

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Distributional learning of recursive structures

Permalink

<https://escholarship.org/uc/item/45221021>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN

1069-7977

Authors

Li, Daoxin
Schuler, Kathryn

Publication Date

2021

Peer reviewed

Distributional Learning of Recursive Structures

Daoxin Li (daoxinli@sas.upenn.edu)

Department of Linguistics, University of Pennsylvania
3401-C Walnut Street, Suite 300, C Wing, Philadelphia, PA 19104 USA

Kathryn Schuler (kschuler@sas.upenn.edu)

Department of Linguistics, University of Pennsylvania
3401-C Walnut Street, Suite 300, C Wing, Philadelphia, PA 19104 USA

Abstract

Languages differ regarding the depth, structure, and syntactic domains of recursive structures. Even within a single language, some structures allow infinite self-embedding while others are more restricted. For example, English allows infinite free embedding of the prenominal genitive *-s*, whereas the postnominal genitive *of* is largely restricted to only one level and to a limited set of items. Therefore, while the ability for recursion is considered as a crucial part of the language faculty, speakers need to learn from experience which specific structures allow free embedding and which do not. One effort to account for the mechanism that underlies this learning process, the distributional learning proposal, suggests that the recursion of a structure (e.g. X_1 's- X_2) is licensed if the X_1 position and the X_2 position are productively substitutable in the input. A series of corpus studies have confirmed the availability of such distributional cues in child directed speech. The present study further tests the distributional learning proposal with an artificial language learning experiment. We found that, as predicted, participants exposed to productive input were more likely to accept unattested strings at both one- and two-embedding levels than participants exposed to unproductive input. Therefore, our results suggest that speakers can indeed use distributional information at one level to learn whether or not a structure is freely recursive.

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

Introduction

This study investigates the learning mechanism that enables speakers to determine which structures are recursive and which are not in a given language. Recursion refers to infinite self-embedding of a particular type of linguistic element or grammatical structure. Many linguists and cognitive scientists agree that the ability for recursion is a crucial part of the language faculty and is universal across languages (e.g. Hauser, Chomsky, & Fitch, 2002; Pinker, 1994)¹. However, languages differ regarding the depth, structure, and syntactic domains of recursive structures (Pérez-Leroux, Peterson, Castilla-Earls, Béjar, Massam, & Roberge, 2018). Indeed, even within a single language, some structures are permitted to freely recurse while others are more restricted. For example, English allows infinite embedding with almost no

constraint using the prenominal genitive *-s*, (1a), whereas the postnominal genitive *of* is limited by a range of semantic, pragmatic, and phonetic constraints and usually does not allow deep embedding, (1b-c), (see Levi (1978), Biber, Geoffrey, Leech, Conrad, & Finegan (1999), and Rosenbach (2014) for extensive discussion). In this paper, we use “recursive structures” to refer to structures that are unrestricted regarding embedding levels and the choice of embedded elements, such as English *s*-genitive, unless otherwise specified. Given the cross- and within-linguistic differences in recursive structures, speakers have to learn from language-specific experience in which syntactic domains the ability of recursion can be freely applied. Thus, what kind of linguistic experience do they use, and how do they make use of it?

- (1) a. the man's neighbor's computer
b. ?the computer *of* the neighbor
c. ??/*the computer *of* the neighbor *of* the man

One line of research has proposed that explicit evidence for deep embedding in the input is necessary to determine that a structure can be used recursively (e.g. Roeper, 2011; Roeper & Snyder, 2005): Under this account, the learner starts with the default assumption that a structure cannot be used recursively, and this assumption is only revised when deep embedding of the structure is observed in the input (e.g. *the man's neighbor's computer*). This mechanism prevents learners from wrongly assuming that limited structures like *of* can freely embed (1b-c); however, it is challenged by the empirical fact that evidence for deep embedding is rarely attested in young children's input. For example, Giblin, Zhou, Bill, Shi, and Crain (2019) examined caregivers' speech in CHILDES (MacWhinney, 2000), where they found only 107 recursive *s*-genitives in 3.1 million English utterances, and no recursive genitives with the productive genitive marker *de* in three Mandarin corpora. These findings predict the acquisition of recursive structures to be very difficult, if not impossible, under the deep embedding approach. Yet, despite the paucity of explicit evidence in the input, several behavioral experiments have reported early acquisition of

¹ We are also aware of a long tradition of research on the learning and processing constraints on recursion, e.g. center embedding (e.g. Christiansen & MacDonald, 2009; Karlsson, 2007; Roth, 1984). Our study, though, does not rely on assumptions of the status of

recursion, and explores the learnability problem from a different approach, namely how can recursive structures be learned from one-level embedding input.

recursive structures. For instance, 4-year-old English- and Mandarin-speaking children can comprehend and produce multi-level recursive *s-* or *de-*genitives (e.g. Giblin et al., 2019; Li, Yang, Roeper, Wilson, Yin, Kim, Merritt, Lopez, & Tero, 2020) — an unexpected finding if children solely rely on direct evidence of deep embedding.

Moreover, there is a logical problem of learning recursive structures: no N -level embedding entails even $N+1$ levels of embedding. Thus, it is difficult for the deep embedding approach to explain how native speakers learn that recursive structures can embed deeper than have been observed in the input, and ultimately, how they learn that recursive structures can stack infinitely when examples in the input are always finite.

Recently, an alternative mechanism for learning recursive structures was proposed (Grohe, Schulz, & Yang, 2021; Li, Grohe, Schulz, & Yang, 2021), which relies on distributional learning (e.g. Brain, 1987; Maratsos & Chalkley, 1980). Specifically, the distributional learning proposal suggests that recursion can be viewed as structural substitutability: That is, for a structure such as X_1 's- X_2 , it is freely recursive if position X_1 and X_2 are productively substitutable, i.e. any noun that appears in one of those positions can also be used in the other position. Therefore, children learn recursion by learning the lexicon for which structural substitutability holds. For example, if the phrases *the mother's car* and *the boy's mother* are attested in one's linguistic input, then the *s-*possessive is recursive at least for the word *mother*, and therefore *mother's mother...* can infinitely embed. Then the learner will seek to form generalizations over attested nouns such as *mother* above: If there is sufficient evidence that structural substitutability is generalizable — that is, if a sufficiently large proportion of nouns attested in one position are also attested in the position, like *mother* — then the child will acquire the generalization that all nouns that appear in one position can also appear in the other, even though not all of them are attested in both positions, and therefore the structure is freely recursive; otherwise, the structure is restricted to certain lexical items. Thus, under this view, children discover whether a structure allows recursion in the same way they discover other productive generalizations in their language.

Importantly, the fact that deep embedding is rarely attested in input to children is no longer a problem under the distributional learning proposal. Learners can discover structural substitutability (and therefore that a structure allows recursion) by utilizing distributional information at one level of embedding. It also addresses the logical problem of learning recursive structures, because it predicts that a structure is either infinitely recursive or must stop at one-level. If a structure is productive at one level, then it allows

infinite embedding; if it does not reach the productivity threshold at one level, then it cannot be freely embedded.

The distributional learning proposal has been supported by findings from corpus studies. Grohe, Schulz, and Yang (2021) found that for *determiner-adjective1-adjective2-noun* strings in English and German input corpora, *adjective1* and *adjective2* are fully substitutable in both languages according to the Tolerance/Sufficiency Principle (Yang, 2016), a model of distributional learning; therefore, the productivity and recursion of prenominal adjective stacking can be learned though distributional cues in the two languages². Li et al. (2021) examined productively recursive and restricted genitive structures in Mandarin Chinese, English, and German, and confirmed for all the recursive structures that a sufficiently large number of nouns appear in both possessor and possessum positions to meet the productivity threshold, in contrast, for all the restricted structures, the proportion of nouns appearing in both positions fail to meet the threshold, thus suggesting that the recursivity of a structure can be learned distributionally from language-specific one-level experience.

In summary, previous corpus studies have provided strong evidence that there is reliable distributional information at first level in the input for learners to acquire recursive structures. However, it is necessary to examine whether learners indeed utilize such distributional information as predicted by the distributional learning proposal. In the present study, we use an artificial language learning paradigm to test the proposal experimentally. In two conditions, participants were exposed to one-level X_1 -*ka*- X_2 strings in an artificial language. We manipulated the distribution of words in the exposure so that the X_1 and X_2 positions are productively substitutable in one condition, but not in the other. At test, we asked participants to rate one- and two-level X_1 -*ka*- X_2 strings that were never attested during exposure, together with attested and ungrammatical controls. If speakers indeed use one-level distributional information to learn recursive structures as predicted by the distributional learning proposal, then participants exposed to productive input should rate the unattested strings higher than participants exposed to unproductive input, since only the former group are predicted to generalize substitutability and recursivity to unattested words. We present and discuss the experiment in the following sections.

Experiment

Methods

Participants 50 adults participated in the experiment: 25 in the Unproductive condition, and 25 in the Productive

² A reviewer pointed out that prenominal adjectives in English do not appear to be freely interchangeable, e.g. 'the big green ball' sounds more natural than 'the green big ball'. We agree with this judgement; but in the definition of the proposal, given 'the big green ball', 'the green big ball' does not have to be attested. It will be sufficient to hear 'adj1 big' (e.g. 'beautiful big') and 'green adj2'

(e.g. 'green wooden'). Besides, although 'the green big ball' is usually not preferred, it is not ungrammatical, and can actually be preferred in certain contexts (e.g. 'Which big ball?' 'The green big ball!'). The distributional learning proposal only concerns whether a structure can possibly be allowed by grammar; the preference will be learned via other mechanisms.

condition. Participants were native English speakers with normal hearing and normal to corrected-to-normal vision recruited online using Prolific (www.prolific.ac). Participants were paid \$9/hour for their participation.

Stimuli We exposed learners to strings generated from an artificial grammar of the form X_1 -ka- X_2 . In addition to the functional morpheme -ka-, the artificial language contained 12 nonsense words adapted from Ruskin (2014), all of which were mono- or bi-syllabic words and conformed to English phonotactics. Some words were more frequent than others in order to imitate word frequency in natural language input, but the total token frequency of each word was kept the same across two conditions.

In both conditions, all 12 words were attested in the X_1 position. Crucially, we manipulated the number of words attested in the X_2 position so that there was sufficient evidence for structural substitutability in the Productive condition (10 of the 12 words attested in X_2) but not in the Unproductive condition (6 of the 12 words attested in X_2). We selected these values because they are consistent with several metrics of productivity. For example, some metrics require a pattern to apply to the majority of types in order to meet the threshold for productive generalization (e.g. Bybee, 1995). Here, such a metric would require at least 7 of our 12 words to be attested in X_2 position. Other metrics require a larger proportion for words to be attested in X_2 position in order to meet the threshold for productivity. For example, the Tolerance/Sufficiency Principle (Yang, 2016) proposes that a rule R defined over N items productively generalizes iff the number of exceptions to the rule is less than or equal to the number of items divided by the natural log of the number of items ($e \leq N/\ln N$). Here, the Tolerance/Sufficiency Principle requires no more than 4 exceptions to structural substitutability ($12/\ln 12 = 4.83$), so at least 8 of our 12 words must also be attested in X_2 position for the rule to generalize. Still other metrics generate an index of productivity — typically a value between 0 and 1 — to capture the intuition that the more items a pattern applies to, the more likely it is to be productive. The *Word-Form Rule* (Aronoff, 1976; Baayen & Lieber, 1991), for example, states that the productivity of a given structure can be quantified as the number of items the structure applies to divided by the number of items it could potentially apply to. Here, our values of 6 (Unproductive condition) and 10 (Productive condition) out of 12 words correspond to a productivity index of 0.50 and 0.83 respectively. We would like to point out that it is not our goal to distinguish the different metrics with the current experiment; instead, our stimuli were designed to meet all of them to make sure that only one condition provides productive input.

In both conditions, the most frequent word was attested in both positions. The words not occurring in position X_2 included both words of higher token frequency and words of lower token frequency. We randomly combined the X_1 words and the X_2 words to create an exposure corpus of 44 X_1 -ka- X_2 strings for each condition. The distribution of the words and their frequencies across conditions and X -positions are shown in Table 1³.

Table 1: The distribution of words in the 44 string exposure corpus and word frequency in X_1/X_2 position

Word	Frequency	Unproductive		Productive	
		X_1	X_2	X_1	X_2
nogi	36	6	30	12	24
sane	10	10	0	10	0
tesa	6	6	0	3	3
waso	6	6	0	3	3
sito	6	2	4	3	3
kosi	6	2	4	3	3
mito	4	2	2	2	2
kewa	4	2	2	2	2
bila	4	2	2	2	2
seta	2	2	0	1	1
sasa	2	2	0	1	1
tana	2	2	0	2	0
Total	88	44	44	44	44

Procedure The experiment consisted of an exposure phase and a test phase. In the exposure phase, participants were told they would hear strings from a new language, and their task was to pay attention to the strings, because they would be tested on their knowledge of the language later. The exposure phase consisted of two repetitions of the exposure corpus (44 X_1 -ka- X_2 strings) presented in pseudo-random order, with no accompanying referential world. There was 1.5s of silence between each string, and participants were offered a break after each repetition of the 44 strings to prevent 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. 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

³ A reviewer suggested that the distribution is potentially problematic since in the Productive condition, we have the vast majority of the stimuli featuring one single very high frequent word. But as shown in Table 1, the most frequent word is productively attested in both X_1 and X_2 positions in the Unproductive condition

as well; and as our results will show, participants in the Unproductive condition do not generalize even though the dominantly frequent word exhibits substitutability. Therefore, we think there is strong evidence that learners use the overall distributional information to learn recursive structures.

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 generated to include either one (X_1 - ka - X_2) or two levels (X_1 - ka - X_2 - ka - X_3) of embedding. At each level, there were three types of test strings: attested strings, unattested strings, and ungrammatical strings. Attested strings were strings or combinations of two strings that had been heard during exposure. For example, in the sample attested test strings in Table 2, *waso-ka-mito*, *sane-ka-kewa*, and *kewa-ka-nogi* have all been attested. Unattested strings were strings or combinations of two strings whose post-*ka* position (X_2 or X_3) was occupied by a word that never appeared in X_2 position during exposure. Thus, in the unattested strings in Table 2, *sane*, *tesa* and *tana* have never been attested after *ka*. Finally, ungrammatical strings were strings with wrong word order, such as *ka-X₁-X₂* or *ka-X₁-X₂-X₃-ka*. There were six test strings of each type at each level, leading to 36 test strings in total. We designed our test strings such that in each string type, there were both words of higher frequency and words of lower frequency, in order to avoid the influence of token frequency in the test. The test strings were delivered in random order.

Table 2: Sample test strings in Unproductive condition.

Type	One-level	Two-level
attested	<i>waso-ka-mito</i>	<i>sane-ka-kewa-ka-nogi</i>
unattested	<i>nogi-ka-sane</i>	<i>waso-ka-tesa-ka-tana</i>
ungrammatical	<i>ka-bila-kosi</i>	<i>ka-waso-kosi-sito-ka</i>

In both conditions, participants are expected to rate attested strings significantly higher than ungrammatical strings at both levels. Of particular interest are the unattested strings. According to the distributional learning proposal, only participants in the Productive condition would learn position X_1 and position X_2 are productively substitutable in the X_1 - ka - X_2 structure, and thus generalize this pattern to unattested words: If a word appeared in position X_1 during exposure, it must be able to appear in position X_2 as well, even though it was never attested there in the input. On the other hand, participants in the Unproductive condition would not learn that position X_1 and position X_2 are productively substitutable: for words that only appeared in position X_1 , they would be more likely to assume that they cannot appear in position X_2 than participants in the Productive condition. Therefore, it is predicted that participants in the Productive condition would rate one-level unattested strings higher than participants in the Unproductive condition. Furthermore, given the productivity of the structure at level-one, participants in the Productive condition would acquire the generalization that all of the 12 words can be freely embedded in recursive X_1 - ka - X_2 structures, but for participants in the Unproductive

condition this structure cannot be generalized beyond the attested examples. Thus, participants in the Productive condition are predicted to rate two-level unattested strings higher than participants in the Unproductive condition as well.

Results

The rating scores for one-level and two-level test strings were shown in Figure 1 and Figure 2 respectively.

We analyzed the results using ordinal regression, with rating score as an ordered factor from 1 to 5, test string Type (attested, unattested, or ungrammatical) and Condition (Unproductive, Productive) as fixed effects, and by-participant random intercepts and random slopes for Type. At one level, there was no main effect of Condition ($\chi^2(1) = 0.02$, $p = 0.90$), but there was a significant main effect of Type ($\chi^2(2) = 253.00$, $p < 0.001$). Post-hoc pairwise comparisons suggested that in the Unproductive condition, unattested strings were rated significantly lower than attested strings ($\beta = 1.65$, $SE = 0.22$, $z = 7.43$, $p < 0.0001$), and significantly higher than ungrammatical strings, but to a lesser degree ($\beta = 0.75$, $SE = 0.21$, $z = 3.57$, $p = 0.001$); in the Productive condition, unattested strings were rated significantly lower than attested strings ($\beta = 1.15$, $SE = 0.22$, $z = 5.20$, $p < 0.0001$), and significantly higher than ungrammatical strings ($\beta = 1.62$, $SE = 0.22$, $z = 7.51$, $p < 0.0001$). There was also a significant interaction between Type and Condition ($\chi^2(2) = 8.67$, $p = 0.01$). Thus, overall, in both conditions, unattested strings were rated lower than attested strings and higher than ungrammatical strings, but unattested strings were regarded more similar to ungrammatical strings in the Unproductive condition than in Productive condition. Comparisons between conditions showed that unattested strings were rated marginally lower in the Unproductive condition than in the Productive condition ($\beta = -0.48$, $SE = 0.28$, $z = -1.71$, $p = 0.09$).

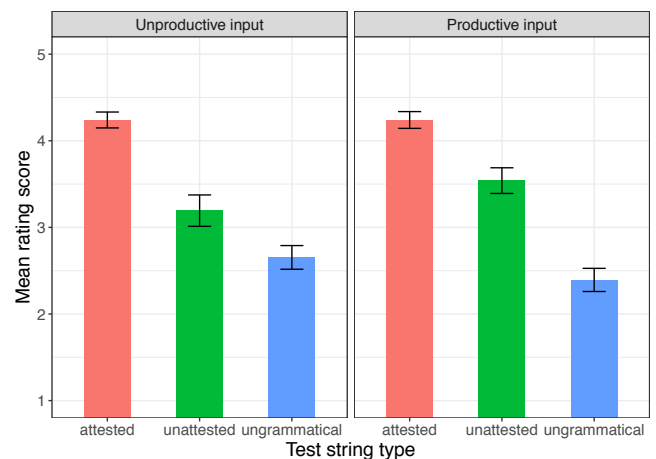


Figure 1: Mean rating scores for each type of one-level test strings. Error bars indicate standard errors of the mean.

For two-level test strings, again there was no main effect of Condition ($\chi^2(1) = 0.06, p = 0.81$), but a significant main effect of Type ($\chi^2(2) = 323.82, p < 0.001$). Post-hoc pairwise comparisons revealed that in the Unproductive condition, unattested strings were rated significantly lower than attested strings ($\beta = 3.35, SE = 0.25, z = 13.45, p < 0.0001$), but were not different from ungrammatical strings ($\beta = 0.26, SE = 0.22, z = 1.18, p = 0.47$); in the Productive condition, unattested strings were rated significantly lower than attested strings ($\beta = 1.10, SE = 0.21, z = 5.22, p < 0.0001$), and significantly higher than ungrammatical strings ($\beta = 1.39, SE = 0.22, z = 6.44, p < 0.0001$). There was a significant interaction between Type and Condition as well ($\chi^2(2) = 52.74, p < 0.001$). Therefore, the two conditions shared the pattern that the rating scores for unattested strings were between those for attested strings and ungrammatical strings, but unattested strings were rated significantly higher than ungrammatical strings only in the Productive condition. Comparison between conditions showed that unattested strings were rated significantly lower in the Unproductive condition than in the Productive condition ($\beta = -1.18, SE = 0.39, z = -3.02, p < 0.01$). These results suggest that, while learners in the Productive condition learned that the X_1 - ka - X_2 structure allows recursion, learners in the Unproductive condition treat additional levels of embedding as ungrammatical.

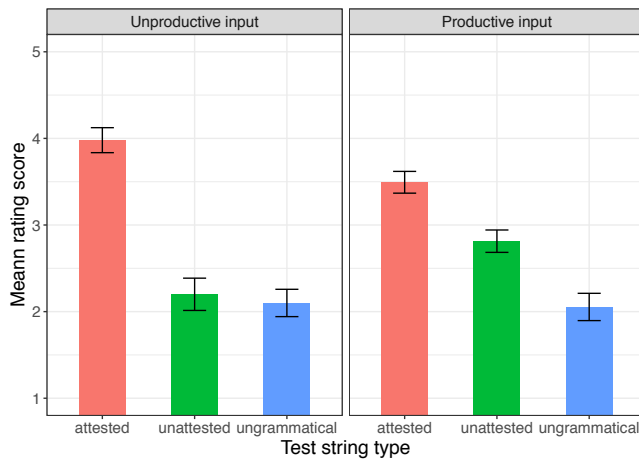


Figure 2: Mean rating scores for each type of two-level test strings. Error bars indicate standard errors of the mean.

In summary, at both one level and two level, participants rated unattested strings lower than attested strings and higher than ungrammatical strings, but unattested strings were rated higher in the Productive condition than in the Unproductive condition. In particular, recursive unattested strings were only distinguished from ungrammatical controls in the Productive condition. Therefore, the results suggest that as the distributional learning proposal predicted, speakers can use one-level distributional information to learn about recursive structures.

General Discussions

In this study, we investigated whether speakers can learn freely recursive structures purely based on the productivity of the structure in simple one-level embedding data. The distributional learning proposal argues that for a structure such as X_1 - ka - X_2 if a large enough proportion of words are attested in both the X_1 and X_2 positions in one-level input, then speakers can acquire the generalization that the two positions are productively substitutable, so a word attested in one position is able to appear in the other position as well, even though it has never been attested in the other position in the input. Furthermore, once a structure is productive at one level, speakers will learn it can be freely embedded to any level. In contrast, if the number of words attested in both positions in the input does not reach the productivity threshold, speakers will assume the positions are not substitutable and thus the structure cannot be embedded further, except for specific items that have been attested in the input. We found that as predicted, participants exposed to productive input were more willing to accept one- and two-level unattested strings than participants exposed to unproductive input; moreover, only participants exposed to productive input treated recursive unattested strings as significantly better than ungrammatical strings. Therefore, our results suggest that learners can indeed access the distributional information and perform the necessary computations as the distributional learning approach proposes. Together with previous corpus studies which demonstrated the availability and reliability of distributional information about structural productivity in naturalistic data (Grohe, Schulz, & Yang, 2020; Li et al., 2020), the findings indicate that the recursivity of a structure can be learned distributionally from language-specific level-one experience. Therefore, overall, the results imply that recursivity can be viewed as a productive generalization, which can be acquired through distributional learning. This learning mechanism also avoids the logical problem of learning recursive structures, since it does not rely on explicit evidence of deep embedding; instead, it predicts that a structure can be freely embedded once it is productive at one level. Therefore, this learning mechanism enables speakers to acquire knowledge of infinite embedding from finite input data.

The results of this study add to a body of work that investigates how distributional information can be utilized to acquire higher-order linguistic representations (e.g., Reeder, Newport, & Aslin, 2013; Schuler, Reeder, Newport, & Aslin, 2017). We would like to make it clear that we are not arguing that children acquire the *ability* of recursion through distributional learning. Instead, we are interested in whether learners can use distributional information to learn to which specific structures recursion can be freely applied, which must be learned from language specific experience. Furthermore, the present study is focused on what speakers can learn about recursive structures from distributional information alone, and our results indicate distributional information itself already allows learners to distinguish structures that can be embedded freely from those that cannot.

However, we do not deny the important role of other factors, such as the well-recorded semantic, pragmatic, and phonetic constraints for the English *of*-genitive (e.g. Rosenbach, 2014), in the acquisition of recursive structures. Rather, we consider this work a first step toward future investigations into how learners coordinate and exploit different cues to learn which structures are recursive and the constraints on this recursion in the language they are acquiring.

There is a curious question about the results: namely, if unattested strings were considered acceptable in the Productive condition, why were they still rated lower than attested strings? Although ideally one might expect unattested strings to receive the same scores as attested strings, we do not find this result surprising, given the complexity of the structure to be learned and the short duration of the exposure phase. Indeed, even experiments with natural language found that ratings from native speakers get lower as the number of embeddings in a sentence go up (e.g. Christianson & MacDonald, 2009). Crucially, only participants in the Productive condition of our experiment distinguished recursive unattested strings from ungrammatical strings, and they always rated unattested strings higher than participants in the Unproductive condition.

Another important question is whether the learners in our experiment acquired a hierarchical structure from the artificial language input or if they simply acquired the linear order of strings. Generalizing the X_1 -*ka*- X_2 structure to X_1 -*ka*- X_2 -*ka*- X_3 involves tail-recursion, which, in the absence of a referential world, could be accomplished with simple iteration. We agree that our design does not rule out the possibility that learners may not have acquired a hierarchical structure from our language input. However, some artificial language learning studies have found that if human learners can apply certain distributional learning strategy to linear strings, they are also able to apply it to hierarchical structures (Takahashi & Lidz, 2008; Thompson & Newport, 2007). Therefore, even though what our participants have learned is a linear structure, we think they are also likely to learn hierarchical structures with the same mechanism. We plan to test this by constructing an explicitly hierarchical language as in Takahashi & Lidz (2008) and Thompson & Newport (2007).

Another possible interpretation of the results is that participants were learning categories: In the Productive condition, they learned all the words belong to one productive category, whereas in the Unproductive condition, they learned the words belong to different categories and are thus uninterchangeable. We suggest this interpretation is not necessarily inconsistent with the distributional learning proposal. For example, the corpus study in Li et al. (2021) showed that for freely recursive structures in English, Mandarin, and German, there is no constraint on the words that can appear in either position, i.e. all words can be viewed as belonging to one productive category; in contrast, for restricted structures, the words which can be used in certain position do form semantic subcategories. For instance, for the restricted possessive structure X_1 's- X_2 in German, words that

are attested in X_1 are limited to close kinship terms. We will examine the exact relation and distinction between categories and recursion in future research.

Finally, the present experiment was conducted with adult participants. However, it is unknown whether young 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. Aslin, 2017; Grevain, Macagno, Cogoi, Pena, & Mehler, 2008; Teinonen, Fellman, Naatanen, Alku, & Huottilainen, 2009). 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.

Acknowledgments

Thanks to our participants for participating in the experiment; to members of the Child Language Lab and the Language and Cognition Lab at University of Pennsylvania for discussion and feedback; to Charles Yang for helpful comments; 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.
- Baayen, H. & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. *Linguistics*, 29(5), 801–844.
- 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. (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.
- Emond, E., & Shi, R. (2020). *Infants' rule generalization is governed by the Tolerance Principle*. Paper presented at the 45th annual Boston University Conference on Language Development.
- Gervain, J., Macagno, F., Cogoi, S., Pena, M., & Mehler, J. (2008). The neonate brain detects speech structure. *PNAS*, 105, 14222–14227.
- Giblin, I., Zhou, P., Bill, C., Shi, J., & Crain, S. (2019). The Spontaneous eMERGEance of recursion in child language.

- Proceedings of the 43rd annual Boston University Conference on Language Development* (pp. 270-286). Somerville, MA: 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 biannual conference on Generative Approaches to Language Acquisition – North America.
- Hausser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579.
- Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, 43, 365–392.
- Levi, J. N. (1978). *The syntax and semantics of complex nominals*. Academic Press.
- Li, D., Grohe, L., Schulz, P., & Yang, C. (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, D., Yang, X., Roeper, T., Wilson, M., Yin, R., Kim, J., Merritt, E., Lopez, D., & Tero, A. (2020). Acquisition of recursion in child Mandarin. *Proceedings of the 44th annual Boston University Conference on Language Development* (pp. 294-307). Somerville, MA: Cascadilla Press.
- MacWhinney, B. (2000). *The CHILDES Project*. Lawrence Erlbaum.
- 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.
- 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.
- Pinker, Steven. (1994). *The language instinct*. William Morrow and Company.
- Rosenbach, A. (2014). English genitive variation – The state of the art. *English Language and Linguistics*, 18, 215-262.
- 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. (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*. John Benjamins.
- 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.
- 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., & Jeffrey, L. (2008). Beyond statistical learning in syntax. Unpublished manuscript.
- 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.
- Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development*, 3, 1-42.
- Yang, C. (2016). *The price of linguistic productivity*. MIT Press.