one-levels strings that were attested during exposure phase: e.g., *nogi-ka-mito* and *mito-ka-tesa* are both selected from the exposure corpus. In unattested strings, the A_2 and A_3 positions were filled by words that never appeared after the category-B word during exposure, such as *seta* and *sasa* in Table 4. The ungrammatical strings were A_1-A_2-A_3-B-B strings, which were impossible for both grammars. There were 6 strings of each type at each level, leading to 54 test strings in total.

All exposure and test strings 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.

Table 4. Sample test strings in A-head language condition.

Type	Zero-level	One-level	Two-level
attested	*sito-mito, ka-kewa*	*nogi-ka-mito*	*nogi-ka-mito-ka-tesa*
unattested		*bila-ka-tana*	*waso-ka-seta-ka-sasa*
ungrammatical	*tesa-ka*	*nogi-tesa-ka*	*nogi-waso-bila-ka-ka*

3.3. Procedure

The experiment consisted of two phases: exposure, in which participants were exposed to the artificial language, and test, in which participants were tested on how well they learned the language and what generalization they have formed. In the exposure phase, participants were told they would hear strings from a new language, and that they need to pay careful attention to the strings, because they would be tested on their knowledge of the language later. During exposure, 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). There was 1.5s of silence between each string, and participants were offered a break after each block of 36 strings to prevent task fatigue. In order to make sure that the participants were paying attention, other sounds were randomly dispersed among the linguistic strings, such as bird chirping sounds, and participants were later asked how many such sounds they heard. The random sounds occurred only rarely so as not to interfere with the learning of the language (i.e., 2 or 3 times per block). All participants answered those questions correctly.

Once the exposure phase was completed, the test phase began. 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 those strings came from the language they had just heard (e.g., whether they think a native speaker of the language would have said that particular string). 1 meant the string was definitely not from the language; 2 meant the string may not have come from the language; 3 meant the string may or may not have come from the language; 4 meant the string may have come from the language; 5 meant the string definitely came from the language. The test strings were delivered in random order.

4. Results

Next, we present the results for zero-level, one-level and two-level test strings. For one- and two-level strings, we computed a learning index and a generalization index to measure how much the participants learned and generalized. 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). Those indices were calculated separately for one-level and two-level test strings. The predictions are as below. First, for zero-level strings, if participants have learned the correct hierarchical structure, then they are predicted to rate attested strings in their language (i.e., *AA* and *BA* in A-head language, *AB* and *BA* in B-head language) significantly higher than the ungrammatical strings (i.e., *AB* in A-head language, *AA* in B-head language). Next, for one-level strings, participants in both conditions should learn the A_1-B-A_2 structure and the substitutability of A_1 and A_2. Therefore, there should be no difference between conditions in either the learning index or the generalization index. In contrast, for two-level strings, participants from the A-head language condition are predicted to rate both attested and unattested strings higher than participants from the B-head language condition: although participants from the B-head language conditions 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; neither would they be willing to allow recursion for unattested words although they have learned substitutability.

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

4.1. Zero-level

The zero-level data are shown in Table 5. A mixed-effects regression model showed a significant main effect of test string Type (attested vs. ungrammatical) ($\chi^2(1) = 587.42$, $p < 0.001$): Ungrammatical strings were rated significantly lower than attested strings ($\beta = -2.20$, $SE = 0.06$, $t = -34.19$, $p < 0.001$). Post-hoc analyses suggested this holds true in both language conditions (A-head language: $\beta = 1.04$, $SE = 0.09$, $t = 11.41$, $p < 0.001$; B-head language: $\beta = 3.37$, $SE = 0.09$, $t = 36.94$, $p < 0.001$). Therefore, people from both conditions have learned the correct hierarchical structure of the language. 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 ($\beta = 2.33$, $SE = 0.13$, $t = 18.05$, $p < 0.001$). Post-hoc analyses confirmed that participants in the B-head language condition rated ungrammatical strings (i.e., *AA*) lower than those in the A-head language condition (i.e., *AB*) ($\beta = -1.83$, $SE = 0.16$, $t = -11.26$, $p < 0.001$), which led to the significant main effect of Condition. 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 condition were *AA*, which did not contain the head at all; in contrast, the ungrammatical strings in the A-head language condition, *AB*, did contain a category A-word, so it is reasonable that participants from the B-head language condition would rate their ungrammatical strings lower. Post-hoc analyses also showed that grammatical strings in the B-head language condition (i.e., *AB* and *BA*) were rated higher than those in the A-head language condition (i.e., *AA* and *BA*) (β = 0.50, *SE* = 0.14, *t* = 3.47, *p* = 0.001). Further examination showed that this can be attributed to *AA* strings: while *BA* strings in both conditions and *AB* strings in the B-head language condition were rated similarly, *AA* strings in the A-head language condition were rated lower than them. This can be explained by the fact that while all combinations for *AB* and *BA* strings were attested during the exposure phase, there were many possible *AA* strings (144) and only a small proportion of them (36) were selected for the exposure corpus. Therefore, it is expected that participants would be less certain about the judgments of *AA* strings and rate them lower than the other grammatical strings.

Table 5. Mean rating scores for zero-level test strings. Standard errors are in parentheses. Ungrammatical strings are in italics.

Condition	AA	AB	BA
A-head	3.88 (0.27)	*3.11 (0.22)*	4.43 (0.20)
B-head	*1.29 (0.15)*	4.75(0.15)	4.56 (0.18)

4.2. One-level

Figure 1 shows the results for one-level test strings. Mixed-effects regression demonstrated that test 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. In specific, post-hoc analyses revealed that there is no significant difference between two conditions for either the learning index (β = -0.13, *SE* = 0.27, *t* = -0.48, *p* = 0.63) or the generalization index (β = -0.29, *SE* = 0.27, *t* = -1.11, *p* = 0.27). Therefore, this suggests that participants in both conditions have learned the A_1-B-A_2 structure, and have generalized the rule of substitutability of A_1 and A_2 to similar extent. The significant main effect of test Type showed that the generalization index was generally lower than the learning index, suggesting participants were more willing to accept attested strings than unattested strings.

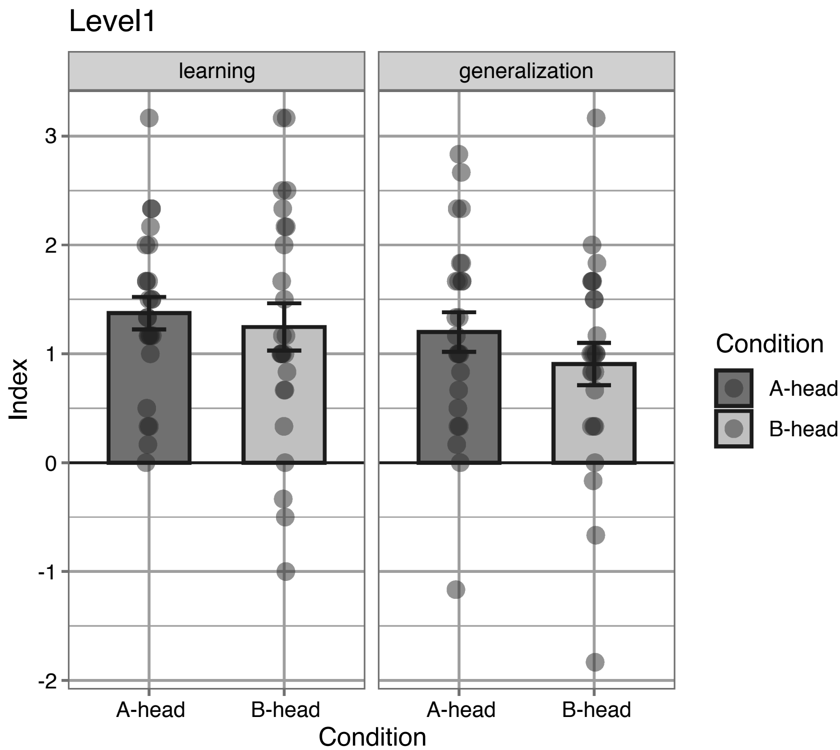


Figure 1. Effects of input condition on learning and generalization at level one. Dots are individual participants and error bars are standard error.

4.3. Two-level

Results at level-two are shown in Figure 2. Mixed-effects regression showed that both Condition ($\chi^2(1) = 5.04$, $p = 0.025$) and test Type ($\chi^2(1) = 12.46$, $p < 0.001$) 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 indices were higher in the A-head language condition than in the B-head language condition ($\beta = 0.74$, $SE = 0.33$, $t = 2.26$, $p = 0.03$). In particular, post-hoc analyses confirmed that the learning index in the B-head language was marginally significantly lower than that in the A-head language ($\beta = -0.62$, $SE = 0.34$, $t = -1.81$, $p = 0.0765$), and the generalization index was significantly lower ($\beta = -0.87$, $SE = 0.34$, $t = -2.52$, $p = 0.01$). Therefore, although participants from both conditions have learned substitutability in one-level strings, participants from the A-head condition were more willing to accept recursively embedded strings for 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 ($\beta = 0.49$, $SE = 0.14$, $t = 3.55$, $p < 0.001$).

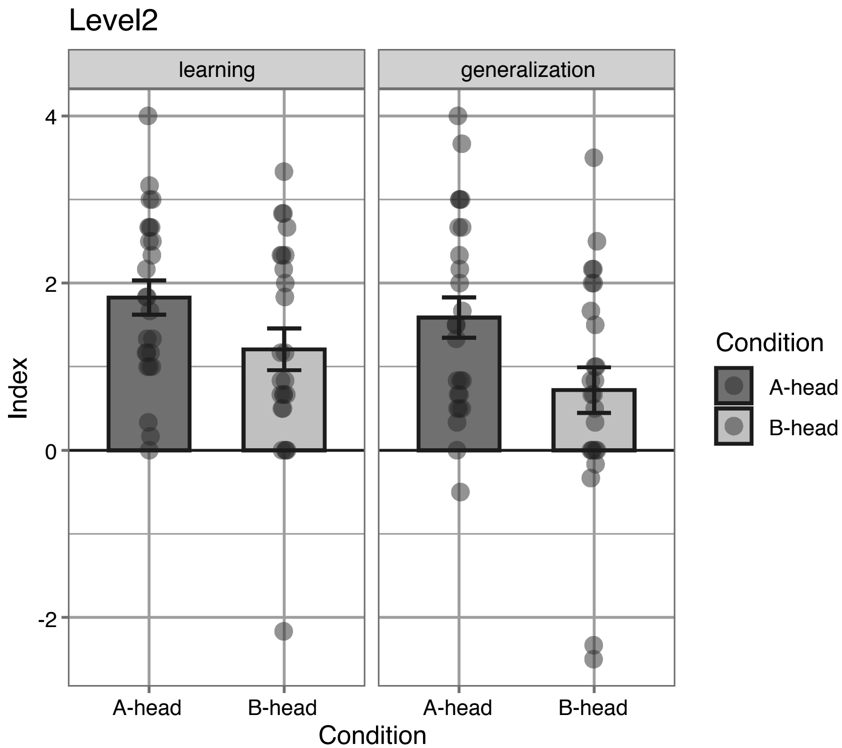


Figure 2. Effects of input condition on learning and generalization at level two. Dots are individual participants and error bars are standard error.

5. General discussion

In summary, through an artificial language learning experiment, we investigated how learners use distributional information about the productivity of structural substitutability to learn recursive structures based on syntactic knowledge of the head of the structure. According to the distributional learning proposal, a prerequisite for structural substitutability to lead to recursion is that the substitutable element must be the head of the structure, because only in that case will self-embedding be involved, which is the definition of recursion. To test the proposal, we exposed participants to two different artificial languages. In both languages, the A_1 and A_2 positions in A_1-B-$A2$ are productively substitutable, but participants could also learn from distributional cues that the head of the structure is A in one language but is B in the other language. At test, we found that as predicted, although participants in both conditions learned substitutability and generalized the rule to unattested words in one-level strings, in the B-head language condition, where recursion was not expected, participants were significantly less likely to accept embedded strings for either attested words or unattested words, thus indicating that learners can integrate knowledge of the syntactic structure to distributionally acquire recursion.

The current results showed that learners can use purely 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., Thompson & Newport, 2007; Takahashi & Lidz, 2008; 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. And as we pointed out earlier, there may be other distributional cues for the head in natural languages in addition to the ones we applied in our current design. The current study only focuses on the role of specific distribution information, but it is worthwhile for future studies to investigate how different types of cues are coordinated and exploited by the learner.

For one- and two-level test strings, besides the predicted effect of condition, we also observed a significant effect of test string type, which indicates that participants tended to rate attested strings higher than unattested strings. It is true that ideally one might expect unattested strings to receive the same scores as attested strings if learners have completely acquired the generalization. However, given the complexity of the linguistic rules to be learned and the short duration of the exposure phase, we do not find this result surprising. In ongoing work, we plan to lengthen the exposure phase so that participants will have longer time and thus more opportunity to learn the rules.

Finally, the present experiment was conducted with adult participants. However, it is unknown whether younger learners can also fully utilize such distributional information, given their more limited cognitive abilities. Previous studies have suggested that children and even infants can learn grammatical rules through distributional learning (e.g., Emond & Shi, 2020; Marcus, Vijayan, Rao, & Vishton, 1999), but the rule to be learned in this study is more abstract than those investigated before. In addition, some studies suggested that distributional learning is an ability available from birth (e.g., Grevain, Macagno, Cogoi, Pena, & Mehler, 2008; Teinonen, Fellman, Naatanen, Alku, & Huotilainen, 2009; Aslin, 2017). Therefore, it is necessary for future research to examine whether young learners exploit the distributional cues in the same way as the adults in the present study, and at what age this distributional learning is available.

References

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

Berwick, Robert, & Chomsky, Noam. (2017). *Why only us.* MIT Press.

Braine, Martin 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, Joan. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5), 425–455.

Emond, Emeryse, & Shi, Rushen. (2021). Infants' rule generalization is governed by the Tolerance Principle. In Danielle Dionne & Lee-Ann Vidal Covas (eds.), *Proceedings of the 45th annual Boston University Conference on Language Development*, 191-204. Somerville, MA: Cascadilla Press.

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

Gervain, Judit, Macagno, Francesco, Cogoi, Silvia, Pena, Marcela, & Mehler, Jacques. (2008). The neonate brain detects speech structure. *PNAS*, 105, 14222-14227.

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

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

Li, Daoxin, & Schuler, Kathryn. (2021). Distributional learning of recursive structures. In *Proceedings of the 43rd Annual Conference of the Cognitive Science Society*, 1437-1443.

Maratsos, Michael P., & Chalkley, Mary 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, Gary F., Vijayan, S., Rao, S. Bandi, & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398), 77-80.

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

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

Ruskin, David. (2014). *Cognitive influences on the evolution of new languages.* Doctoral dissertation, University of Rochester.

Schuler, Kathryn, Reeder, Patricia A., Newport, Elissa L. & Aslin, Richard N. (2017). The effect of Zipfian frequency variations on category formation in adult artificial language learning. *Language Learning and Development*, 13, 357–374.

Takahashi, Eri, & Jeffrey, Lidz. (2008). *Beyond statistical learning in syntax.* Unpublished manuscript.

Teinonen, Tuomas, Vineta Fellman, Risto Naatanen, Paavo Alku & Minna Huotilainen. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10, 21.

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

Yang, Charles. (2016). *The price of linguistic productivity.* MIT Press.

Yang, Charles. (2022). *Productivity and recursion in English compounding.* Paper presented at *the 47th annual Boston University Conference on Language Development.*

Proceedings of the 47th annual Boston University Conference on Language Development

edited by Paris Gappmayr
and Jackson Kellogg

Cascadilla Press Somerville, MA 2023

Copyright information

Proceedings of the 47th annual Boston University Conference on Language Development

ISSN 1080-692X
ISBN 978-1-57473-087-6 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com